

Antioch University

AURA - Antioch University Repository and Archive

Antioch University Full-Text Dissertations &
Theses

Antioch University Dissertations and Theses

2022

Examining Gender Differences in a Forensic Sample Using the Personality Assessment Inventory

Rebecca Elliott

Follow this and additional works at: <https://aura.antioch.edu/etds>



Part of the [Clinical Psychology Commons](#)

EXAMINING GENDER DIFFERENCES IN A FORENSIC SAMPLE USING THE
PERSONALITY ASSESSMENT INVENTORY

A Dissertation

Presented to the Faculty of
Antioch University New England

In partial fulfillment for the degree of

DOCTOR OF PSYCHOLOGY

by

Rebecca Jean Elliott

ORCID Scholar No. 0000-0001-5235-4618

November 2022

EXAMINING GENDER DIFFERENCES IN A FORENSIC SAMPLE USING THE
PERSONALITY ASSESSMENT INVENTORY

This dissertation, by Rebecca Jean Elliott, has
been approved by the committee members signed below
who recommend that it be accepted by the faculty of
Antioch University New England
in partial fulfillment of requirements for the degree of

DOCTOR OF PSYCHOLOGY

Dissertation Committee:

Kathi Borden, PhD, Chairperson

Monique Bowen, PhD

Thomas Powell, PhD

Copyright © 2022 by Rebecca Jean Elliott
All Rights Reserved

ABSTRACT

EXAMINING GENDER DIFFERENCES IN A FORENSIC SAMPLE USING THE PERSONALITY ASSESSMENT INVENTORY

Rebecca Jean Elliott

Antioch University New England

Keene, NH

Rates of female criminality appear to be rising (The Sentencing Project, 2022), and thus more women are likely to present for evaluation in a forensic capacity. A majority of research in the field of criminal behavior has focused exclusively on male populations. A dearth of research examining risk factors among female offenders has led to a lack of empirically validated tools used to assess women. Additionally, research on gender differences within a forensic population representing a wide range of referral questions has not been conducted. Women offenders present with unique and different risk factors than men (Grimbos et al., 2016); thus, gender-responsive assessment is warranted to facilitate the development of informed clinical opinions regarding risk, needs, and recommended treatment. The Personality Assessment Inventory (PAI) is a widely used self-report measure of personality (Morey, 2007) that assists forensic evaluators by offering additional clinical information relevant to psychopathology and treatment. The present quantitative study was designed to examine gender differences in a unique forensic sample using the PAI, providing information about the profiles of males and females referred for forensic evaluation. This study expanded the knowledge base regarding differences and similarities in psychopathology between men and women within a forensic population. These findings highlight the importance of maintaining a gender-responsive approach to forensic assessment that will lead to more valid assessment results and treatment recommendations. The PAI is discussed regarding its utility within gender-responsive assessment in a sample of this

kind. Recommendations for forensic evaluators are also offered. This dissertation is available in open access at AURA (<https://aura.antioch.edu>) and OhioLINK ETD Center (<https://etd.ohiolink.edu>).

Keywords: forensic evaluation, personality assessment inventory, gender

Dedication

This dissertation is dedicated to my family. Without your support, this dissertation would not exist. To my parents, I would not have believed I could pursue this degree had it not been for your never ending love and guidance. Even when I doubted myself, you were there to lift me up and push me forward. No words can convey how grateful I am for everything you have done. To my sister, my twin flame, I am who I am because of you. You traveled across the country with me because of this degree, and I cannot imagine having anyone else by my side. After everything you all have sacrificed, this one's for you.

Acknowledgments

Something that has been quoted by many different people, many times, over the years is that you never accomplish anything alone; community truly is the foundation where everything else can grow. There are so many people who have supported me over the last five years. To Dr. Kathi Borden, my dissertation chair and one of the hardest working individuals I have ever (and likely will ever) meet. Your kind words and critical eye have guided me throughout this entire process. To Dr. Thomas Powell, one of my first ever practicum supervisors and the sole reason I pursued forensic psychology, I am forever grateful. To Dr. Monique Bowen, I am lucky to have had you as a professor in the course where I first discovered my passion for assessment. Janelle, Jordyn, and Sara—our friendship has enabled me to grow as a person and influenced who I am as a clinician. I am so grateful to know each of you. I could not ask for better colleagues and friends. To my doc sem cohort, Emily and Quynh, thank you for all of the laughs and listening ears. To Lindsay, I would never have discovered Antioch without you. I am here because of you. And to the rest of my friends and family members who have lovingly supported me over the last five years—it means more than you know. Again, to you all, thank you, thank you, thank you.

Table of Contents

Abstract	iv
Dedication	vi
Acknowledgments.....	vii
List of Tables	x
CHAPTER I: INTRODUCTION.....	1
Purpose of the Present Study	3
Definition of Gender.....	4
Research Questions and Overall Hypotheses	4
CHAPTER II: REVIEW OF THE LITERATURE	6
Gender Differences in Forensic Populations	6
Unique Considerations of Gender in Forensic Assessment.....	8
The Personality Assessment Inventory (PAI).....	11
Summary.....	19
CHAPTER III: METHOD	21
Design.....	21
Participants	21
Measures.....	22
Procedure.....	23
Analysis	23
Inferential Tests of VTFA Sample	23
Research Question One and Two	24
Research Question Three.....	24
CHAPTER IV: RESULTS.....	25
Demographics.....	25

Screening for Outliers.....	25
Descriptive Statistics	26
Comparison of Women in the Archival vs. Women in the Normative Sample on the PAI	26
Comparison of Men in the Archival vs. Men in the Normative Sample on the PAI.....	28
Comparison of Men vs. Women in the Archival Sample on the PAI.....	29
CHAPTER V: DISCUSSION.....	31
Comparisons Between the Archival and Normative Sample on the PAI	32
Gender Differences within the Archival Sample on the PAI.....	40
Limitations.....	43
Future Considerations.....	44
Clinical Implications and Recommendations for Evaluators	46
Conclusion.....	47
References.....	50
APPENDIX A: PERMISSIONS FORM	54
APPENDIX B: DESCRIPTION OF PAI SCALES	55
APPENDIX C: SYNOPSIS OF MODAL CLUSTER PROFILES	57
APPENDIX D: TABLES.....	58

List of Tables

Table 1 Demographic Characteristics of Sample by Gender..... 58

Table 2 PAI Sample Means by Gender and Normative Comparisons..... 59

Table 3 PAI Sample Means and Differences by Gender 60

CHAPTER I: INTRODUCTION

The rate of female incarceration has been rising at a rate greater than males since 2000 (Davidson et al., 2015). Research has shown an upward trend of female criminality across numerous countries (Ndung'u et al., 2021). Over the past 20 years, the number of women who have committed violent offenses has been escalating (The Sentencing Project, 2022). Although women still comprise the minority of the forensic mental health and prison populations, the rate of women involved in these services has increased over the last two decades (de Vogel & Nicholls, 2016). Given this information, it is likely that more women are presenting for forensic evaluation than in the past.

Forensic evaluation (also referred to as a forensic assessment) is a category of psychological assessment in which forensic psychologists gather information to assist legal decision makers, including giving expert testimony in court. As a result, forensic psychologists must display excellent clinical skills in conjunction with knowledge of the relevant legal systems (i.e., criminal and civil). Forensic assessment is comprehensive and integrates multiple sources of information when addressing the referral question, such as reviewing relevant records, conducting clinical interviews, and administering psychological tests. Opinions and recommendations offered in these evaluations have significant implications for legal outcomes, such as sentencing.

When evaluating an individual in a forensic capacity, consideration of the individual's cultural identities, including gender, is essential (American Psychological Association [APA], 2013). Most research on pathways to offending has exclusively focused on men (McKeown, 2010), which is understandable given that men comprise the majority of incarcerated individuals (The Sentencing Project, 2022). In response to the increasing prevalence of justice-involved

women, there has been increased interest in understanding female offending behavior over the last several years, including consideration of which factors are relevant to female criminal behavior (Rettinger & Andrews, 2010). Research shows that female pathways to offending differ from those of males (McKeown, 2010), and female offenders present with different risk factors for violence compared to male offenders (Grimbos et al., 2016). Given this information, understanding gender differences within various forensic settings and how they impact forensic evaluation is of greater relevance today than ever.

Men and women share some similarities in the factors that influence antisocial behavior; however, assessments with females should be approached differently than with males (de Vogel & Nicholls, 2016). Arguably, one of the most essential methods of assessment in forensic evaluation is the administration of psychological tests. Despite the recent surge of research in female offending, there remains a dearth of forensic assessment tools validated explicitly for use among women. Empirically validated assessments are necessary to avoid evaluators overly relying on unstructured clinical judgment, leading to inaccurate conclusions and recommendations when, for example, predicting future violence and determining risk for recidivism (Hilton et al., 2006; Monahan, 1981). Given the paucity of gender-specific assessment tools, forensic evaluators must utilize other tools that have not been specifically validated for women in formulating clinical profiles.

One such tool is The Personality Assessment Inventory (PAI), a self-assessment inventory of personality widely accepted in various forensic and psycholegal applications (Morey, 2007). While the PAI is not designed to be used as a risk assessment tool, it plays a role in forensic risk assessment by providing other clinical information about the examinee. The PAI addresses psychopathology and treatment planning rather than specific risk-related factors.

Similar to most other available assessment tools, the PAI is not a gender-specific assessment tool. Knowledge of gender differences on the PAI may facilitate gender-responsive forensic assessment, reducing the potential for inappropriate recommendations and decision making. Gender differences observed on PAI scales were examined in the current study, and suggestions related to the importance of using the PAI in a gender-responsive manner are discussed.

Purpose of the Present Study

Women in forensic populations have more significant mental health needs; thus, they should be assessed accordingly to ensure valid clinical profiles and appropriate recommendations for treatment (de Vogel & Nicholls, 2016). The increased prevalence of females presenting for forensic evaluation and the importance of appropriate gender-responsive assessment was the impetus for the current study. The present study investigated gender differences within a unique forensic sample using the PAI to provide information about how men and women differ on measures of validity of responses, psychopathology, treatment considerations, and interpersonal style when presenting for forensic evaluation.

Most research on women within forensic settings has focused exclusively on offender or inpatient populations, limiting the understanding of women to those specific contexts. The present archival sample is unique as it consists of individuals presenting for various purposes necessitating forensic evaluation in a myriad of settings. Research describing gender differences within a forensic sample of this kind has not been examined. This study addressed that gap in the literature. This study was exploratory in nature and contributed to research within the realms indicated above.

Definition of Gender

Important to discuss as it relates to the present study is how gender is defined. Gender and sex do not refer to the same construct (de Vogel & Nicholls, 2016). Sex is specific to the biological aspects of males and females, while gender refers to the “psychological, social, and cultural aspects associated with the biological aspects (de Vogel & Nicholls, 2016, p. 2). The terms gender-specific, gender-sensitive, and gender-responsive are terms used throughout this study (adopted from de Vogel & Nicholls, 2016), which views gender as “a factor that goes both ways” (p. 2), meaning practices and policies adopted should be best suited for men, and vice versa for women. Despite established gender differences, there is debate on whether a gender-neutral or gender-specific perspective is more helpful in understanding female criminal behavior (Rettinger & Andrews, 2010). This study argues that a gender-neutral approach is not sufficient in capturing the nuances and implications that gender has on forensic evaluation. In the present study, for ease of statistical analysis and comparison, gender is defined using binary labels (i.e., men and women; male and female) due to the limited diversity of gender identity in the data that will be analyzed. Given that gender identity exists on a spectrum rather than as a binary, this study’s narrow and outdated view of gender is considered a limitation, which is discussed further in Chapter V.

Research Questions and Overall Hypotheses

1. How do the overall mean scores on any PAI scales for females within the archival sample compare to the test developer’s normative community sample scores among females?
2. How do the overall mean scores on any PAI scales for males within the archival sample compare to the test developer’s normative community sample scores among males?

3. Are there significant differences between males and females within the archival sample on any PAI validity, clinical, treatment, or interpersonal scales?

