

7-22-2015

Common Profiles and Clinical Utility of the MMPI-2-RF in Competency to Stand Trial Evaluations

Laura K. Gilman
Indiana University of Pennsylvania

Follow this and additional works at: <http://knowledge.library.iup.edu/etd>

Recommended Citation

Gilman, Laura K., "Common Profiles and Clinical Utility of the MMPI-2-RF in Competency to Stand Trial Evaluations" (2015).
Theses and Dissertations (All). 412.
<http://knowledge.library.iup.edu/etd/412>

This Dissertation is brought to you for free and open access by Knowledge Repository @ IUP. It has been accepted for inclusion in Theses and Dissertations (All) by an authorized administrator of Knowledge Repository @ IUP. For more information, please contact cclouser@iup.edu, sara.parme@iup.edu.

COMMON PROFILES AND CLINICAL UTILITY OF THE MMPI-2-RF IN
COMPETENCY TO STAND TRIAL EVALUATIONS

A Dissertation

Submitted to the School of Graduate Studies and Research

in Partial Fulfillment of the

Requirements for the Degree

Doctor of Psychology

Laura K. Gilman

Indiana University of Pennsylvania

August 2015

Indiana University of Pennsylvania
School of Graduate Studies and Research
Department of Psychology

We hereby approve the dissertation of

Laura K. Gilman

Candidate for the degree of Doctor of Psychology

June 12, 2015

Signature on File

David J. LaPorte, Ph.D.
Professor of Psychology, Advisor

June 12, 2015

Signature on File

Margaret Reardon, Ph.D.
Associate Professor of Psychology

June 12, 2015

Signature on File

John A. Mills, Ph.D.
Professor of Psychology

ACCEPTED

Signature on File

Randy L. Martin, Ph.D.
Dean
School of Graduate Studies and Research

Title: Common Profiles and Clinical Utility of the MMPI-2-RF in Competency to Stand Trial Evaluations

Author: Laura K. Gilman

Dissertation Chair: Dr. David J. LaPorte

Dissertation Committee Members: Dr. Margaret Reardon
Dr. John A. Mills

Competency to stand trial evaluations are the most common psychiatric evaluations ordered by criminal courts, and there is considerable variation in competency evaluation procedures and judgment processes. The MMPI-2 is utilized by forensic evaluators more frequently than any other test, and the MMPI-2-RF is gaining popularity as well. Although these measures do not assess psycholegal knowledge, they provide relevant information about personality characteristics and psychological functioning that underlie competency abilities. Findings regarding the MMPI-2 as it relates to competency, however, have been sparse and inconsistent, and there are no peer-reviewed data regarding MMPI-2-RF profiles among competent versus incompetent individuals. In the current project, the utility of the MMPI-2 and the MMPI-2-RF in differentiating competent and incompetent to stand trial defendants was investigated, and the two tests were then compared. The results demonstrated that overall, incompetent individuals had higher scale elevations, suggesting this group experienced more general maladjustment. Incompetent individuals scored significantly higher than competent individuals on MMPI-2 scales Infrequency (F), Psychasthenia (7), and Paranoia (6). For the MMPI-2-RF, incompetent individuals displayed relative elevations on scales Infrequent Responses (F-r), Demoralization (RCd), and Dysfunctional Negative Emotions (RC7) as compared to competent individuals. Competency prediction models were created using the MMPI-2

and MMPI-2-RF, and both tests substantially increased accuracy for the prediction of competency status. The overall competency classification rate of just over 70% accuracy is similar to the reliability rates among forensic evaluators in routine practice. Although the two tests had similar rates of significance, the MMPI-2-RF was marginally superior in predicting competency.

ACKNOWLEDGEMENTS

There are a number of people whom I would like to thank, as I received a great deal of support throughout the course of this project. First, I am grateful for my significant other, Peter Stewart, who has provided me with unconditional support, guidance, grace, and sanity as I faced and conquered various obstacles. I cannot imagine how different this project and my life would be without his companionship, and I am so thankful for his presence. I would also like to thank my mother and father, Beth and Steve Gilman, for their undying encouragement, love, pride, and patience, as well as my brother, Daniel, for his geniality and warmheartedness.

Dr. Dave LaPorte was extremely supportive throughout this entire endeavor, and his assistance was important as I ran into a particularly large hurdle that I feared would impede the project. He reassured me that we would figure something out, and he was right. He has also encouraged my pursuit of forensic psychology throughout my graduate school experience, all the while maintaining an excellent sense of humor. Thank you also to Dr. Randy Martin, who has provided a great deal of wisdom while also assuring me that a project's limitations are expected and okay. Dr. Jay Mills backed me at critical moments in this project, including when I struggled to find a data sample as well as when I imminently needed a new committee member. I am quite grateful for his help, encouragement, and insightful comments. I would also like to thank Dr. Margaret Reardon for serving on my committee, as she was consistently friendly and open to sharing her great knowledge. Finally, thank you to those at Saint Elizabeths Hospital— notably, Dr. Richard Gontang, Dr. Christine Kelley, and Mrs. Tanya Hatcher for making

it possible for me to collect data at the distinguished hospital where I have had the pleasure of training and growing.