Given that this was an exploratory analysis, no specific formal hypotheses were offered.

However, given prior research regarding the higher rates of antisocial personality disorder and alcoholism among men than women, it was expected that Antisocial Features (ANT) and Alcohol Problem (ALC) scores would be higher among men than women.

CHAPTER II: REVIEW OF THE LITERATURE

Gender Differences in Forensic Populations

Research has shown how the development of women's criminal behavior differs from the pathways into criminality men typically follow. Female offenders usually have histories of sexual and physical abuse, financial difficulties, and substance use (McKeown, 2010).

Additionally, research has shown that the developmental course of antisocial behavior differs significantly in males versus females (Rettinger & Andrews, 2010). One research study looking at gender differences among forensic psychiatric patients found that women had more multifaceted histories of victimization, had higher rates of borderline personality disorder diagnoses, were more likely to commit homicide and arson, and were less likely to commit sexual offenses when compared to men (de Vogel et al., 2016). Another study confirmed that female patients have higher incidents of victimization than male patients and discovered that female forensic patients with a history of victimization were also more likely to be diagnosed with borderline personality disorder than women without a history of victimization. Furthermore, men with a history of physical abuse were more likely to be diagnosed with antisocial personality disorder than men without a history of victimization (Bohle & de Vogel, 2017).

Another research study conducted by Grimbos et al. (2016) used the Historical, Clinical, Risk Management-20 (HCR-20), a structured professional judgment (SPJ) tool regarding violence risk assessment (Douglas et al., 2013) and found similar gender differences within forensic psychiatric patients. Similar to de Vogel et al. (2016), the study found a higher prevalence of borderline personality disorder among female than male patients. In contrast, antisocial personality disorder, substance use disorders, and a history of problems with violence were more often found among male patients (Grimbos et al., 2016). Although it seems

contradictory given the previously stated findings, research shows that women are more likely to engage in inpatient aggression than male patients (de Vogel et al., 2016; de Vogel & Nicholls, 2016). This finding is also interesting because in the same study, men within the sample primarily exhibited externalizing problems (i.e., aggression), while many female patients showed severe internalizing problems (i.e., self-harm and depression; de Vogel et al., 2016).

Overall, there has been an increase in prevalence rates among female offenders (typically with a later onset of criminality). In summary, there are distinctive differences between men and women within the forensic field. Research shows women have a higher rate of victimization, comorbidity of psychiatric diagnoses, internalizing or dissociative disorders (and fewer externalizing disorders), and instances of self-harm (de Vogel & Nicholls, 2016). Influences on female offending typically include a history of trauma, comorbid mental health diagnoses, and relationship problems (de Vogel et al., 2022).

Notably, gender-neutral assessments have identified common risk factors that appear to impact both men and women, such as criminal history, antisocial attitudes, financial and employment challenges, education, problematic substance use, and antisocial peers (Van Voorhis et al., 2010). However, these factors often exist in a different context dependent on gender and may manifest differently, highlighted in gender-responsive assessment. For example, issues with substance use are relevant risk factors for both genders; however, the onset of use, dependence, and relapse appear gender-sensitive (de Vogel et al., 2022). Very little research has focused on gender differences in forensic populations outside of prison or inpatient settings, which this current study aims to do.

Unique Considerations of Gender in Forensic Assessment

Despite these documented gender differences within forensic populations, a majority of the literature about criminal behavior has focused on males. This discrepancy has resulted in a lack of validated tools that are appropriate to use when assessing females in a forensic capacity. Forensic psychologists must use assessment instruments that are reliable and valid for the population they are assessing. According to the Code of Ethics, assessments must consider individual differences, including gender (APA, 2013). Knowledge transfer from the male offender literature to females is not applicable. Furthermore, it is unethical for forensic psychologists to utilize assessment instruments designated for a population of male offenders with female offenders (APA, 2013).

For example, there is no database for determining the sexual recidivism of female sexual offenders. Conversely, the Static-99 is an actuarial tool designed to assess risk of recidivism among male sex offenders. Actuarial tools use fixed algorithms based on nomothetic data from groups of recidivistic and nonrecidivistic offenders to estimate the probability or likelihood that a person will reoffend in the future (Hart et al., 2007). As previously stated, the number of female offenders is rising, making the need for appropriate instruments increasingly important. Some forensic psychologists use assessment tools that have been normed with a male offender population to determine the level of risk and, in turn, recidivism rates for female offenders. This generalization to women is inappropriate as it cannot guarantee an accurate risk estimate due to the gender differences previously discussed. Risk score estimates produced by the algorithms are based on group-based norms; therefore, they do not consider gender-specific risk factors. For example, within sexual offending, women have much lower recidivism rates, so using

male-based actuarial assessment instruments (such as the Static-99) would significantly overestimate the risk among these women (Cortoni & Gannon, 2013).

Actuarial risk assessments are mainly focused on examining static risk factors associated with the risk of future violence. Static risk factors are generally unchanging and challenging to address in treatment. Structured professional judgment (SPJ) tools typically rely on dynamic risk factors when assessing recidivism, allowing more individualized nuances to be detected. Dynamic risk factors are generally conceptualized as an offender's psychological or behavioral features that are potentially changeable and may lead to reduced recidivism when addressed in treatment (Cortoni, 2018). SPJ tools combine static and dynamic risk factors to inform interventions and guide practitioners. There is consensus that structured behavioral approaches are more predictively valid than unstructured ones (Poldrack et al., 2018). Furthermore, SPJ measures help assess individuals with unique risk profiles, such as considering gender-specific risk factors (Grimbos et al., 2016). For example, an SPJ tool may be useful if an actuarial assessment tool is unavailable for a specific population. Using structured tools when assessing for risk reduces the likelihood that the evaluator's estimate of future risk or recidivism is influenced by bias or stereotypes. However, it does not eliminate this probability.

Research shows dynamic factors do not regularly outperform static variables in more accurately assessing risk for future violence; however, they are critical in treatment planning and intervention (McKeown, 2010). Static and dynamic risk factors for female offenders are not as well established as for male offenders, but recent research efforts have begun to provide some direction (Gannon & Cortoni, 2010). Some structured assessment tools for use with female offenders include the Women's Risk Needs Assessment, The Security Reclassification Scale for Women, The Early Assessment Risk List for Girls, and The Female Additional Manual (which

can accompany the HCR-20). Although a step in the right direction, there remains a deficiency of tools used to evaluate women. Additionally, most risk assessment tools (SPJ and actuarial tools) were developed and validated with male offenders. Since many commonly used risk assessment tools are not yet empirically sound for female populations, forensic psychologists must remain wary when interpreting the results (de Vogel & Nicholls, 2016), mainly because research has shown that clinicians are limited in their ability to assess psychiatric patients' risk of future violence (Skeem et al., 2005).

Gender Issues in Forensic Assessment

A lack of assessment tools appropriate for women may result in forensic psychologists relying on unstructured clinical judgment, thus reinforcing stereotypes that are likely inaccurate. For example, gender bias may result in forensic psychologists ignoring specific problems of females, as females are more likely than males to internalize problems, resulting in misdiagnosis (de Vogel & Nicholls, 2016). Recognizing the gender differences between males and females in terms of both static and dynamic risk factors (McKeown, 2010), as well as differences in manifestations of violence (Grimbos et al., 2016), have compelled forensic psychologists to take a gender-responsive (i.e., gender-sensitive) approach to risk assessment during the forensic evaluation process (Van Voorhis et al., 2010). The most successful method for treating and rehabilitating female offenders recognizes that "gender makes a difference" (McKeown, 2010, p. 423). However, most risk assessment instruments fail to address common female issues such as self-esteem, self-efficacy, victimization, and trauma exposure (Rettinger & Andrews, 2010).

Within forensic risk assessment, women and men present with different clinical risk profiles, suggesting a greater need for gender-sensitive approaches to risk assessment (Grimbos et al., 2016). Risk assessments of women must focus on risk factors relevant to female offenders,

including prior criminal history, prior child abuse (nonsexual), antisocial attitudes, women's unique health, substance abuse, problematic relationships (antisocial associates), emotional dyscontrol, intimacy deficits, and victimization (Cortoni, 2018; Cortoni & Hanson, 2005; Gannon & Cortoni, 2010; McKeown, 2010). An effective risk assessment process for females would incorporate factors most notably associated with female misconduct (e.g., trauma, substance misuse, mental illness, and family support) into existing gender-neutral assessment instruments, or develop entirely new measures specifically for women, to provide a more valid clinical picture during evaluation (Davidson et al., 2015).

Furthermore, gender-sensitive assessments would result in evaluators being less likely to rely on unstructured clinical judgment leading to more ethical and nonbiased reports. Although not a risk assessment tool, the PAI provides additional clinical information about individuals presenting for forensic evaluation and is typically utilized alongside other SPJ and actuarial instruments. Understanding the nuances in personality and psychopathology between men and women helps forensic psychologists conduct ethical and nonbiased forensic evaluations, contributing to valid clinical profiles. Less biased assessment will lend itself to more appropriate and practical recommendations for treatment.

The Personality Assessment Inventory (PAI)

The PAI is a measure of personality widely accepted in various forensic and psycholegal applications (Morey, 2007). The PAI offers the opportunity to assess psychopathology and overall personality functioning. Dr. Leslie Morey developed this tool, and its initial introduction was cited as an "exciting new personality test" (Morey, 2007). There are several general legal issues for which forensic psychologists may use personality tests (like the PAI) as part of the

evaluation process, including competency, criminal responsibility, risk assessment, sentencing mitigation, treatment amenability, and validity of responses (Edens et al., 2001).

The PAI is a psychological test that offers information about diagnostic considerations, psychopathology, and treatment planning (Morey, 2007). The PAI can be essential in a comprehensive assessment for treatment planning and decision-making within forensic populations. A growing body of literature supports the PAI utility among women and men for examining distorted responses, psychiatric diagnosis, character pathology, substance abuse, risk assessment, and treatment considerations (Morey & Quigley, 2002).

This self-report instrument has 344 items, organized into 22 scales (four validity scales, 11 clinical scales, five treatment scales, and two interpersonal scales). Responses are indicated on a four-point Likert scale ranging from “false” to “very true.” The PAI was developed using clinical diagnoses and associated symptomatology as a guide. The PAI is written at a fourth grade reading level and can be administered to adults aged 18 years or older.

Developed using a construct validation framework, the PAI emphasizes content such as diagnostic criteria and personality traits. This framework emphasizes convergent and discriminant validity of specific scales within the PAI structure (Morey, 2007), meaning each item addresses a particular scale or diagnostic criterion, and items do not overlap. Furthermore, to strengthen content validity, the Likert scale guarantees sufficient depth by offering information about the intensity and severity of symptomatology on each construct.

The PAI has three different normative samples for comparison. The community sample was a U.S. census-matched standardization sample recruited from various sites and states. Clinicians enlisted the clinical sample from a variety of treatment settings. Lastly, the university sample was collected and drawn from seven universities for research purposes. The U.S.

census-matched standardization sample is the normative sample used in the present study.

Raw scores on the PAI are converted to T-scores using the computer software developed by test creators. Elevations in T-scores are interpreted by comparing the gathered T-scores to those of the normative community (average score of 50T) or clinical sample (average score of 70T). When comparing to the community sample, the PAI scale and subscale raw scores are transformed to T-scores (with a mean of 50 and a standard deviation of 10) relative to the standardization sample of 1,000 community-dwelling adults. Unlike other clinical measures, the PAI does not calculate T-scores differently for men and women; combined norms are used for both genders. A score of 60T depicts an individual at the 84th percentile regarding symptom experience and problems of a particular type. A score of greater or equal to 70T represents a level of reported problems rarely encountered in the general population, and a score at about the 96th percentile for most scales indicates a significant clinical concern within that domain (Morey, 2007).