TABLE OF CONTENTS

Chapter		Page
1	INTRODUCTION	1
2	LITERATURE REVIEW	5
	Competency to Stand Trial	5
	The History of Competency	5
	The Current Competency Standard	7
	Competency to Stand Trial in Pennsylvania	12
	Competency to Stand Trial Evaluations	13
	Minnesota Multiphasic Personality Inventory (MMPI)	25
	The MMPI-2	29
	The MMPI-2-RF	34
	Functional Utility of the MMPI-2/MMPI-2-RF With Forensic Populations	37
	MMPI in Classification and Prediction Among Forensic Populations	38
	MMPI and Competency Assessment	42
	Hypotheses	47
3	METHODS	49
	Participants	49
	Patients	49
	Setting	50
	Procedure	50
	Planned Analyses	51
4	RESULTS	56
	Analysis of Hypothesis 1	70
	Analysis of Hypothesis 2	72
	Analysis of Hypothesis 3	73
	Power Analysis	76
5	DISCUSSION	81
	Summary and Review	81
	The Present Sample and Descriptive Statistics	84
	Interpretation of Significant Results	95
	Limitations and Future Directions	103
	REFERENCES	108

APPENDICES	122
Appendix A- WAIS and BPRS	122
Appendix B- MMPI-2/ MMPI-2-RF Scales	124

LIST OF TABLES

Table		Page
1	Demographic Characteristics.....	58
2	Psychological Characteristics.....	61
3	Means and Standard Deviations of the MMPI-2.....	62
4	Means and Standard Deviations of the MMPI-2-RF	64
5	MMPI-2 Scale Intercorrelations.....	68
6	MMPI-2-RF Scale Intercorrelations.....	69
7	Between-Group Multivariate Analysis of Variance With MMPI-2.....	71
8	Between-Group Multivariate Analysis of Variance With MMPI-2-RF.....	73
9	Observed Power of MMPI-2 Scales.....	77
10	Observed Power of MMPI-2-RF Scales.....	78
11	Power for Tests of Model Coefficients.....	79

LIST OF FIGURES

Figure		Page
1	Mean scores of MMPI-2 scales by competency status.....	63
2	Mean scores of MMPI-2-RF scales by competency status.....	65
3	Central and non-central distributions for a sample size of 64.....	80

CHAPTER 1

INTRODUCTION

Competency to stand trial doctrine protects the welfare of individuals and enforces fairness within the legal system. The Sixth Amendment guarantees criminal defendants the right to a fair trial, and serious mental deficiencies may render defendants incapable of properly defending themselves (Grisso, 1988). The current standard in the U. S. is based on the *Dusky v. United States* (1960) decision that defendants must not only be oriented to time and place, but they must be able to consult with counsel and have a rational as well as factual understanding of the proceedings against them. If a “bona fide doubt” of the defendant’s competency is raised, the judge may order an evaluation at any point during the adjudicative process (Grisso, 1988).

Competency to stand trial evaluations are by far the most common psychiatric evaluation ordered by criminal courts (Miller, 2003), as it has been estimated that there are approximately 60,000 competency evaluations each year (Bonnie & Grisso, 2000). Although competency is a legal decision, judges defer to the opinions of examiners in approximately 89% of cases (Gowensmith, Murrie, & Boccaccini, 2012). Given the abundance and weight of the decisions assigned to evaluators, it is imperative that their methods, judgments, and reports be accurate and thorough.

The standards established in *Dusky* are considered somewhat ambiguous (Mossman et al., 2007). Consequently, there is considerable variation in competency evaluation procedures and judgment processes (Cooper & Zapf, 2003). A number of forensic evaluators use instruments designed specifically for competency evaluations, including the MacArthur Competency Assessment Instrument Tool—Criminal

Adjudication (MacCAT-CA). Nevertheless, traditional measures such as the Minnesota Multiphasic Personality Inventory-2 (MMPI-2) are more commonly used. In fact, surveys of forensic psychologists and psychiatrists have demonstrated that the MMPI-2 is likely the most frequently utilized instrument in competency to stand trial evaluations (Borum & Grisso, 1995). More current data regarding the frequency of MMPI-2 use is less specific, indicating that the MMPI-2 is commonly used across various types of forensic evaluations (Lally, 2003), and it is the most frequently used multiscale inventory in forensic evaluations of adults (Archer, Buffington-Vollum, Stredny, & Handel, 2006).

In addition to exploring which tests are commonly used in competency evaluations, various characteristics regarding incompetent and competent defendants have been presented. One of the most substantial findings is that defendants with a psychotic disorder diagnosis are eight times more likely to be deemed incompetent than those without (Pirelli, Gottdiener, & Zapf, 2011). Some studies have also presented MMPI profile patterns among competent and incompetent defendants. For example, a recent meta-analysis found that scales F (Infrequency), 6 (Paranoia), and 8 (Schizophrenia) are significantly higher among incompetent defendants (Pirelli, Gottdiener, & Zapf, 2011). Other findings concerning the MMPI as it relates to competency have generally been sparse and inconsistent. Numerous investigators have presented meaningful findings concerning the MMPI's utility in classifying and predicting in other forensic domains, although these data tend to be dated. For example, it has been used to distinguish types of criminals (Megargee & Bohn, 1979), subgroups among sex offenders (Anderson, Kuncze, & Rich, 1979; Kalichman, Szymanowski,

McKee, Taylor, & Craig, 1989; Herkov, Gynther, Thomas, & Myers, 1996), and in predicting offenders' treatment participation (Kalichman, Shealy, & Craig, 1990).

Considering the fact that the MMPI-2 is the primary instrument used to evaluate competency to stand trial, research regarding patterns and guidelines among incompetent and competent defendant groups is sorely lacking. Furthermore, the recently developed and significantly shorter MMPI-2-RF presents newfound potential in describing such defendants with improved Restructured Clinical (RC) Scales and substantive scales such as Thought Dysfunction scale (THD). Various MMPI-2-RF validity scales have been found to be effective in detecting deceit (Sellbom, Toomey, Wygant, Kucharski, & Duncan, 2010), although literature examining the MMPI-2-RF's utility specifically in competency to stand trial evaluations is absent.

The purpose of the present study is a) to present the prototypical MMPI-2 and MMPI-2-RF profiles of incompetent and competent defendants, b) to determine if there are significant differences between the profiles of such groups, and c) to examine whether the MMPI-2-RF differentiates groups more clearly than the MMPI-2. Profiles of competent individuals will be from individuals who have been found Not Guilty by Reason of Insanity (NGRI). These individuals are presumed to be competent, because competency is essential in order for the court hearing to occur. It should be noted that some NGRI individuals may have been incompetent at one time, in which case they likely underwent competency restoration treatment. However, the present setup ensures that both groups are composed primarily of severely mentally ill individuals who are mandated to receive treatment at the same psychiatric hospital. The aim was for these groups to be as similar as possible, so that competency status is the distinguishing factor.