Within the interpretive report generated by the computer software, the scoring profile offers a “skyline” of T-scores as a reference point. The skyline makes it easy to see which scores are two standard deviations above the mean for a sample of 1,246 clinical patients (70T), offering an immediate understanding of the person being evaluated. This bar graph “skyline” representation provides a picture of how obtained T-scores compare to the clinical sample provided by test developers, which assists in making clinical judgments regarding the severity level within each scale for the person being assessed (Morey, 2007). Approximately 98% of clinical individuals will score below this skyline. For example, it is apparent in the “skyline” depiction to see where the individual’s T-scores fall compared to the clinical population versus just looking at scores alone.

Description of PAI Scales

Refer to Appendix B for a table containing brief descriptions of the PAI scales. The validity scales include scales of Inconsistency (INC), Infrequency (INF), Negative Impression (NIM), and Positive Impression (PIM). According to the test developers (Morey, 2007), the INC scale reflects carelessness or confusion in responding as it considers how the respondent answered items with similar content. The INF scale measures carelessness in responding and identifies respondents who answered atypically (i.e., confusion, reading difficulties, carelessness, random responding). The NIM scale is not necessarily a malingering scale. This scale encompasses items that present an exaggerated unfavorable impression OR represents unusual or atypical symptoms that are not likely. The PIM scale contains items that present a favorable impression OR the denial of relatively minor faults.

The clinical scales include scales of Somatic Complaints (SOM), Anxiety (ANX), Anxiety-Related Disorders (ARD), Depression (DEP), Mania (MAN), Paranoia (PAR), Schizophrenia (SCZ), Borderline Features (BOR), Antisocial Features (ANT), Alcohol Problems (ALC), and Drug Problems (DRG). The SOM scale contains items that reflect concerns about physical functioning and worries related to health. The ANX scale provides an overall indication of anxiety within the clinical picture. The ARD scale measures the range of behavioral expressions of anxiety related to specific disorders (i.e., fears, phobias, obsessive compulsive thoughts, and bothersome experiences linked to some traumatic event). The DEP scale reflects clinical features common to depression (i.e., pessimism, negative expectations, subjective feelings of unhappiness, and physical signs related to depression, such as low energy and appetite changes). The MAN scale assesses the extent of the clinical picture as it relates to mania and hypomania (i.e., elevated mood, grandiosity, heightened activity levels, irritability). The

PAR scale is a measure of interpersonal mistrust and hostility and emphasizes the symptoms and chronic features of paranoia. The SCZ scale provides an assessment of numerous features of schizophrenia. The BOR scale considers several features related to severe personality disorder, precisely characteristics of borderline personality disorder, although some are common to other personality disorders. The ANT scale appraises the features related to antisocial personality and psychopathy constructs. The ALC scale measures behaviors and the effects of alcohol use, abuse, and dependence. The DRG scale measures behaviors and the effects related to drug use, abuse, and dependence. Of note, both the ALC and DRG scales can be elevated in respondents who have a history of alcohol or drug use yet are not currently using substances. Furthermore, items within the ALC and DRG scales are particularly vulnerable to instances of denial due to the self-report and directive nature of the items. The test developers suggest that raters be cautious in interpreting ALC and DRG scores when the respondent has elevated PIM scores.

The treatment consideration scales include scales of Aggression (AGG), Suicidal Ideation (SUI), Stress (STR), Nonsupport (NON), and Treatment Rejection (RXR). Of note, treatment consideration scales do not correlate directly with a specific diagnostic category, as is the case with the clinical scales. These scales touch on fundamental affects and behaviors involved across different groupings. The AGG scale evaluates attitudes and behaviors significant to aggression, anger, and hostility. The SUI scale assesses thoughts and ideas related to death and suicide. The STR scale evaluates the presence of life stressors that the client presently or has previously encountered, such as financial difficulties or problems with employment. The NON scale measures the perceived lack of social support, including the availability and quality of the respondent's social relationships. Items within the RXR scale reveal a disinclination to engage in

treatment actively, a lack of acknowledgment of personal problems, and a hesitancy to take accountability for problems in one's life.

The interpersonal scales include scales of Dominance (DOM) and Warmth (WRM). The DOM scale measures the range in which a person is controlling, submissive, or autonomous in interpersonal relationships. The WRM scale measures the range in which a person is empathic and engaging or withdrawing, rejecting, and mistrustful in interpersonal relationships.

Configural Profile Interpretation

Along with interpreting scale scores independently, interpretation is offered through profile configuration. This configural approach is founded on the principle that examining the pattern of information presented within multidimensional inventories is more beneficial than scrutinizing the parts (Morey, 2007). The test developers offer 10 empirically determined PAI configurations based on clusters derived from statistical procedures outlined in the professional manual (Cluster 1 through Cluster 10). For example, the Cluster 1 modal profile is notable for the absence of prominent elevations, defined as scale mean averages no more than one standard deviation above the mean, suggesting this individual is overall functioning relatively well. The Cluster 8 modal profile is notable for marked elevation on the Somatic Complaints (SOM) scale, suggesting this individual is concerned about their physical functioning. See Appendix C for a synopsis of the 10 clusters.

Influence of Demographic Variables

Throughout the development of the PAI, test developers were conscious of the necessity to minimize the risk of bias due to age, gender, race, and ethnicity. The test developers assembled a bias review panel encompassing citizens and professionals from diverse backgrounds and occupations to eliminate possible bias. The panel reviewed PAI items and

identified any that were deemed offensive or misleading, and these items were then eliminated or revised. The test developers acknowledged that using a bias review panel does not necessarily assure that the final items included will be inherently bias free on a practical level (Morey, 2007).

The procedures to reduce bias outlined above were used to guarantee that PAI items were equally useful indicators of psychopathology across different demographic groups. However, since certain disorders are associated with specific demographic groups (i.e., an antisocial personality disorder is far more common in men than women and younger than older patients), test developers did not endeavor to accurately equate mean values across demographic variables. The test developers suggest that “applying separate normative transformations for different demographic groups and using a similar cutoff score would lead to prevalence rates that would be similar in these different groups, yet the similarity would not be consistent with available epidemiologic data” (Morey, 2007, p. 89). Given this information, PAI T-scores are calculated considering the normative group used for comparison instead of specific demographic groups (i.e., gender). Even though gender-specific information is provided in the test manual for the U.S. census-matched standardization sample, there are no gender-based norms and scores are interpreted using combined norms based on the comparison sample. The present study examined gender differences, so the normative sample used in the study was the different means according to gender rather than the U.S. census-matched standardization sample as a whole.

According to research conducted by the test developers, the influence of demographic variables, such as gender, appears to be negligible on the PAI scale and subscale scores. Means and standard deviations (SDs) for the PAI scale and subscale T scores as a function of gender were considered. The only differences in the standardization sample between men’s and

women's scores were in the Antisocial Features (ANT) and Alcohol Problems (ALC) subscales. This difference is consistent with a higher rate of antisocial personality disorder and alcoholism among men than women (Grimbos et al., 2016; Morey, 2007).

Validity and Reliability as Reported in the Literature

The PAI has good internal consistency and test-retest reliability, as presented in the *Personality Assessment Inventory (PAI): Professional Manual, 2nd Edition* (Morey, 2007). Studies have estimated that the average internal consistency values for the scales and subscales range from .70 to .80. Regarding test-retest reliability, most scales illustrated reliabilities in the .70 to .80 range. One study raised concerns regarding construct validity as the PAI factor structure was not replicated for the standardized clinical sample. Furthermore, the confirmatory factor analysis using the normative correlational data provided by test developers revealed poor fit indices with the clinical sample in the study (Boyle & Lennon, 1994). Additionally, a study conducted by Salvin-Mulford et al. (2012) did not support construct validity for either the Mania (MAN) scale or the Anxiety (ANX) scale. However, the same study supported the PAI scales' overall criterion validity (Salvin-Mulford et al., 2012). Additionally, one study examining differences between the PAI and the MMPI-2 found the PAI to be preferable regarding internal consistency and test-retest reliability (Wise et al., 2010). Further, all 22 scales are nonoverlapping, supporting high discriminant validity.

Strengths

The PAI is easy to administer and score. A computer program is used to record responses and generate interpretive hypotheses. The fourth grade level readability is a strength, which is a crucial consideration in forensic settings where education levels may be lower than found in the general population (Morey & Quigley, 2002). The standardization groups are also a strength of

the PAI as the clinician can use the standardization sample that most closely fits the person being assessed (Wise et al., 2010). Furthermore, having multiple normative groups is useful when assessing broad and diverse populations, such as those presenting for forensic evaluation.

Limitations

Psychologists should take care when interpreting the PAI, and multiple data sources should be considered (i.e., case histories, clinical interviews with the patient, collateral interviews, consultation with treatment team members, historical data, and reviewing available records) before drawing conclusions or making recommendations. The interpretative hypotheses generated by the PAI should be used in conjunction with other assessment tools and structured clinical judgment when making diagnostic and treatment decisions. Furthermore, the PAI is a self-report measure, and social desirability may impact the results. This consideration is of particular concern within a forensic sample where individuals present for legal matters and may attempt to fake bad or fake good to influence verdicts or sentencing. However, validity scales are built-in, and concerns around validity are noted in the overall interpretative hypotheses. Furthermore, the PAI incorporates several response style indices (carelessness, random responding, and minimization or exaggeration of symptoms) to address response bias and social desirability (Davidson et al., 2015).

Summary

Research regarding gender differences among forensic populations continues to grow within the field; however, there remain gaps in the knowledge (de Vogel & Nicholls, 2016), particularly in a sample such as the one used in the present study. Men and women have different mental health needs; thus, they must be assessed accordingly to facilitate nonbiased and ethical clinical formulations (de Vogel & Nicholls, 2016). Overall, there has been a growing interest in

gender-sensitive assessment tools because most assessment instruments are normed with male populations and fail to incorporate the specific factors relevant to females, resulting in clinical profiles of questionable validity. While not a gender-specific instrument, the PAI assesses psychopathological syndrome and provides information pertinent to screening for psychopathology (Morey, 2007). Overall, the PAI appears to have strong psychometric properties and is a valid and reliable measure of personality. Very few studies discuss the implications of gender differences on the PAI. Specifically, no studies have explored the PAI's clinical application when explicitly used with a forensic sample of this kind. Research shows good utility of the PAI with an offender population (Morey, 2007). Nevertheless, limited studies have used a forensic sample in which individuals have presented for evaluation for a myriad of reasons, particularly those not already institutionalized or jailed. The current study addresses that gap in the literature.

CHAPTER III: METHOD

Design

This study employed a quantitative, descriptive design using existing clinical data from Vermont Forensic Assessment (VTFA), a private group forensic psychology practice located in Shelburne, VT. Based on the archival and deidentified nature of the data, as well as permission from VTFA, these data were exempt from IRB review.

Participants

Demographic data examined in this study included age, gender, race/ethnicity, marital status, years of education, and occupation. Age at the time of test administration was determined by subtracting the individual's date of birth from the test date (month, day, and year).

Archival Sample

Individuals in the archival sample were referred for assessment at a private group practice consisting of licensed forensic psychologists in Vermont. The referral sources include the Department of Children and Families, the Department of Corrections, individuals, companies, lawyers, and courts. Referral requests included parenting capacity evaluations, fitness for duty evaluations, risk assessments, psychological evaluations, and psychosexual evaluations. Assessments were conducted in an office setting and in jails and prisons.

Refer to Table 1 for a breakdown of the sample's demographic characteristics by gender (female and male). Additional demographic information is available in the Results section. The archival sample was acquired from existing clinical data from VTFA and consisted of 336 adults. Females comprised 29.5% of the sample, and 70.5% were males. The sample was predominantly white (90.9% Caucasian).

Additionally, a majority of the sample was single (54.5%). Unfortunately, information related to gender was analyzed using a binary system. In the present study, gender differences will be explored according to comparisons between the male sex and female sex (defined as men and women within this study for ease of understanding). The sample consists of a diverse array of occupational statuses. Of the 65.5% of individuals that listed occupation, 8% were not employed, 5.4% were incarcerated, and 4.8% indicated having a disability. Occupations included teacher, student, truck driver, manager, nun, business owner, army, police officer, correctional officer, electrician, engineer, laborer, retiree, salesperson, and others.