The standards regarding competency to stand trial will be presented in turn, including a brief history, current doctrine, and Pennsylvania criteria. Competency to stand trial evaluations will then be discussed, including common procedures, measures, and outcomes. Next, an in-depth explanation of the MMPI, MMPI-2 and MMPI-2-RF will be provided. The utility of the MMPI among forensic populations will then be reported. Lastly, a review of the MMPI in competency evaluations will be presented.

CHAPTER 2

LITERATURE REVIEW

Competency to Stand Trial

The History of Competency

The doctrine of competency in the American system of criminal law is based on the underlying values of fairness and accuracy in criminal trial processes and results. Competency questions all reflect a concern for the defendant's meaningful participation, although they may be raised at various stages of criminal justice processes. During pretrial investigations, for example, competency to confess or to waive rights requires that the defendant appreciate the potential consequences of making self-incriminating statements. Other specific competencies include competency to plead guilty, to waive right to counsel, to stand trial, to be sentenced, and to be executed. Somewhat different demands are required of defendants at the various stages in the criminal trial, and this is reflected in the specific competencies. Competency to stand trial is the present focus, which involves having the ability to assist an attorney in developing and presenting a defense, and understanding the nature of the trial and its potential consequences (Grisso, 1988).

The origin of the rule denoting that an individual must be competent during criminal processes goes back to at least the 17th century. The question of whether a defendant is competent is believed to have begun in English courts, at times when defendants stood mute in court rather than making the required plea. At such times, the court deciphered whether the defendant was "mute of malice" or "mute by visitation of God." If the former was decided, the court ordered increasingly heavy weights to be

placed on the individual's chest until a plea was finally given. If the latter was deemed, the individual was not subjected to the weights. Initially, this category referred to deaf or mute individuals, but it later expanded to include the "lunatic" (Melton, Petrila, Poythress, & Slobogin, 2007).

In the 18th century, Blackstone documented that a "mad" defendant should not be arraigned "because he is not able to plead to it with the advise and caution that he ought" and should not be tried because "how can he make his defense?" (Blackstone 1783, p. 94). Early English courts reflected this idea, such as *Frith's Case* (1790), in which the court decided:

no man shall be called upon to make his defense at a time when his mind is in that situation as not to appear capable of so doing for... the inquiring of his guilt must be postponed to that season, when by collecting together his intellects, and having them entire, he shall be able so to model his defense and to ward off the punishment of the law. (p. 318)

In American courts, the incompetency plea was recognized in *United States v. Lawrence* (1835) when a man who attempted to assault President Andrew Jackson was declared unfit to stand trial. Deciphering competency has also been addressed in various other cases, such as *United States v. Chisolm* (1906), when the federal appeals court stated:

Does the mental impairment of the prisoner's mind, if such there be, whatever it is, disable him ... from fairly presenting his defense, whatever it may be, and make it unjust to go on with his trial at this time, or is he feigning to be in that condition... (p. 298)

Competency doctrine is likely an endeavor to uphold the Sixth Amendment, which guarantees criminal defendants the right to have effective counsel, to confront their accusers, and to present evidence. Furthermore, the defendant must be competent in order for the criminal proceedings to be accurate, fair, and dignified, serving societal as well as individual interests (Winick, 1983). Even a proceeding that resulted in an accurate outcome would be immoral if the convicted individual was unaware of what was happening or why (Melton et al., 2007).

The Current Competency Standard

Examiners' evaluations are needed to protect the welfare of individuals and to enforce fairness within the legal system. The thoroughness and quality of evaluations are momentous given their impact. A failure to identify marked impairments in a defendant's ability to meaningfully participate in adjudicative processes could compromise the fairness of the proceedings and outcomes. On the other hand, erroneously concluding that a defendant is incompetent would lead to unnecessary suspension of the proceedings as well as unwarranted confinement and treatment (Hoge, Bonnie, Poythress, Monahan, Eisenberg, & Feucht-Haviar, 1997).

The current legal standard for competency in the United States was established in *Dusky v. United States* (1960), in which the defendant, Milton Dusky, was charged with transporting a girl across state lines and subsequently raping her. The pretrial psychiatric evaluation assigned him a diagnosis of "schizophrenic reaction, chronic undifferentiated type." In a separate psychiatric report, it was stated that Dusky was unable to properly assist his attorney due to paranoid thoughts including being framed. Nevertheless, the court deemed Dusky competent to stand trial. He was then convicted of rape. The case

was appealed to the U.S. Supreme Court, which ruled in favor of Dusky and established new guidelines for competency:

It is not enough for the district judge to find that 'the defendant is oriented to time and place and has some recollection of events,' but that the test must be whether he has sufficient present ability to consult with his lawyer with a reasonable degree of rational understanding—and whether he has a rational as well as factual understanding of the proceedings against him. (p. 402)

The Dusky formulation is now considered the minimum constitutional standard in determining competency, and many states supplement this with more detail. For example, several states have added the requirement that deficiencies in the ability must be due to “mental disorder,” “mental disease or defect,” or other similar wording (Grisso, 1988).

The standard established in *Dusky* points to three prongs that clinicians should closely follow in determining a defendant’s competency: factual understanding of the proceedings; rational understanding of the proceedings; and the ability to consult with counsel. Assessing a defendant’s factual understanding involves evaluating his or her comprehension of his or her charges, his or her legal rights throughout the trial, the adversarial nature of the legal proceedings, the roles of the various participants in the courtroom, potential penalties, and the concept of a plea bargain. Lacking some factual knowledge is acceptable if the noted deficits are not a result of psychiatric impairment, and if the defendant is capable of learning the essential information. A rational understanding involves using reason and logic to appreciate how the criminal process applies to him/ herself. For example, a defendant with grandiose religious delusions may have a correct factual understanding of the legal process but may irrationally believe that

he or she is immune from punishment. This could impair the defendant's ability to participate effectively. Finally, the defendant must also have the ability to consult with counsel to make legal decisions, to prepare for his or her defense, and to participate in other activities that counsel may require. Such activities may include testifying, waiving a jury trial, and filing a plea bargain. To assist in assessing this domain, the evaluator may contact the defendant's attorney to learn about the defendant's behavior with the attorney thus far (Mossman et al., 2007).

In interpreting the *Dusky* standard, a number of other points should be recognized. A reasonable degree of understanding is required, as opposed to a perfect or complete understanding. The defendant's capacity is also emphasized, rather than the defendant's willingness to relate to counsel and to understand the proceedings (Melton et al., 2007). In addition, stating that the defendant must have "sufficient present ability" to work with his or her attorney provides limited guidance for the level of capacity that justifies a finding of competence (Mossman et al., 2007). The ambiguity inherent in this allows for flexibility in interpretation, although most observers agree that the threshold for a competency finding is not very high. Furthermore, the attorney's personality and the specifics regarding the case may greatly affect the defendant's ability to communicate with counsel. Simple charges may require a much lower level of understanding than more complicated offenses. Moreover, the presence of mental illness and a need for treatment are not sufficient for an incompetent determination. The emphasis on having both a "rational" and "factual" understanding implies that cognitive functioning is also important (Melton et al., 2007).

The question of whether someone is competent to stand trial can be raised at any point during the adjudication process, and warrants an examination of the defendant's abilities by a mental health professional (Grisso, 1988). The defense attorney most frequently raises the question, and in many states the prosecutor or judge may as well (Melton et al., 2007). If there is evidence that raises a "bona fide doubt" about the defendant's competency, most states require the judge to order an evaluation. This is considered a very low standard, and may be ordered on very little evidence (Grisso, 1988). It is also important to recognize that, at times, lawyers raise the question for tactical reasons such as to delay the trial (Roesch & Golding, 1980). Therefore, it is critical that examiners do not assume probable incompetency.

Once the Judge grants the motion for a competency evaluation, one or more clinicians examine the defendant. In the past, these evaluations were conducted by psychiatrists, and the defendants were generally hospitalized for one to three months. Since the 1970s, however, evaluations have begun to be more commonly conducted at outpatient facilities by psychologists and social workers (Grisso, 1988).

In formulating the opinion about competency, Mossman et al. (2007) summarized three questions the mental health professional must consider:

1. What symptoms does the defendant have, and what is the defendant's psychiatric diagnosis?
2. What is the relationship, if any, between the symptoms or diagnosis and the mental capabilities required under the jurisdiction's standard for competence to stand trial?

3. If the defendant appears incompetent to proceed with adjudication, how likely is it that appropriate restoration services would restore his competence, and what is the appropriate, least restrictive setting for such services?

If the defendant is judged to be competent, the criminal proceedings continue. If the defendant is decidedly incompetent, proceedings are suspended, and the defendant is treated until competency is restored. Until the 1970s, incompetent defendants were often committed to long-term or even lifetime confinement in a maximum-security unit, and were sometimes even forgotten by the court system (Roesch, Zapf, Golding, & Skeem, 1999). This changed in the 1972 case *Jackson vs. Indiana*. In this case, the court held that a defendant who is committed because of questioned competency to stand trial “cannot be held more than a reasonable period of time necessary to determine whether there is a substantial probability that he will attain the capacity in the foreseeable future” (pp. 737-738). That is, if there is not at least substantial probability of restoring competency, committing the defendant would be an unconstitutional deprivation of liberty. The literature is relatively consistent in identifying six months as a sufficient or “reasonable” amount of time to determine whether competency can be restored. If the defendant’s competency is not restored in this amount of time, the state may then detain the individual in a hospital only upon the determination of him or her being danger to self or others (Melton et al., 2007).

In addressing the probability of restoration, the evaluator should consider whether the defendant is incompetent because of a ‘treatable’ deficit, such as psychiatric symptoms caused by an illness that typically responds to medication, or a lack of prior exposure to information about the trial process. This would differ from a static,

irremediable condition such as mental retardation. The evaluator should also consider the defendant's previous psychiatric treatment and responses to treatment, as well as the nature of the presenting symptoms and current scientific knowledge about how well those symptoms typically respond to treatment (Mossman et al., 2007). Finally, the evaluator should indicate whether the defendant's treatment needs may be administered on an outpatient basis and the length of time that would likely be sufficient (Melton et al., 2007).

Competency to Stand Trial in Pennsylvania

In Pennsylvania, the Mental Health Procedures Act (1976), Section 402, defines incompetency to proceed on criminal charges as the following:

Whenever a person who has been charged with a crime is found to be substantially unable to understand the nature or object of the proceedings against him or to participate and assist in his defense, he shall be deemed incompetent to be tried, convicted or sentenced so long as such incapacity continues (p. 17).

Regarding the competency examination report, the following should be included:

1. Diagnosis of the person's mental condition
2. An opinion as to his capacity to understand the nature and object of the criminal proceedings against him and to assist in his defense
3. When so required, an opinion as to his mental condition in relation to the standards for criminal responsibility as then provided by law if it appears that the facts concerning his mental condition may also be relevant to the question of legal responsibility

4. When so requested, an opinion as to whether he had the capacity to have a particular state of mind, where such state of mind is a required element of the criminal charge. (pp. 19-20)

It is further specified that the report should be conducted by at least one psychiatrist, and the court is to determine competency within twenty days after receiving it.