Normative Sample

The normative sample used as a comparison in this study was the standardized community sample presented in the *Personality Assessment Inventory (PAI): Professional Manual, 2nd Edition* (Morey, 2007). This community sample is a subset of a group of 1,462 community dwelling adults (i.e., not residing in an inpatient facility or hospital). Ultimately, 1,000 individuals were chosen based on cross-stratification for gender, race, and age variables who had left no more than 33 items blank. These individuals were from rural and urban settings across 12 states, and all were 18 years of age and older (42.7% of the sample was between ages 30–49). Within the normative sample, 41.1% were white males, and 44% were white females. Twenty-one and a half percent of the sample were never married, 52.2% were married, 9.5% were divorced, and the remaining were remarried, separated, widowed, or other.

Measures

The PAI is administered to most individuals referred to VTFA for an evaluation as part of the assessment process. The PAI is typically completed via paper and manually entered through the PAI computer software and interpretation services. PAI responses are manually entered into

the computer twice and compared to ensure accuracy via the software. All records were stored securely in the VTFA office.

Procedure

The sample for this study included 378 individuals who were administered the PAI at VTFA between December 2014 and October 2019. The T-scores for all individuals who completed the PAI in its entirety from 2014 to 2019 were included in the analysis. In line with the PAI manual, any profiles missing more than 33 items were excluded from the analyses. Upon visual examination of the data set, several cases did not have the gender listed. Additionally, some cases were either missing the birth date or test date. Cases were excluded from the study if age or gender could not be discerned. Also, some cases appeared to be duplicate entries (i.e., contained the same test date, demographic information, and response set), and those were excluded as well. This process resulted in 336 unique PAI administrations within the period described above. The resulting sample included 99 females and 237 males.

Analysis

The deidentified data set was downloaded to SPSS, which was used to conduct all analyses. Various statistical procedures were utilized to explore how profiles did or did not differ for men and women within this forensic sample, as well as how they compared to the overall normative sample.

Inferential Tests of VTFA Sample

A Chi-Square Test of Independence was used to determine if the distribution of marital status and ethnicity was consistent for females and males. Independent *t*-tests were conducted to determine if the average age and number of years of education were the same for females and males.

Research Question One and Two

1. How do the overall mean scores on any PAI scales for females within this sample compare to the normative community sample scores among females provided by the test developer?
2. How do the overall mean scores on any PAI scales for males within this sample compare to the normative community sample scores among males provided by the test developer?

One-sample *t*-tests were used to compare the archival mean to the normative mean for females only for each of the 22 scales to assess research question one. One-sample *t*-tests were also used to compare the archival mean to the normative mean for males only for each of the 22 scales.

Research Question Three

3. Are there significant differences between males and females within this archival forensic sample on any PAI validity, clinical, treatment, or interpersonal scales?

To determine if there were significant differences between females and males within the VTFA sample on any of the PAI validity, clinical, treatment, or interpersonal scales, a multivariate analysis of variance (MANOVA) was conducted using the mean T-scores per scale for each gender. Based on the sample size and the number of analyses conducted, a more conservative *p*-value of 0.01 was used. The MANOVA was employed to determine if there was any indication that at least one subscale was different for females and males.

CHAPTER IV: RESULTS

Demographics

A sample of 336 clinical records was available based on the criteria that all scales of the PAI were completed and the gender of the individual was recorded. The sample consisted of 99 females (29.5%) and 237 males (70.5%). Age was significantly different across gender ($t(334) = -2.99, p = .003$). The average age of female respondents was 33.5 years ($SD = 10.9$), and male respondents were 38.2 years ($SD = 14.1$). The average years of education for females were 11.7 ($SD = 4.4$) and males were 12.4 ($SD = 2.9$). Differences in education were not significant. The sample predominantly consisted of white females and males (87.5% and 92%, respectively). The distribution of race/ethnicity among males and females was assumed to be the same. Table 1 summarizes the distribution of marital status and ethnicity across gender. Chi-square tests of independence indicated that the distributions do not differ across genders, suggesting that the ethnic and marital composition of the gender samples were equivalent.

Screening for Outliers

According to Tabachnick and Fidell (2019), outliers can be easily identified by converting raw scores into standardized z -scores to determine those scores that exceed a z -score of 3.29 ($p < .001$, two-tailed). They suggest comparing analyses with outliers removed to analyses with outliers included to determine if analyses show any difference in significant findings. In the present analyses, 14 cases were identified as outliers. Comparison of analyses with and without outliers showed no difference in the levels of significance. Therefore, analyses were conducted with all cases included, which maximized the sample size.

Descriptive Statistics

Table 2 reports the descriptive statistics and normative means for each scale of the PAI across gender. Normative means for the PAI were reported in the *Personality Assessment Inventory (PAI): Professional Manual, 2nd Edition* (Morey, 2007). One-sample *t*-tests were conducted to compare sample means for each scale to the corresponding normative mean and the results are depicted in Table 2. A more conservative *p*-value of .01 was used due to the number of one-sample *t*-tests conducted. Lastly, the minimum value to determine a small effect size was .20, a minimum value of .50 for a medium effect size, and a value of at least .80 to signify a large effect size.

Comparison of Women in the Archival vs. Women in the Normative Sample on the PAI

Validity Scales

Based on the one-sample *t*-tests, females within the archival sample had significantly different mean scores than the normative sample on the Infrequency (INF) and Negative Impression (NIM) scales. On the INF scale, females within the archival sample had significantly higher mean INF scores ($M = 53.86, SD = 9.17$) and NIM scores ($M = 56.40, SD = 13.16$) than the normative sample; $t(98) = 5.32, p < .001$ and $t(98) = 5.35, p < .001$, respectively. Medium effect sizes were also found for the INF and NIM scales (.53 and .54, respectively). No significant differences were found between the archival sample and the normative means within the Positive Impression (PIM) scale.

Clinical Scales

Many clinical scales were significantly different between females in the archival sample and females in the normative sample. Results of the one-sample *t*-tests indicate that females in the archival sample had significantly higher mean scores when compared to the normative mean

on the following scales: Somatic Complaints (SOM) ($M = 57.16$, $SD = 14.96$); $t[98] = 4.64$, $p < .001$), Anxiety (ANX) ($M = 59.62$, $SD = 15.57$); $t[98] = 5.67$, $p < .001$), Anxiety-Related Disorders (ARD) ($M = 61$, $SD = 15.40$); $t[98] = 6.48$, $p < .001$), Depression (DEP) ($M = 61.57$, $SD = 16.67$); $t[98] = 6.79$, $p < .001$), Paranoia (PAR) ($M = 57.43$, $SD = 14.70$); $t[98] = 5.74$, $p < .001$), Schizophrenia (SCZ) ($M = 53.02$, $SD = 13.62$); $t[98] = 2.77$, $p = .007$), Borderline Features (BOR) ($M = 58.46$, $SD = 15.13$); $t[98] = 5.83$, $p < .001$), Antisocial Features (ANT) ($M = 50.04$, $SD = 11.70$); $t[98] = 2.66$, $p = .009$), and Drug Problems (DRG) ($M = 56.21$, $SD = 18.59$); $t[98] = 3.80$, $p < .001$). Females in the archival sample had significantly lower mean scores than the normative sample on the Mania (MAN) scale ($M = 45.72$, $SD = 9.94$; $t[99] = -2.84$, $p = .006$). Small effect sizes were found on the following scales: SOM, MAN, SCZ, ANT, and DRG. Medium effect sizes were found on the following scales: ANX, ARD, DEP, PAR, and BOR. No significant differences were found between females and the normative sample on the Alcohol Problems (ALC) scale, and a trivial effect size was found. See Table 2 for statistical results for each scale.

Treatment Consideration Scales

Females in the archival sample had significantly higher scores than the normative sample on the following treatment scales: Suicidal Ideation (SUI), Stress (STR), and Nonsupport (NON). There were no significant differences between the archival and normative samples within the Aggression (AGG) scale scores. Females had significantly lower scores than the normative sample on the Treatment Rejection (RXR) scale ($M = 46.80$, $SD = 11.53$, $t[98] = -2.67$, $p = .009$; $M = 44.46$, $SD = 12.19$; $t[236] = -7.19$, $p < .001$). A medium effect size difference was found on the STR scale. Small effect size differences were found on the SUI, NON, and RXR scales.

Interpersonal Scales

There were no significant differences within the interpersonal scales (Dominance [DOM] and Warmth [WRM]) for females compared to the normative sample. Effect sizes less than .20 were observed on the DOM scale for women, and a small effect size was found on the WRM scale.

Comparison of Men in the Archival vs. Men in the Normative Sample on the PAI

Validity Scales

Based on one-sample *t*-tests, males within the archival sample had significantly higher mean scores than the normative sample on the INF scale ($M = 53.97$, $SD = 9.30$; $t[236] = 4.71$, $p < .001$) and NIM scale ($M = 56.43$, $SD = 13.96$; $t[236] = 6.89$, $p < .001$). Additionally, results show males within the archival sample had higher scores on the Inconsistency (INC) scale as compared to the males within the normative sample ($M = 52.86$, $SD = 9.07$; $t[236] = 3.80$, $p < .001$). Small effect sizes were found for the INF and NIM scales. Trivial effect sizes were observed on the INC and PIM scales (less than .20 on Cohen's *d*). No significant differences were found between the archival and normative samples on the Positive Impression (PIM) scale.

Clinical Scales

Many clinical scales were significantly different between males in the archival sample and males in the normative sample. Males in the archival sample had significantly higher mean scores when compared to males in the normative sample on the following clinical scales: SOM ($M = 55.53$, $SD = 12.26$; $t[236] = 6.95$, $p < .001$), ANX ($M = 56.54$, $SD = 14.46$; $t[236] = 7.76$, $p < .001$), ARD ($M = 57.09$, $SD = 14.84$; $t[236] = 8.47$, $p < .001$), DEP ($M = 60.63$; $SD = 15.39$; $t[236] = 10.92$, $p < .001$), PAR ($M = 57.59$, $SD = 13.96$; $t[236] = 7.13$, $p < .001$), SCZ ($M = 54.80$, $SD = 13.56$; $t[236] = 4.50$, $p < .001$), BOR ($M = 58.01$, $SD = 14.05$; $t[236] = 8.23$,

$p < .001$), ANT ($M = 56.26$, $SD = 11.68$]; $t[236] = 3.80$, $p < .001$), and DRG ($M = 58.35$, $SD = 18.17$]; $t[236] = 5.87$, $p < .001$). No significant differences between males and the normative sample on the MAN or ALC scales were found. Small effect sizes were found on the SOM, PAR, ANT, ALC, and DRG scales, while medium effect sizes were found on the ANX, ARD, DEP, MAN, and BOR scales. Trivial effect sizes were found on the SCZ scale. See Table 2 for statistical results for each scale.

Treatment Consideration Scales

Males in the archival sample had significantly higher scores than the normative sample on the following treatment scales: Suicidal Ideation (SUI), Stress (STR), and Nonsupport (NON). There were no significant differences on the Aggression (AGG) scale score, but a small effect size was found. Males had significantly lower scores than the normative sample on the Treatment Rejection (RXR) scale ($M = 46.80$, $SD = 11.53$; $t[98] = -2.67$, $p = .009$; $M = 44.46$, $SD = 12.19$; $t[236] = -7.19$, $p < .001$). A medium effect size difference was found for the STR scale. Small effect size differences were found on the AGG, SUI, NON, and RXR scales.

Interpersonal Scales

The only significant result among the interpersonal scales for males only was that the archival sample scored significantly lower on the DOM scale than the normative sample ($M = 47.97$, $SD = 10.37$; $t[236] = -5.56$, $p < .001$). A trivial effect size was found for the WRM scale among men in each sample and a small effect size was observed on the DOM scale.