Competency to Stand Trial Evaluations

There is no standard protocol for competency evaluations. This has led to a great deal of variation in competency evaluation procedures and judgment processes, and clinicians have a high degree of discretion and responsibility in their competency decisions (Cooper & Zapf, 2003). However, Melton and colleagues (2007) provide six elements that they recommend including in any assessment. First, pre-evaluation preparation and consultation involves obtaining court documents and evaluating the defendant's understanding of the charges. The clinician should then notify the defendant about the purpose and nature of the evaluation, as well as the limitations on confidentiality. Next, the clinician may obtain a brief social history and conduct a mental status examination to assess current cognitive, emotional, and behavioral functioning. Administration of a competency assessment measure may follow, as well as an evaluation of the defendant's understanding of, and reasoning about, the charges and case. Additional elements may potentially involve collateral psychological testing and issues related to amnesia and the defendant's statement about the offense.

Competency assessment standards. Regarding assessment measures, the Supreme Court provided guidelines concerning admissibility of scientific evidence in *Daubert v. Merrell Dow Pharmaceuticals, Inc.* (1993). In this case, the plaintiffs were

Jason Daubert and Eric Schuller, both born with serious limb reduction birth defects. These defects were allegedly caused by their mothers' ingestion of Bendectin, a drug intended to alleviate severe morning sickness. They sued Merrell Dow Pharmaceuticals, who subsequently submitted documents demonstrating that no published scientific study showed a link between Bendectin and birth defects. The Dauberts submitted expert evidence suggesting that Bendectin could cause birth defects, but such evidence was based on in vitro and in vivo animal studies, chemical structural analysis, and a reanalysis of previously published epidemiological studies. The vast majority of this evidence was judged to be either inadmissible at trial or insufficient to prove that Bendectin "more likely than not" caused the birth defects. The judge was assigned the role of gatekeeper in ensuring that expert evidence is consistent with the following scientific guidelines: (a) the technique can be and has been empirically tested, (b) it has been subjected to peer review, (c) the error rates of the technique are known, (d) there are standards for applying the technique, and (e) the technique is generally accepted in its scientific discipline. This standard is upheld in all federal courts (Archer et al., 2006).

Competency to stand trial assessment instruments. Many tools have been developed to assist in evaluating competency. Some are brief screening instruments designed to quickly identify defendants in need of a more extensive evaluation. Examples of these include the Competency Screening Test, the Georgia Court Competency Test, and the Computer-Assisted Determination of Competency to Proceed. These offer brief administration and standardized administration and scoring, but they have demonstrated poor validity beyond screening out defendants who are "obviously competent." Another type is a non-standardized semi-structured interview, such as the Competency

Assessment Instrument and the Interdisciplinary Fitness Interview. These more adequately tap into relevant legal issues, but the lack of standardized administration and criterion-based scoring limits their utility. For example, it is unclear whether many of these instruments evaluate defendants' abilities rather than their existing knowledge. Finally, there are standardized interviews created to systematically assess Dusky-related abilities, such as the MacArthur Competency Assessment Tool—Criminal Adjudication and the Evaluation of Competency to Stand Trial—Revised. These tools are more effective in distinguishing between a lack of knowledge and an incapacity for legal information, and they also provide comparison with relevant populations (Melton et al., 2007).

While various competency measures have been developed recently, their widespread use has not necessarily followed. More specifically, Pirelli, Gottdiener, and Zapf (2011) reported that twelve competency assessment instruments have been developed over the past forty years; however, traditional assessment instruments, generally designed to measure broad psychological constructs, are more widely accepted and commonly used in competency evaluations. This likely reflects the notion that trial competency is a socially constructed, open, context-specific concept and, therefore, cannot be reduced to a fixed set of psycholegal abilities. The traditional clinical assessment instruments most commonly used for competency are the Minnesota Multiphasic Personality Inventory, the Wechsler Adult Intelligence Scales, and the Brief Psychiatric Rating Scale (Pirelli, Gottdiener, & Zapf, 2011).

In an attempt to better understand test usage patterns, Borum and Grisso (1995) surveyed forensic psychologists and psychiatrists regarding their use of psychological

testing in competency to stand trial evaluations. These professionals were board certified or had at least five years of pertinent experience. Although their findings were certainly informative, their survey lacked precision, and the conclusions that can be drawn are limited. For example, asking participants to provide “never, rarely, sometimes, frequently, or almost always” ratings allows for validity concerns.

Nevertheless, approximately half of evaluators considered psychological testing essential or recommended. Consistent with this, about half reported using tests frequently or almost always, while about half viewed it as optional. Ninety percent of psychologists and 85% of psychiatrists reported using objective personality inventories, intellectual testing, or both, making them the most popular types of instruments. Over a third of evaluators also noted using neuropsychological tests and projective tests at times. Regarding forensic instruments, about one third of psychologists reported almost always using them, while another third reported never using them. Only 11% of psychiatrists usually use forensic tests, while 80% reported never or rarely using them. In sum, these findings are consistent in suggesting that traditional personality tests are more commonly used than competency-specific forensic measures. Furthermore, a substantial proportion of evaluators do not (or rarely) use forensic measures at all.

Competency to stand trial evaluation outcomes. There have been two meta-analyses comparing competent and incompetent defendants. Nicholson and Kugler (1991) reviewed thirty studies, and reported that an average of just over 30% of defendants evaluated were found incompetent to stand trial. The vast majority of the subjects in the studies were men, and fewer than 40% were members of minority groups. The individuals had an average of less than ten years of education, more than half had

never been married, and about two thirds did not have steady employment. Over half of the defendants had previous arrests or convictions, and more than half were currently charged with violent offenses. Almost 40% had previous psychiatric hospitalizations and had received a psychotic disorder diagnosis.

The factors most highly correlated with incompetency were a psychotic diagnosis, poor scores on forensic psycholegal assessment instruments (the Competency Screening Test, Georgia Court Competency Test, or the Competency Assessment Instrument), and symptoms reflecting severe psychopathology (disorientation, delusions, hallucinations, impaired memory, and disturbed behavior). These symptoms “reflect cognitive dysfunction and behavioral dyscontrol that might impair a defendant's understanding of legal proceedings and ability to assist in his or her defense” (Nicholson & Kugler, 1991, p. 360). Incompetent defendants were significantly less likely to have prior legal involvement and were significantly more likely to have at least one previous psychiatric hospitalization. On average, they had lower IQ scores and elevated MMPI scales F (“Infrequency,” assessing random responding, over-reporting, or genuine psychopathology), 5 (“Masculinity/Femininity,” evaluating stereotypical masculine or feminine interests and behaviors), 6 (“Paranoia,” measuring abnormal suspiciousness and sensitivity, as well as possible delusions of persecution or grandeur), and 8 (“Schizophrenia,” indicating bizarre or unusual thinking and behavior, inappropriate affect, and possible hallucinations or delusions).

A number of demographic characteristics are worth noting as well. Defendants judged to be incompetent were more likely to be older, female, of a minority group, and not married. There were no significant effects for educational achievement or

employment status. Finally, the correlation of a diagnosis of mental retardation and competency status was not statistically significant, and neither was violent/ nonviolent offense type and competency (Nicholson & Kugler, 1991).

In the most recent meta-analysis, Pirelli, Gottdiener, and Zapf (2011) examined 68 studies published between 1967 and 2008 comparing competent and incompetent defendants. Consistent with previous findings, the base rate of incompetency was 27.5%. The most prominent findings were that defendants with a psychotic disorder diagnosis were eight times more likely to be deemed incompetent than those without such a diagnosis, and the likelihood of being found incompetent was approximately doubled for defendants with a previous psychiatric hospitalization as well as for defendants who were unemployed. Three MMPI scales were significantly higher among incompetent defendants: F (Infrequency), 6 (Paranoia), and 8 (Schizophrenia). There were small to medium standardized effect sizes for each scale. On average, competent defendants' IQ scores were five to six points higher. Effect sizes were substantially larger (approximately one Cohen's *d*-point) for the association between competency status and competency assessment measures than for the relationship between competency status and traditional measures. This included twelve competency measures and three types of traditional measures most commonly researched in the competency arena: the MMPI/MMPI-2; the Wechsler Abbreviated Scales of Intelligence (WASI), Wechsler Adult Intelligence Scale (WAIS), Wechsler Adult Intelligence Scales-Revised (WAIS-R), Wechsler Adult Intelligence Scales-III (WAIS-III); and the Brief Psychiatric Rating Scale (BPRS). (See Appendix A for brief description of each.)

In this review, it was similarly found that incompetent defendants were slightly older on average, predominantly non-white, and not married. However, in contrast with the earlier meta-analysis, female defendants were about as likely to be found incompetent. Finally, defendants with a current violent criminal charge were 1.25 times more likely to be found competent than those with a current nonviolent charge (Pirelli, Gottdiener, & Zapf, 2011).

Although not included in the meta-analyses, Rosenfield and Wall (1998) reported analogous findings regarding a sample of 138 pretrial criminal defendants. A significant correlation was demonstrated between incompetency determinations and symptoms of disorientation, delusions, and hallucinations, as well as higher ratings of thought disorder and paranoia. In fact, they concluded that the presence of such symptoms had greater predictive value than diagnostic categories. For patients with a psychotic disorder, competency determinations were best predicted by symptoms of thought disorder and delusional beliefs. In contrast, among patients without a psychotic disorder, disorientation, mania, hallucinations, and prior judgments of incompetence were the best predictors of incompetence.

Various other findings are relevant as well. Rates of psychosis among incompetent and competent defendants have been reported to be 88% and 39%, respectively (Stafford & Wygant, 2005), and other researchers similarly found 78% of incompetent and 37% of competent defendants to be psychotic (Johnson et al., 1990). Additionally, defendants with a previous history of psychiatric hospitalization have been found to be twice as likely to be deemed incompetent as those without (Melton et al., 2007). A diagnosis of schizophrenia specifically is also highly associated with

incompetence (Hart & Hare, 1992; Rogers, Gillis, McMain, Dickens, 1988; Warren, Fitch, Dietz, & Rosenfield, 1991).

Regarding other mental diagnoses, Stafford and Wygant (2005) observed that defendants judged to be incompetent were significantly less likely to have a personality disorder diagnosis. Conversely, approximately half of defendants judged to be competent were diagnosed with a personality disorder. Research has also suggested that defendants with a substance abuse disorder are significantly more likely to be deemed competent (Rosenfield & Wall, 1998; Rogers, Gillis, McMain, Dickens, 1988; Cooper & Zapf, 2003). Stafford and Wygant's results similarly revealed that 40% of competent defendants, compared to 20% of incompetent defendants, were diagnosed with substance abuse problems.

Conclusions about mood disorders seem to vary. Rogers, Gillis, McMain, and Dickens (1988) observed that there were substantially more defendants with major mood disorders judged to be incompetent (37%) than those deemed competent (14%). Although Warren, Fitch, Dietz, and Rosenfield's (1991) sample demonstrated a lower rate of mood disorders (9% of incompetent cases), this diagnostic category was still much greater among incompetent defendants. Other researchers found that there was not a significant relationship between competency status and depressive symptoms (Rosenfield & Wall, 1998).

Taken together, the most prominent finding is the clear association between psychosis and incompetency status. While this may not be surprising, the abundant extent to which this was found is striking. In a similar vein, incompetent individuals are more likely to be unemployed and to have a history of hospitalizations. Additionally, substance

abuse and personality disorders seem to serve as protective factors, as those groups are more commonly found competent to stand trial.