Comparison of Men vs. Women in the Archival Sample on the PAI

The primary purpose of the present study was to examine gender differences within a forensic sample using the Personality Assessment Inventory. A MANOVA was conducted to determine if any of the multiple PAI scales were significantly different across gender. The

MANOVA was used rather than multiple independent-sample *t*-tests to reduce the likelihood of Type I errors. The first step in MANOVA is to assess if a linear composite score, which weights the multiple PAI scores to create a single score, is significantly different across gender using Wilks' lambda. MANOVA results for the PAI scales indicated significant differences across gender [Wilks' lambda = 0.722, $F(22, 313) = 5.478, p < .001$] on the composite PAI dependent variable. The individual scales found to be significant across gender were noted in Table 3. These scales include Inconsistency (ICN), Mania (MAN), Antisocial Features (ANT), and Alcohol Problems (ALC). Males reported significantly higher scores than females on each of the scales ($[M = 52.86, SD = 9.07]$, $[M = 49.91, SD = 10.67]$, $[M = 56.26, SD = 11.68]$, $[M = 54.37, SD = 15.42]$, respectively). Of note, effect sizes among a majority of the 22 scales were trivial (all less than .01) except for the Inconsistency (INC), Anxiety-Related Disorders (ARD), Mania (MAN), Antisocial Features (ANT), Alcohol Use Problems (ALC), and Aggression (AGG) scales.

CHAPTER V: DISCUSSION

The present study investigated gender differences within a unique forensic sample using the PAI to provide information about how men and women present on measures of validity of responses, psychopathology, treatment considerations, and interpersonal style when presenting for forensic evaluation. To assess how the archival forensic sample differed from the PAI normative sample as posited in research questions one and two, statistical analyses compared mean scores between the archival and community samples according to gender. To examine gender differences as posited in research question three, mean scores were compared between men and women in the archival sample.

Overall, the mean differences found in this study add to the existing literature by describing information about a unique forensic sample and how the PAI may look alike or dissimilar based on gender. This study supported some of the previous research highlighting the differences observed between men and women within forensic populations while also describing how the archival sample differs from other types of previously studied populations according to gender. Overall, the present study's findings emphasize that gender makes a difference when conducting forensic evaluation, as men and women present differently on a measure assessing personality structure and psychopathology. The results underline the need for additional appropriately validated tools to improve the assessment of women in a forensic population.

The differences between men and women in the archival sample on the PAI scales might shed light on the existing literature regarding the contrast between internalizing and externalizing factors of psychopathology that has been observed in prior research with forensic populations. The main implication of the current study is that forensic psychologists must attend to gender differences when interpreting the results of the PAI to formulate more appropriate and clinically

relevant diagnoses. Moreover, PAI scale elevations not only reflect fundamental clinical constructs related to psychopathology, but the PAI also provides context for how individuals with certain scores might manage their problems, thus aiding in making treatment recommendations. Given that the PAI does not perform similarly for men and women in this kind of forensic sample (which will be explained in greater detail below), psychologists should use clinical judgment when evaluating an individual and must not rely solely on test results. These findings highlight the utilization of multiple data sources and awareness of gender differences as essential components of gender-responsive assessment.

Important to note is that these findings should be interpreted with caution due to the sample size. Although a more conservative *p*-value of .01 was used, statistical significance is not synonymous with clinical relevance. Additionally, some analyses yielded trivial effect sizes, with others finding small to medium effect sizes. Thus, the clinical implications should be considered judiciously. Regardless, the impact on clinical work and future research is relevant and valuable and will be discussed throughout this chapter. Of note, scales will hereafter be referred to as the construct of which they measure (i.e., the Depression scale, also known as DEP, will be referred to as a measure of depression) to facilitate ease of understanding and to communicate implications and recommendations for clinical practice better.

Comparisons Between the Archival and Normative Sample on the PAI

Research question one asked, “How do the overall mean scores on any PAI scales for women within the archival sample compare to the test developer’s normative community sample scores among women?” Research question number two asked, “How do the overall mean scores on any PAI scales for men within the archival sample compare to the test developer’s normative

community sample scores among men?” Many PAI scales were significantly different between the archival and the normative sample (see Table 1).

Although the archival sample scored significantly higher on many scales compared to the normative sample, according to the range of scores per scale provided by the test developers that reflect severity of symptoms, group means in the archival sample reflected no markedly severe scores. Thus, the archival sample appears to fit the Cluster 1 profile described in Chapter II. In other words, this broad forensic sample seems to look most similar to how individuals present in the community (i.e., relatively free of prominent symptoms) rather than individuals in a clinical setting. Yet, this interpretation cannot be stated with statistical certainty because a comparison to other kinds of samples (i.e., the clinical sample provided by the test developers) was not analyzed in the present study and is discussed further in the Future Considerations section. The similarity between the archival sample and the Cluster 1 profile is likely a reflection of the diversity of referral reasons in the archival sample (i.e., a mix of offenders and nonoffenders).

Nonetheless, upon visual examination of isolated scores, prominent score elevations were present among many individuals within the archival sample. Additionally, the archival sample scored significantly higher than the normative sample on several scales that reflect overall psychopathology. Meaning individuals in the archival sample appear to have higher levels of distress and psychopathology than the normative sample, regardless of appearing similar to the Cluster 1 profile as a whole. This is discussed further in the Levels of Psychopathology section.

The present study used the normative community sample as a basis for comparison because individuals in the archival sample were a mix of community-dwelling and incarcerated adults. Evaluators should always choose the most relevant sample for the evaluated individual, which the test developers assert in the professional manual. Specifically, the utility and validity

of the PAI are reliant on and directly correlated with the expertise and competency of the professional who interprets the PAI profile (Morey, 2007). In line with this statement, the forensic psychologists at VTFA carefully compare individuals with the appropriate normative sample, which is not always the community sample. One should decide which sample is most appropriate as it relates to the context of the examinee. For example, an individual in an inpatient setting would likely be compared to the clinical sample.

According to the test developers (Morey, 2007), the overall satisfaction in functioning and subjective contentment reported by individuals in the Cluster 1 profile type should be considered in the context of additional information, including historical information, records reviewed, or collateral interviews. Relying on multiple data sources is a vital component of forensic evaluation. Given the nature of the archival sample (i.e., individuals presenting for forensic assessment), relatively average and nonpathological scores would be important to consider within the context of other sources of information, including the reason for referral, rather than relying on the results of the PAI alone. Incorporating multiple data sources, such as mental health records, within a forensic assessment is an essential ethical consideration (APA, 2013).

Overall Approach to the PAI and Response Styles (Validity Scales)

Women and men within the archival sample seemed to present themselves in a more exaggerated unfavorable manner than the normative sample, as evidenced by scores on a scale related to a negative impression. This finding appears consistent with research provided by the test developers, which states negative impression scores are typically more elevated in clinical patients than in a general population of adults (Morey, 2007) due to increased levels of distress; however, the archival sample is different than the clinical sample described by test developers

(i.e., forensic setting versus an inpatient/outpatient treatment setting). Despite this difference, individuals with severe emotional problems are more likely to have higher scores on this scale, reflecting overall more significant impairment. Since individuals in the archival sample had higher scores on this scale, it follows that these individuals may be more impaired than individuals in a community setting. Furthermore, scores on measures related to a favorable impression response style were relatively similar between the samples. It seems counterintuitive since individuals in the archival sample were presenting for forensic evaluation and an attempt to portray themselves more favorably would be understandable given the possible legal ramifications. These findings suggest that individuals in the archival sample are not attempting to malingering per se; instead, they may present with a higher rate of distress and mental disorders than the normative sample resulting in a distorted sense of self and others. This finding is important for forensic evaluators to bear in mind, as social desirability may impact the results of the PAI. Validity of responses is particularly important in a forensic setting, as examinees may attempt to fake bad or fake good given the legal implications. Based on the findings stated above, forensic evaluators must use caution when trying to determine if scores related to impression management on the PAI are related to malingering or not.

Regarding response styles, women and men in the archival sample responded to questions in a more atypical style than the normative sample, and scores among men in the archival sample also suggested they responded in a way that reflected more carelessness or confusion than the normative sample. This finding indicates that while men and women in the archival sample responded somewhat unusually, men appeared more negligent when responding to items on the PAI than women. Thus, women presenting for forensic evaluation may be more careful in their responses than men, suggesting women may be more concerned with the outcome

of the assessment. This finding is explored further when discussing gender differences within the archival sample. Also, these differences may be better explained by educational attainment as a mediating factor. For example, higher levels of education in the normative sample than in the archival sample may have resulted in a better understanding of the PAI items. The impact of education on PAI scores would be beneficial to explore in future studies, as this was not examined in the present analyses. Nevertheless, considering the examinee's level of education is imperative prior to the onset of an evaluation to determine the appropriate measures to use for the assessment.

Levels of Psychopathology (Clinical Scales)

Both women and men in the archival sample had significantly higher scores than the normative sample on measures related to somatic complaints, anxiety, anxiety-related disorders, depression, paranoia, schizophrenia, borderline features, antisocial features, and drug problems. Score elevations in borderline and antisocial features likely reflect a greater severity of psychopathology in the archival sample. The archival sample endorsed more symptomology consistent with anxiety, depression, schizophrenia, and somatic concerns, suggesting the archival sample has higher rates of diagnoses related to these symptoms than the normative sample. These findings appear fitting and are expected given the forensic nature of the sample. Elevated paranoia scores in the archival sample may suggest some level of mistrust regarding the forensic assessment process among those who were mandated for an evaluation or referred through the court system. Forensic evaluators must consider what elements of the individual's presentation are leading to elevated scores in this regard, as it might impact how the person will interact with the court system. For example, this mistrust may hinder their ability to assist legal counsel.

Individuals in the archival sample appear more likely to experience issues related to drug use than the community sample, but problems related to alcohol use were similar between the samples. Cultural perception of alcohol use as more normative and common than illicit substances may have influenced this finding. As individuals in the archival sample appeared to present themselves in a more unfavorable light, they may be more forthcoming about their drug use. Of note, PAI items related to alcohol and drug use cannot discriminate between current and past use, so higher drug problem scores may reflect a more remarkable history of past, but not present, drug use in the archival sample than the normative sample. Also, this scale does not discriminate between legalized and illicit substance use and cannabis use is legal in the state of Vermont where this archival sample was collected. These findings regarding substance use highlight the importance of multiple evaluation methods to determine the precise reasoning behind elevated scores in these domains.

Women in the archival sample had lower scores on items related to mania than women in the normative sample, which may reflect a lower prevalence of symptoms of mania among women in the archival sample than the normative sample. Men in both samples appeared to respond similarly to items on this measure. The finding among men is consistent with research that average scores on the Mania scale within the representative clinical sample and normative community sample are typically the same, which is not seen within any other PAI scale (Morey, 2007). Further, elevations on the Mania scale tend to be rarer in the clinical sample than in any of the other clinical scales (Morey, 2007). The rationale for this observation is not explained by the test developers, although it may reflect the potential that items within the Mania scale do not accurately reflect manic symptoms and may be related to the pervasiveness of misdiagnosis

when it comes to bipolar disorder. Again, forensic psychologists must use all available sources of information when making a diagnosis to ensure accuracy.

Factors Related to Treatment (Treatment Consideration Scales)

Differences in mean group scores suggest that the archival sample endorsed more thoughts and ideas related to suicide, a greater presence of life stressors, and perceived themselves as lacking social support compared to the normative sample. These findings are important for forensic psychologists to consider when determining treatment recommendations, particularly the importance of bolstering an individual's support system. Additionally, forensic psychologists should pay greater attention to possible endorsements of critical items related to suicide risk when assessing an individual in this kind of forensic population. Therefore, these items should be checked immediately upon completion of the PAI so the examiner can intervene and provide necessary support if warranted.

Risk for violence is a common form of risk assessment in forensic psychology; therefore, scores related to aggression are essential to consider in this population. Both samples scored similarly on a measure of aggression, suggesting attitudes and behaviors significant to anger and hostility were not vastly different in the archival sample compared to the normative sample. This finding is somewhat surprising as the archival sample contains some individuals referred for an evaluation following a violent offense and may reflect that most individuals in this forensic sample were seen in an outpatient setting. Furthermore, this highlights the clinical disparities between this archival sample and the samples primarily studied in forensic research, typically comprised of strictly violent offenders.