Restorability among incompetent defendants. Regarding competency restoration, studies reveal that rates are generally high. For example, it has been estimated that 75-90% of individuals are restored in approximately six months of inpatient treatment (Zapf & Roesch, 2011). Factors correlating with successful restoration among incompetent defendants are substance abuse and personality disorders (Morris & DeYoung, 2012; Colwell & Giancesini, 2011) as well as mood disorder diagnoses (Morris & Parker, 2008, as cited in Morris & DeYoung, 2012). Psycholegal abilities are predictive of successful restoration on a continuum, with the factors of factual legal understanding, basic behavior and outlook, and rational attorney assistance progressively demonstrating success in competency restoration (Morris & DeYoung, 2012).

Research suggests that two types of incompetent defendants are unlikely to be restored: chronically psychotic defendants with lengthy histories of inpatient hospitalizations, and defendants with irremediable cognitive disorders, such as mental retardation, severe dementia and brain injury (Mossman, 2007; Wolber, 2008; Colwell & Giancesini, 2010). Other related characteristics that have been associated with a lower probability of restoration include older age (Mossman, 2007), being prescribed more medications, having lower Global Assessment of Functioning scores, and having a less serious charge (Colwell & Giancesini, 2011; Mossman, 2007).

Accuracy of competency to stand trial evaluations. It is difficult, for various reasons, to assess the validity of competency decisions. Judges rarely challenge clinicians' opinions, and there is no independent criterion with which to compare such

opinions (Melton et al., 2007). It is impossible to fully capture the predictive validity because only defendants deemed competent proceed in court; incompetent defendants' judicial proceedings are suspended, and they are referred for treatment. Researchers have noted that the "openness" of the construct of competency causes substantial confusion among mental health and legal professionals. As such, competency evaluations are stated to be inefficient since the methods for arriving at competency decisions vary so greatly (Johnson, Nicholson, & Service, 1990). Assessing concurrent validity is also challenging, because measures such as intelligence and psychopathology are adopted on a functional, case-by-case basis (Roesch, Zapf, Golding, & Skeem, 1999).

Skeem, Golding, Cohn, and Berge (1998) analyzed community examiners' reports on competency to stand trial by having expert raters code 100 randomly selected competency to stand trial reports. It was concluded that examiners' reports often failed to incorporate critical issues such as a defendant's higher order decisional capacities, which entail cognitive tasks involved in understanding and rationally choosing legal options. Instead, examiners operationalized competency in terms of minimal abilities defendants must possess: the defendant's appreciation of charges, understanding court personnel roles; and, ability to disclose information with counsel. In addition, although evaluators provided reasoning to substantiate their clinical conclusions, they did not present a link between deficits in competency abilities and symptoms of psychopathology. Most reports included intelligence and personality measures, yet less than 30% related the test results to the defendant's competence.

Despite these findings, Skeem and colleagues also concluded that examiners generally agree on defendants' global competence. In line with this, Roesch and Golding

(1980) reported that evaluators agreed in at least 80% of cases. More recent studies show a decrease; Golwensmith, Murrie, and Boccaccini (2012) found that evaluators agreed unanimously on competency opinions in 70.9% of initial cases. Additionally, among individuals initially found incompetent, evaluators re-assessing them for competency at a later time only agreed in 61% of cases (Golwensmith, Murrie, & Boccaccini, 2012). Studies have shown that evaluators tend to have considerably divergent bases for such opinions, and agreement on specific psycholegal deficits averages only about 25%. Although this has not been extensively or recently explored, a large proportion of the differences between examiners regarded aspects of psychopathology (Skeem, Golding, Cohn, & Berge, 1998).

Base-rate decisions alone allow for high levels of agreement with examining clinicians, as approximately 72% (Pirelli, Gottdiener, & Zapf, 2011) to 80% (Roesch, Zapf, Golding, & Skeem, 1999) of all referred defendants are decidedly competent. Furthermore, practicing clinicians featured in competency decision studies often worked and trained together, likely contributing to their high rates of agreement. At times, institutional standards of practice were identical, interview content was identical, and evaluators even consulted with each other to ensure that they arrived at the same conclusion. It stands to reason that independent clinicians would be likely to agree less often (Gowensmith et al., 2012), particularly in ambiguous cases.

The 70-80% rate of agreement on competency decisions is unimpressive, even problematic, given the high base rate of competency determinations and the lack of uniform reasoning behind such outcomes. The fact that there is no standard protocol for competency evaluations suggests a need for, at a minimum, published information

regarding specific practices that are common and corresponding reasoning. Borum and Grisso's (1995) survey reporting the psychological tests used in criminal forensic evaluations is one of the few articles to present findings detailing specific practices of competency evaluators. One conclusion was that approximately half of the competency to stand trial evaluators considered psychological testing to be "optional," which the authors defined in the survey as: "inclusion of this information would not affect the overall quality of the report" (p. 466). Furthermore, limited findings available suggest that evaluators usually do not agree on the specific psycholegal deficits when they are present. It is, therefore, unsurprising that other researchers have found that evaluators differ considerably in their bases for competency opinions (Golwensmith, Murrie, & Boccaccini, 2012). This presents a threat to the validity of competency decisions, and furthermore, it suggests that evaluators are usually unable to give useful feedback regarding areas to focus on in helping individuals to regain competency.

Competency to stand trial decisions are definitive, dichotomous legal determinations, despite the fact that the psychological factors influencing them are inherently ambiguous and multifaceted. Furthermore, competency to stand trial evaluations lack standardization in assessment measures, reported findings, and rationales; and the validity of corresponding decisions can only be speculated. There are numerous assessment tools designed to measure psycholegal abilities and broad psychological constructs, and the appropriateness of a particular tool for a single evaluation undoubtedly depends on the specific factors regarding the defendant and related circumstances. However, it is unfortunate that guidelines regarding how to navigate such factors have not been established.