Overall, the archival sample had lower treatment rejection scores than the normative sample. Per the test developers (Morey, 2007), higher scores on the treatment rejection scale

suggest a lower motivation for treatment. This interpretation would suggest that men and women in the archival sample had a higher motivation for treatment than the normative sample. The results are understandable as men and women in the archival sample appear to be experiencing distress at a higher rate than individuals in normative sample (as stated in previous findings) and may recognize the need for treatment. Another factor impacting these scores may be the examinee's potential court involvement. For example, these individuals may be more motivated to participate in treatment to appease the court or stay out of further trouble. Scores indicating lower motivation for treatment would not have been surprising because some individuals presenting for forensic evaluation may attempt to portray themselves in a more favorable light, given the legal ramifications of a risk assessment. Ultimately, no definitive statements can be made as some individuals in the archival sample self-referred for evaluation, and scores on a separate measure related to favorable impression management were not different when comparing the archival sample to the normative sample. Overall, considering motivation for treatment is an important component of assessment. Specifically, research has shown that female offenders who engage in aftercare community treatment following release from prison are less likely to reoffend (Grella & Rodriguez, 2011). Important to note is that motivation for treatment "does not always equal a good outcome" (Parker et al., 2020, p. 2050).

Factors Related to Interpersonal Functioning (Interpersonal Scales)

Women in both samples had similar scores on measures of dominance and warmth. Men in the archival sample scored lower on measures of dominance than the normative sample, although they scored similarly on measures of warmth. These findings are somewhat surprising as it would be expected that men and women in the forensic sample would score higher on measures of dominance and lower on measures of warmth. Prior research examining the PAI

among male criminal offender populations has shown scores on these measures to be higher and these factors appeared related to antisocial traits and aggression (Parker et al., 2020). This unexpected finding may indicate that men in the archival sample may be more self-conscious and less confident than men in the normative sample, reflecting shame related to their legal problems or perceived inability to function competently in society compared to others.

Further, interpersonal functioning is noted as relevant to treatment (Morey, 2007). One research study looking at PAI scales and interpersonal characteristics among male sex offenders found that dominance had a positive relationship with the treatment rejection scale (Parker et al., 2020), which seems consistent with the findings of the present study; men in the archival sample appeared motivated for treatment and scored lower on measures of dominance. Again, this finding highlights the importance of considering multiple sources of available information when determining risk and motivation for treatment within a forensic population rather than making conclusive statements about risk and recidivism based on PAI results alone.

Gender Differences within the Archival Sample on the PAI

The third research question asked whether there were differences between men and women within the archival sample. Overall, men scored significantly higher than women on measures related to inconsistent response style, mania, antisocial features, and alcohol problems (see Table 2). The results of this study somewhat represented gender differences that would be expected based on identified gender traits and research regarding internalizing and externalizing factors summarized previously (de Vogel & Nicholls, 2016). For example, men scored higher than women on scales involving externalizing symptoms (i.e., scores related to antisocial features and alcohol problems). This finding aligns with previous research, which shows that men are more likely than women to be diagnosed with antisocial personality disorder and have

substance use-related diagnoses (Grimbos et al., 2016). Women did not score higher on scales representing internalization of symptoms, such as anxiety, anxiety-related disorders, depression, and somatic complaints. This finding is inconsistent with prevalence data from the DSM-5-TR (American Psychiatric Association [APA], 2022), which states that women are more likely than men to receive diagnoses related to anxiety, depression, and somatic disorders. The current findings add to the research regarding the complex nature of the relationship between internalizing and externalizing psychopathology and gender. Men do not exhibit solely externalizing symptoms, and women do not display exclusively internalizing symptoms.

Another interesting finding in the present study that differs from previous research is that men had significantly higher scores on the PAI measure of mania than women. This finding is inconsistent with the DSM-5-TR (APA, 2022), which states that men and women have similar rates of bipolar I disorder. Further, some clinical samples suggest bipolar II disorder is more common in women and they are more likely to endorse symptoms related to hypomania (APA, 2022). These conflicting findings seem to reflect prior research findings that did not support construct validity for the PAI Mania scale (Salvin-Mulford et al., 2012). This finding may be explained by considering the subscales of the PAI Mania scale, which include Activity Level, Grandiosity, and Irritability. Since there are higher rates of antisocial personality disorder symptoms among men, which include criteria related to irritability, impulsivity, and arrogant self-appraisal, PAI mania subscale elevations in Grandiosity and Irritability may have been higher in men than women leading to overall higher elevations on mania. The present study did not analyze subscales, so this is merely conjecture. A forensic evaluator may still find it helpful to examine Mania subscales to further understand overall symptom presentation. Future research might examine the relationship of gender to subscale scores. On the other hand, this may suggest

that men in this sample experienced more symptoms related to bipolar disorder and may have been more likely to be diagnosed with bipolar disorder than women.

Of interest, borderline features were not different between men and women in this study, which is inconsistent with prior research citing the higher prevalence of borderline personality disorder among women than men in a forensic population (Bohle & de Vogel, 2017; de Vogel et al., 2016; Grimbos et al., 2016). These results may reflect that men and women in this forensic context present similarly regarding borderline traits. Conversely, it may suggest that the PAI is not sensitive at detecting borderline personality trait differences among genders. Ultimately, this discrepant finding emphasizes the need for further research in this regard. This finding also illuminates issues related to misdiagnosis. In response to the lack of validated assessment tools for women, examiners may rely on unstructured clinical judgment, leading examiners to attribute women's symptomatology to borderline personality rather than consider other possible explanations (Morey & Benson, 2016). The pervasiveness of misdiagnosis highlights the importance of using structured professional judgment and multiple assessment tools when evaluating an individual. An approach with these findings in mind will facilitate nonbiased forensic evaluations and valid clinical diagnoses, one of the overarching conclusions of the present study.

Lastly, men appeared more inconsistent in their responses to the PAI items than women. According to the test developers, higher scores on their measure of inconsistency may suggest that the reader was not attending carefully to items and may have been negligent when responding to questions (Morey, 2007). This finding raises the question of whether men who present for forensic evaluation are more likely than women to be careless or confused in their response style. The present study cannot answer this question with certainty; still, the question

possesses clinical relevance. As stated previously, women presenting for forensic evaluation may be more careful in their responses than men, suggesting women may be more concerned with the outcome of the assessment. Forensic evaluators may find it beneficial to be more explicit in their explanation of the purpose and procedure of the evaluation with male examinees. The evaluator should also remain available to the examinee throughout the administration to answer any questions about the wording of the PAI questions. Lawyers who refer their male clients for forensic evaluation may find it helpful to highlight the importance of the evaluation as it relates to legal implications to illicit more thorough attention to PAI items. Despite the importance of these findings among men in particular, these recommendations for evaluators should also apply to women.

Limitations

A major limitation of the present study is the lack of diversity in gender identity. As a result of the binary categorization of the archival data and the normative data, the labels male and female were used to define gender and were used synonymously with the biological sex categories of men and women. Nevertheless, biological sex is not the same as gender. In the present study, gender is discussed as though it exists solely along a binary, neglecting the full spectrum of gender identity in the real world. As a result, the results of this study are not indicative of scores that may present themselves among gender minority populations, such as transgender or nonbinary individuals. Thus, this study does not represent all possible gender differences present within a forensic population.

Another limitation of the present study is that the archival sample was from one geographic area, and was predominantly white, which suggests the current sample may not represent gender differences observed in a more racially or geographically diverse sample. By

contrast, the sample was quite diverse in occupation, which may reflect diversity in socioeconomic status. However, demographic information related to socioeconomic status was not available in the present study and limited the claims that can be made.

Lastly, the archival sample used in the present study consisted of individuals referred for forensic evaluation for various reasons, including violence risk assessment, psychosexual assessment, parenting capacity evaluation, fitness for duty evaluations, and general psychological evaluation. The reason for referral is both a novel aspect and a limitation of the present study. Making claims about a specific group of individuals is difficult since the diversity of referrals is vast. The present findings may be a function of the diverse nature of referrals rather than gender. Despite these limitations, the results of the present study offer a unique perspective of gender differences, as measured by the PAI, that may present themselves in a private group practice setting.

Future Considerations

In light of the limitations of the present study and the current research in this area, there are several avenues that future research might explore. The present study focused on gender differences observed among the 22 main scales of the PAI and did not look at specific subscales. A future study could focus on analyzing gender differences within the PAI subscales to identify the nuances of the reported gender differences observed in the present study. Additionally, the present study could be replicated using the different samples provided by the test developers (i.e., clinical and college samples). Additional research in this regard may highlight the necessity of another normative comparison sample explicitly comprised of a forensic population.

Considering the diversity of referrals in the archival sample, examining whether the type of

referral was different across gender would be interesting to investigate and determine whether the referral type or source was a mediator of gender differences.

Since the PAI is quite broad in the number of constructs it assesses, another future study may find it beneficial to explore constellations of scales as they relate to psychopathologies, such as examining the validity of responses and the function of gender or potential complications in treatment based on gender. Approaching from the view of those factors as a whole rather than individual variables, which was not done in the present study, would help in examining the data in less of a monolithic manner.

The notion that self-report inventories may contain gender-biased items has mainly been ignored (Lindsay et al., 2000). Furthermore, some items within currently used personality disorder self-report inventories may contain gender bias (Lindsay et al., 2000), including the PAI. Significant gender differences may represent gender bias as a form of test bias in the PAI. Based on the significant gender differences found in the present study, future studies could employ a confirmatory factor analysis to compare the overall factor structure of the archival sample to the structures presented by Morey (2007). An analysis of this kind may address whether the test accurately measures what it is designed to measure among men and women.

Lastly, the present study serves as a call to action for researchers within forensic psychology to explore gender minority populations in more depth, as the present study is limited to two genders. Prior research is also limited to male and female samples when discussing gender differences. One potential benefit of the PAI using standard norms for comparison rather than separate norms based on binary gender is that it may be more applicable to gender minority individuals than other means of personality assessment.

Clinical Implications and Recommendations for Evaluators

The gender differences found within the archival sample support the idea that evaluators need to consider the impact of gender on forensic assessment. The findings that men and women differ in their responses to the PAI, a widely used instrument in forensic assessment, further highlight the importance of incorporating multiple data sources to understand the factors most relevant to the examinee and aid in diagnostic formulation. In addition, they support the need to use validated measures using samples of women, thus allowing gender-specific interpretation. As stated by the test developers and reiterated by the present study, PAI results should never be interpreted in isolation (Morey, 2007), nor should any single measure. Specifically, when using the PAI, evaluators should consider input from a variety of sources and include the PAI as part of a more holistic assessment. Being aware of gender differences on various assessment instruments will ultimately allow evaluators to develop more ethical, nonbiased, and clinically appropriate recommendations. Assessment tools containing gender-specific norms and predictive validity should be used to appropriately assess individuals with unique risk profiles and minimize bias.

Although not a gender-specific tool, the PAI can assess both men and women in the realm of psychopathology and overall personality structure and functioning. Given the dearth of actuarial assessment tools for women, the PAI can play a role in forensic evaluation. The PAI is extremely useful in the information it provides related to personality and psychopathology. However, despite the helpful information the PAI offers, having gender-specific assessments would be more useful than looking at scores on the PAI alone. Furthermore, given that some of the findings regarding gender differences conflicted with prior research, the PAI may not be as sensitive as detecting nuances of gender. Although the PAI is a personality measure frequently used within a psycholegal context, this study found that the PAI does not perform similarly for

men and women in a broad forensic sample, suggesting that gender-specific norms may be more useful when interpreting test scores rather than comparing scores to the publisher's mixed-gender comparison sample. Additional studies looking at gender differences among forensic samples of this kind will further add to the utility of this measure.

Conclusion

In conclusion, research on gender differences in a forensic population is growing, although it remains limited, most notably in samples like the one analyzed in the present study. This study aimed to add to the existing literature regarding the clinical profiles of men and women presenting for forensic evaluation. Awareness of the need for gender-sensitive assessment is increasing in forensic psychology as research has shown that men and women present with different risk factors (Van Voorhis et al., 2010). In terms of previously identified shared risk factors, research has shown that the development and course look different between men and women (de Vogel et al., 2022). Overall, significant gender differences were found on a widely used measure in the present study, supporting previous research that women and men in forensic populations have different clinical presentations and consequently, have gender-specific needs for assessment and treatment. Maintaining a gender-responsive approach to forensic assessment is more appropriate than taking a gender-neutral approach, as it will lead to more valid assessment and recommendations for treatment (McKeown, 2010; Van Voorhis et al., 2010), which may in turn lead to reductions in recidivism.

Being aware of general patterns of personality functioning among men and women is essential when conducting forensic evaluations; however, it is of even more importance to simultaneously recognize the uniqueness of each individual. As the findings of this study illustrate, while there are differences between genders, there are also similarities. Some research

has cautioned against a “dramatic interpretation of the pervasive gender differences in personality” as the distributions of traits for men and women “are largely overlapping” (Weisberg et al., 2011, p. 10). Although there is some debate on the cause and development, previous research has shown that gender differences exist within personality structure and functioning (Weisberg et al., 2011). Despite the previous statement, men and women can also exhibit individual characteristics that do not follow the general patterns. It must be understood that while there may be significant differences between genders, there may also be coinciding similarities.

Evaluators must always remain mindful of their own biases regarding gender when determining risk, offering diagnostic considerations, and formulating treatment recommendations. The legal ramifications based on opinions stated in forensic evaluations should not be minimized. Ultimately, there are very limited tools that look at psychopathology according to a set of biological experiences (i.e., sex assigned at birth), let alone how society interacts and responds with cultural identities (i.e., gender identity) that may be based on preconceived notions or biases. The role of the forensic examiner is to be aware of the limitations of the measures they use to assess an individual and bear in mind the impact of cultural factors, including gender when forming conclusions about risk and treatment planning. When a biased decision occurs in a forensic setting, it can have profound consequences, and the examinee is the one who will suffer.

The present findings highlight the complex nature of forensic evaluation and seek to enrich the field’s understanding of issues related to gender that are present within this population. The hope is that this study will reinforce the importance of taking a

gender-responsive approach to assessment. Assessment tools must be developed and used in a way that is appropriate to the population being considered. Specifically, gender-sensitive assessment tools are necessary to produce valid clinical profiles, particularly in the assessment of women. Since men and women were found to score differently on the PAI, the test developers should consider developing gender-specific norms for comparison when interpreting the results. One of the main conclusions of the present study is that investigating subtleties in instrumentation associated with gender may aid in leveling the information that exists about different groups, consequently diminishing possible injustice that could transpire by means of systematic bias based on test scores.

Finally, the field of forensic psychology must consider the spectrum of identities that encompass the construct of gender. The current literature is minimal regarding gender identities outside of the binary, particularly the forensic assessment of gender minority individuals (Saleh et al., 2021). The limited literature may reflect the overrepresentation of males in the criminal justice system or that most forensic settings (i.e., prisons) typically categorize and house individuals solely based on sex assigned at birth. The present study is similarly limited. The field cannot say we fully understand gender differences in personality structure if we do not consider marginalized gender identities.

References

- American Psychiatric Association. (2022). *Diagnostic and statistical manual of mental disorders* (5th ed., Text Revision). <https://www.appi.org/dsm-5-tr>
- American Psychological Association. (2013). Specialty guidelines for forensic psychology. *American Psychologist*, 68(1), 7–19. <https://doi.org/10.1037/a0029889>
- Bohle, A., & de Vogel, V. (2017). Gender differences in victimization and the relation to personality disorders in forensic psychiatry. *Journal of Aggression, Maltreatment & Trauma*, 26(4), 411–429. <https://doi.org/10.1080/10926771.2017.1284170>
- Boyle, G. J., & Lennon, T. J. (1994). Examination of the reliability and validity of the personality assessment inventory. *Journal of Psychopathology and Behavioral Assessment*, 16(3), 173–187. <https://doi.org/10.1007/BF02229206>
- Cortoni, F. (2018). *Women who sexually abuse: Assessment, treatment & management*. Safer Society Press.
- Cortoni, F., & Gannon, T. A. (2013). What works with female sexual offenders. In L. A. Craig, L. Dixon, & T. A. Gannon (Eds.), *What works in offender rehabilitation* (pp. 271–284). Wiley. <https://doi.org/10.1002/9781118320655.ch15>
- Cortoni, F., & Hanson, R. K. (2005). A review of the recidivism rates of adult female sexual offenders (Report No R-169). Correctional Service of Canada.
- Davidson, M., Sorensen, J. R., & Reidy, T. J. (2015). Gender-responsiveness in corrections: Estimating female inmate misconduct risk using the Personality Assessment Inventory (PAI). *Law and Human Behavior*, 40(1), 72–81. <https://doi.org/10.1037/lhb0000157>
- de Vogel, V., & Nicholls, T. L. (2016). Gender matters: An introduction to the special issues on women and girls. *International Journal of Forensic Mental Health*, 15(1), 1–25. <https://doi.org/10.1080/14999013.2016.1141439>
- de Vogel, V., Stam, J., Bouman, Y. H. A., Ter Horst, P., & Lancel, M. (2016). Violent women: A multicentre study into gender differences in forensic psychiatric patients. *The Journal of Forensic Psychiatry & Psychology*, 27(2), 145–168. <https://doi.org/10.1080/14789949.2015.1102312>
- de Vogel, V., Stam, J., Bouman, Y. H. A., Ter Horst, P., & Lancel, M. (2022). Gender differences in substance abuse history and offending behavior: A multicentre study in Dutch forensic psychiatry. *Journal of Forensic Psychology Research and Practice*, 22(1), 1–17. <https://doi.org/10.1080/24732850.2021.1945833>

- Douglas, K. S., Hart, S. D., Webster, C. D., & Belfrage, H. (2013). HCR-20V3: Assessing risk of violence – User guide. Burnaby, Canada: Mental Health, Law, and Policy Institute, Simon Fraser University.
- Edens, J. F., Cruise, K. R., & Buffington-Vollum, J. K. (2001). Forensic and correctional applications of the personality assessment inventory. *Behavioral Sciences & the Law*, *19*(4), 519–543. <https://doi.org/10.1002/bsl.457>
- Gannon, T. A., & Cortoni, F. (2010). *Female sexual offenders: Theory, assessment and treatment*. John Wiley & Sons Ltd. <https://doi.org/10.1002/9780470666715>
- Grella, C. E., & Rodriguez, L. (2011). Motivation for treatment among women offenders in prison-based treatment and longitudinal outcomes among those who participate in community aftercare. *Journal of Psychoactive Drugs*, *43*(sup1), 58–67. <https://doi.org/10.1080/02791072.2011.602275>
- Grimbos, T., Penney, S. R., Fernane, S., Prosser, A., Ray, I., & Simpson, A. I. F. (2016). Gender comparisons in a forensic sample: Patient profiles and HCR-20: V2 reliability and item utility. *International Journal of Forensic Mental Health*, *15*(2), 136–148. <https://doi.org/10.1080/14999013.2016.1152617>
- Hart, S. D., Michie, C., & Cooke, D. J. (2007). Precision of actuarial risk assessment instruments: Evaluating the ‘margins of error’ of group v. individual predictions of violence. *British Journal of Psychiatry*, *190*(S49), s60–s65. <https://doi.org/10.1192/bjp.190.5.s60>
- Hilton, N. Z., Harris, G. T., & Rice, M. E. (2006). Sixty-six years of research on the clinical versus actuarial prediction of violence. *The Counseling Psychologist*, *34*(3), 400–409. <https://doi.org/10.1177/0011000005285877>
- Lindsay, K. A., Sankis, L. M., & Widiger, T. A. (2000). Gender bias in self-report personality disorder inventories. *Journal of Personality Disorders*, *14*(3), 218–232. <https://doi.org/10.1521/pedi.2000.14.3.218>
- McKeown, A. (2010). Female offenders: Assessment of risk in forensic settings. *Aggression and Violent Behavior*, *15*(6), 422–429. <https://doi.org/10.1016/j.avb.2010.07.004>
- Monahan, J. (1981). *Predicting violent behavior: An assessment of clinical techniques*. Sage Press.
- Morey, L. C. (2007). *Personality Assessment Inventory (PAI): Professional manual* (2nd ed.). Psychological Assessment Resources.

- Morey, L. C., & Benson, K. T. (2016). An investigation of adherence to diagnostic criteria, revisited: Clinical diagnosis of the DSM-IV/DSM-5 section II personality disorders. *Journal of Personality Disorders, 30*(1), 130–144. https://doi.org/10.1521/pedi_2015_29_188
- Morey, L. C., & Quigley, B. D. (2002). The use of the personality assessment inventory (PAI) in assessing offenders. *International Journal of Offender Therapy and Comparative Criminology, 46*(3), 333–349. <https://doi.org/10.1177/0306624X02463007>
- Ndung'u, J. W., Lunjalu, P. M., & Wamalwa, D. N. (2021). Analysis of personality types among female offenders in Lang'ata women prison in Kenya. *Journal of Educational Management and Leadership, 2*(2), 50–57. <https://doi.org/10.33369/jeml.2.2.50-57>
- Parker, A. J., Mulay, A. L., & Gottfried, E. D. (2020). The personality assessment inventory (PAI): Treatment scales and interpersonal characteristics in a sample of men charged with or convicted of a sexual offense. *Journal of Forensic Sciences, 65*(6), 2050–2057. <https://doi.org/10.1111/1556-4029.14512>
- Poldrack, R. A., Monahan, J., Imrey, P. B., Reyna, V., Raichle, M. E., Faigman, D., & Buckholz, J. W. (2018). Predicting violent behavior: What can neuroscience add? *Trends in Cognitive Sciences, 22*(2), 111–123. <https://doi.org/10.1016/j.tics.2017.11.003>
- Rettinger, L. J., & Andrews, D. A. (2010). General risk and need, gender specificity, and the recidivism of female offenders. *Criminal Justice and Behavior, 37*(1), 29–46. <https://doi.org/10.1177/0093854809349438>
- Saleh, F. M., Bradford, J. M., & Brodsky, D. J. (2021). *Sex offenders: Identification, risk assessment, treatment, and legal issues*. Oxford University Press.
- Skeem, J., Schubert, C., Stowman, S., Beeson, S., Mulvey, E., Gardner, W., & Lidz, C. (2005). Gender and risk assessment accuracy: Underestimating women's violence potential. *Law and Human Behavior, 29*(2), 173–186. <https://doi.org/10.1007/s10979-005-3401-z>
- Slavin-Mulford, J., Sinclair, S. J., Stein, M., Malone, J., Bello, I., & Blais, M. A. (2012). External validity of the personality assessment inventory (PAI) in a clinical sample. *Journal of Personality Assessment, 94*(6), 593–600. <https://doi.org/10.1080/00223891.2012.681817>
- Tabachnick, B. G., & Fidell, L. S. (2019). *Using multivariate statistics* (7th ed.). Pearson.
- The Sentencing Project. (2022, May). Incarcerated women and girls. <https://www.sentencingproject.org/publications/trends-in-u-s-corrections/>

Van Voorhis, P., Wright, E. M., Salisbury, E., & Bauman, A. (2010). Women's risk factors and their contributions to existing risk/needs assessment: The current status of a gender-responsive supplement. *Criminal Justice and Behavior*, 37(3), 261–288.
<https://doi.org/10.1177/0093854809357442>

Weisberg, Y. J., DeYoung, C. G., & Hirsh, J. B. (2011). Gender differences in personality across the ten aspects of the Big Five. *Frontiers in Psychology*, 2(178), 1–11.
<https://doi.org/10.3389/fpsyg.2011.00178>

Wise, E. A., Streiner, E. L., & Walfish, S. (2010). A review and comparison of the reliabilities of the MMPI-2, MCMI-III, and PAI presented in their respective test manuals. *Measurement and Evaluation in Counseling and Development*, 42(4), 246–254.
<https://doi.org/10.1177/0748175609354594>

APPENDIX A: PERMISSIONS FORM



Sent Via Email:

May 28, 2021

Rebecca J. Elliott
Antioch University New England
40 Avon Street
Keene, NH 03431

Dear Ms. Elliott:

In response to your recent request, permission is hereby granted to you to include up to a total of three (3) sample items from the Personality Assessment Inventory (PAI) in your dissertation/thesis titled, *Examining Gender Differences within a Forensic Sample Using the Personality Assessment Inventory*. Any further use or publication in a Journal (or otherwise) will require additional permission.

This Agreement is subject to the following restrictions:

- (1) Any and all materials used will contain the following credit line:
"Reproduced by special permission of the Publisher, Psychological Assessment Resources, Inc. (PAR), 16204 North Florida Avenue, Lutz, Florida 33549 from the Personality Assessment Inventory by Leslie C. Morey, PhD, Copyright 1991 by PAR. Further reproduction is prohibited without permission of PAR."
- (2) None of the material may be sold, given away, or used for purposes other than those described above.
- (3) Payment of a permission fee will be waived.
- (4) One copy of any of the material reproduced will be sent to PAR to indicate that the proper credit line has been used.

TWO COPIES of this Permission Agreement should be signed and returned to me to indicate your agreement with the above restrictions. I will then sign it for PAR and return a fully executed copy to you for your records.

Sincerely,

Permissions Specialist

ACCEPTED AND AGREED:

BY: REBECCA J. ELLIOTT
DATE: 6/11/2021

ACCEPTED AND AGREED:

BY:
DATE: June 17, 2021

APPENDIX B: DESCRIPTION OF PAI SCALES

Scale (Acronym/No. of Items)	Description
Validity Scales	
Inconsistency (INC/10 item pairs)	Indicates if client is answering consistently throughout inventory. Each pair consists of highly correlated (i.e., positively, negatively) items.
Infrequency (INF/8)	Indicates if client is responding carelessly, randomly, or idiosyncratically. Items are neutral with respect to psychopathology and have either extremely high or low endorsement rates.
Negative Impression (NIM/9)	Suggests an exaggerated, unfavorable impression or malingering.
Positive Impression (PIM/9)	Suggests the presentation of a very favorable impression or reluctance to admit minor flaws.
Clinical Scales	
Somatic Complains (SOM/24)	Focuses on preoccupation with health matters and somatic complaints typically associate with somatization or conversion disorders.
Anxiety (ANX/24)	Focuses on phenomenology and observable signs of anxiety with an emphasis on assessment across different response modalities.
Anxiety-Related Disorders (ARD/24)	Focuses on symptoms and behaviors related to specific anxiety disorders, particularly phobias, traumatic stress, and obsessive-compulsive disorders.
Depression (DEP/24)	Focuses on symptoms and phenomenology of depressive disorders.
Mania (MAN/24)	Focuses on affective, cognitive, and behavioral symptoms of mania and hypomania.
Paranoia (PAR/24)	Focuses on symptoms of paranoid disorders and more enduring characteristics of paranoid personality.
Schizophrenia (SCZ/24)	Focuses on symptoms relevant to the broad spectrum of schizophrenic disorders.
Borderline Features (BOR/24)	Focuses on attributes indicative of a borderline level of personality functioning, including unstable and fluctuating interpersonal relations, impulsivity, affective

- lability and instability, and uncontrolled anger.
- Antisocial Features (ANT/24) Focuses on history of illegal acts and authority problems, egocentrism, lack of empathy and loyalty, instability, and excitement-seeking.
- Alcohol Problems (ALC/12) Focuses on problematic consequences of alcohol use and features of alcohol dependence.
- Drug Problems (DRG/12) Focuses on problematic consequences of drug use (i.e., prescription, illicit) and features of drug dependence.

Treatment Scales

- Aggression (AGG/18) Focuses on characteristics and attitudes related to anger, assertiveness, hostility, and aggression.
- Suicidal Ideation (SUI/12) Focuses on suicidal ideation, ranging from hopelessness to thoughts and plans for the suicidal act.
- Stress (STR/8) Measures the impact of recent stressors on major life areas.
- Nonsupport (NON/8) Measures a lack of perceived social support, considering both the level and quality of available support.
- Treatment Rejection (RXR/8) Focuses on attributes and attitudes indicating a lack of interest and motivation in making personal changes of a psychological or emotional nature.

Interpersonal Scales

- Dominance (DOM/12) Assesses the extent to which a person is controlling and independent in personal relationships. A bipolar dimension, with a dominant style at the high end and a submissive style at the low end.
- Warmth (WRM/12) Assesses the extent to which a person is interested in supportive and empathic personal relationships. A bipolar dimension, with a warm, outgoing style at the high end and a cold, rejecting style at the low end.

Note: Reproduced by special permission of the Publisher, Psychological Assessment Resources, Inc. (PAR), 16204 North Florida Avenue, Lutz, Florida 33549 from the Personality Assessment Inventory by Leslie C. Morey, PhD, Copyright 1991 by PAR. Further reproduction is prohibited without permission of PAR.

APPENDIX C: SYNOPSIS OF MODAL CLUSTER PROFILES

Modal Profile Configurations	Brief Description
Cluster 1	“This pattern suggests a person who is reporting some turmoil in his or her life that might be the source of some stress, but not to the point where prominent symptoms are observed” (Morey, 2007, p. 50).
Cluster 2	“This pattern suggests a person who is severely depressed and withdrawn” (Morey, 2007, p. 53).
Cluster 3	“Individuals with this pattern tend to currently have, or once have had, a severe drinking problem and have experienced a number of physiological problems that may be related to their drinking” (Morey, 2007, p. 53).
Cluster 4	“This pattern suggests a person with a history of acting-out behavior, most notably in the area of substance abuse but probably involving other behaviors as well” (Morey, 2007, p. 58).
Cluster 5	“This pattern suggests an individual who is experiencing a more or less acute reaction to current stressors, with anxiety and moodiness being the most prominent complaints” (Morey, 2007, p. 58).
Cluster 6	“This pattern suggests a person who is somewhat socially isolated and experiencing some confusion and difficulties in thinking and concentration” (Morey, 2007, p. 63).
Cluster 7	“This pattern suggests an individual who is severely depressed, anxious, and agitated” (Morey, 2007, p. 63).
Cluster 8	“This pattern suggests an individual who is reporting marked concerns about his or her physical functioning” (Morey, 2007, p. 68).
Cluster 9	“This pattern suggests an individual who currently has, or has had, a severe drinking problem and has experienced a number of adverse consequences related to his or her drinking, including disruptions in relationships and work,” (Morey, 2007, p. 68).
Cluster 10	“This pattern suggests an individual who is unhappy, angry, resentful, and confused” (Morey, 2007, p. 73).

This table was created using descriptions taken directly from the test developers (Morey, 2007). Of note, these descriptions do not include all available information provided by the test developers (i.e., specific scale elevations) about each cluster, rather these are brief explanations.

APPENDIX D: TABLES

Table 1

Demographic Characteristics of Sample by Gender

	Female	Male
Marital Status		
Single	54.5%	54.4%
Married	16.2%	20.7%
Divorced	20.2%	17.7%
Widowed	1.0%	2.1%
Other	8.1%	8.4%
Race/Ethnicity		
African American	2.8%	5.0%
Asian	4.2%	1.0%
Caucasian	87.5%	92.0%
Hispanic	1.4%	0.0%
Other	4.2%	2.0%

¹ There were no significant differences.

Table 2*PAI Sample Means by Gender and Normative Comparisons*

PAI scale	Female				Male			
	Sample Mean	SD	Normative Mean	Cohen's d	Sample Mean	SD	Normative Mean	Cohen's d
<i>Validity Scales</i>								
ICN	49.91	7.76	49.77	.018	52.86**	9.07	50.62	-.092
INF	53.86**	9.17	48.96	.534	53.97**	9.30	51.12	.299
NIM	56.40**	13.16	49.33	.538	56.43**	13.96	50.18	.473
PIM	49.63	12.78	49.93	-.024	48.58	11.42	50.05	-.033
<i>Clinical Scales</i>								
SOM	57.16**	14.96	50.19	.466	55.53**	12.26	49.99	.479
ANX	59.62**	15.57	50.75	.569	56.54**	14.46	49.26	.665
ARD	61.00**	15.40	50.97	.651	57.09**	14.84	48.93	.784
DEP	61.57**	16.67	50.19	.682	60.63**	15.39	49.71	.711
MAN	45.72*	9.94	48.55	-.285	49.91	10.67	51.55	-.587
PAR	57.43**	14.70	48.95	.577	57.59**	13.96	51.12	.429
SCZ	53.02*	13.62	49.23	.278	54.80**	13.56	50.83	.161
BOR	58.46**	15.13	49.6	.586	58.01**	14.05	50.5	.526
ANT	50.04*	11.70	46.91	.268	56.26**	11.68	53.38	-.286
ALC	48.07	10.85	47.33	.068	54.37	15.42	52.77	-.433
DRG	56.12**	18.59	49.03	.381	58.35**	18.17	51.42	.253
<i>Treatment Consideration Scales</i>								
AGG	46.77	12.21	48.47	-.139	50.39	13.06	51.65	-.400
SUI	55.25**	16.39	49.08	.377	57.81**	18.21	50.57	.286
STR	57.45**	11.10	50.22	.652	58.30**	11.92	49.61	.707
NON	54.44**	13.29	48.71	.432	54.15*	12.68	51.48	.223
RXR	46.80*	11.53	49.89	-.268	44.46**	12.19	50.15	-.291
<i>Interpersonal Scales</i>								
DOM	46.40	12.14	48.35	-.160	47.97**	10.37	51.72	-.438
WRM	48.97	11.24	51.37	-.214	46.76	11.92	48.45	.046

** indicates sample means that were significantly different than normative means based on $p < .001$

* indicates sample means that were significantly different than normative means based on $p < .01$

Table 3*PAI Sample Means and Differences by Gender*

PAI scale	Female		Male		Significance (p-value)	Partial Eta Squared
	Sample Mean	SD	Sample Mean	SD		
Validity Scales						
ICN	49.91	7.76	52.86	9.07	.005*	.023
INF	53.86	9.17	53.97	9.30	.923	.000
NIM	56.40	13.16	56.43	13.96	.989	.000
PIM	49.63	12.78	48.58	11.42	.462	.002
Clinical Scales						
SOM	57.16	14.96	55.53	12.26	.298	.003
ANX	59.62	15.57	56.54	14.46	.084	.009
ARD	61.00	15.40	57.09	14.84	.030	.014
DEP	61.57	16.67	60.63	15.39	.620	.001
MAN	45.72	9.94	49.91	10.67	<.001**	.032
PAR	57.43	14.70	57.59	13.96	.929	.000
SCZ	53.02	13.62	54.80	13.56	.275	.004
BOR	58.46	15.13	58.01	14.05	.793	.000
ANT	50.04	11.70	56.26	11.68	<.001**	.056
ALC	48.07	10.85	54.37	15.42	<.001**	.039
DRG	56.12	18.59	58.35	18.17	.309	.003
Treatment Consideration Scales						
AGG	46.77	12.21	50.39	13.06	.019	.016
SUI	55.25	16.39	57.81	18.21	.227	.004
STR	57.45	11.10	58.30	11.92	.548	.001
NON	54.44	13.29	54.15	12.68	.849	.000
RXR	46.80	11.53	44.46	12.19	.104	.008
Interpersonal Scales						
DOM	46.40	12.14	47.97	10.37	.230	.004
WRM	48.97	11.24	46.76	11.92	.117	.007

** indicates scales that were significantly different between gender means based on $p < .001$

* indicates scales that were significantly different between gender means based on $p < .01$