Although there is a great deal of variation in competency to stand trial evaluations, a number of findings regarding common measures and outcomes have been presented. Both traditional assessment measures and competency-specific tools are used, and despite the suggested superiority of competency instruments, traditional measures remain more popular. Among the most robust findings concerning evaluated defendants is that incompetent defendants tend to have psychotic diagnoses and corresponding symptoms. Just over 70% of defendants evaluated are deemed competent, and the vast majority of incompetent defendants are restorable within approximately six months. Finally, although it is difficult to judge the accuracy of evaluation outcomes, it is clear that there is much room for improvement. The focus will now turn to the MMPI. As shall be demonstrated, the present study will involve using findings from the MMPI-2 and MMPI-2-RF to shed light on their clinical utility.

Minnesota Multiphasic Personality Inventory (MMPI)

Personality assessments became popular during World War I as a way to screen large numbers of individuals. R. S. Woodworth developed the first personality questionnaire in 1920, the Personal Data Sheet, which sought to screen out “unfit” soldiers (Butcher, 2005). This self-rating scale for detecting neurotic individuals consisted of 116 “yes” or “no” questions. Items were selected based on whether Woodworth thought that they assessed psychological maladjustment; no empirical or theoretical basis was used (Greene, 2000).

Personality assessments continued to be developed after the war, and they were generally constructed on a rational basis. For example, the Bernreuter Personality Inventory was developed to measure neuroticism, dominance, introversion, and self-

sufficiency. Bernreuter used clinical experience to select items that seemed to tap a trait or construct (Bernreuter, 1933, as cited in Greene, 2000). However, investigators soon demonstrated that the scale was inadequate in correctly classifying neurotic, psychotic, and normal individuals. Subsequently, J. Page and colleagues constructed a rationally based personality scale. One hundred commonly accepted symptoms of schizophrenia were selected from psychiatric literature, and fifty schizophrenic traits were combined into a questionnaire. After being administered to a group of schizophrenic patients, manic-depressive patients, and normal individuals, the test was found to be inadequate in differentiating groups (Page et al., 1934, as cited in Greene, 2000). It became apparent that early rationally developed personality inventories were generally unsuccessful in appraising the characteristics they attempted to assess.

Starke Hathaway and J. C. McKinley sought to address these problems by developing a measure that would validly and objectively portray patients. Their goal was to develop a single inventory that included a large pool of items from which a variety of personality scales could be constructed. They began creating the Minnesota Multiphasic Personality Inventory in 1940. They selected over 1,000 items from psychiatric examination direction forms, textbooks from psychiatry, earlier personality inventories, and clinical experience. Such items included symptoms, attitudes, beliefs, problems, etc. (Butcher, 2005). They reduced this number to 504 by deleting duplicate items and items considered less significant (Greene, 2000).

Twenty-five headings were then arbitrarily picked to classify the items, and a series of quantitative scales were constructed to assess the various categories of psychopathology (Hathaway & McKinley, 1940, as cited in Greene, 2000). Using an

empirical criterion keying approach, items for each scale were selected based on how well they differentiated a clinical (criterion) group from a group of normal individuals (Butcher, 2005). Those items that consistently differentiated the clinical sample from the normative group were selected for the scale, even if the content did not appear to be related to the clinical syndrome. The rationale behind this was that when a test-taker's pattern of responding results in an elevated clinical scale score, inferences can be made about the test-taker's behavior typical of the groups used to derive these scales (Ben-Porath, 2012).

The clinical groups were chosen based on standard clinical criteria at the time, which included hypochondriacal, depressed, hysteroid, sociopathic, paranoid, psychasthenic, schizophrenic, and hypomanic patients. In order for scales to be effective in personality assessment, criterion groups needed to contain highly similar cases. As such, homogenous, "pure" cases maximizing diagnostic similarity were chosen to increase the likelihood of validly discriminating other cases. The normative sample consisted of 724 individuals who were visitors of patients in the University Hospitals in Minneapolis, and they were considered to be a good estimate of people in general at the time. The typical individual in the Minnesota normative group was Caucasian, approximately thirty-five years old, married, living in a small town, with eight years of general schooling, and work as a skilled or semiskilled operative (Greene, 2000). However, by current standards, the sample was relatively small and overly narrow (Butcher, 2005). Furthermore, the visitors were primarily friends and relatives of the patients (Greene, 2000). Given that there is often a genetic component related to mental

illnesses, the original normative sample is considered to be highly problematic in certain regards.

Ten clinical scales were established to measure different dimensions of personality. Scale 1, “Hs” (Hypochondriasis) reflects abnormal concern over bodily functions and preoccupation with physical complaints. Scale 2, “D” (Depression) reveals a pessimistic world-view, feelings of hopelessness, and self-depreciation. Scale 3, “Hy” (Hysteria) measures the use of physical or mental symptoms to avoid stressful conflicts. Scale 4, “Pd” (Psychopathic Deviate) uncovers a tendency toward conflict with authority, disregard of social conventions and laws, and shallowness in personal attachments. Scale 5, “MF” (Masculinity-Femininity) differentiates a tendency toward traditional masculine or feminine interests, attitudes, and self-expression. Scale 6, “Pa” (Paranoia) indicates abnormal suspiciousness and sensitivity, and possible delusions of persecution or grandeur. Scale 7, “Pt” (Psychasthenia) reflects a tendency for obsessive rumination, guilty feelings, anxiety, and compulsive ritualistic behavior. Scale 8 “Sc” (Schizophrenia) measures bizarre or unusual thinking and behavior, interpersonal withdrawal and alienation, inappropriate affect, and possible hallucinations or delusions. Scale 9, “Ma” (Hypomania) reveals high activity level without productivity, emotional agitation, euphoria, and flight of ideas. Scale 10, “Si” (Social Introversion) indicates shyness, social withdrawal and insecurity, and disinterest in others.

Four validity scales were developed to capture test-taking attitudes that may influence the validity of the clinical scale scores. “?” (Cannot Say) measures non-responding, “L” (Lie) detects under-reporting, “F” (Frequency or Infrequency) assesses

random responding, over-reporting, or genuine psychopathology), and “K” (Correction) measures more subtle under-reporting (Ben-Porath, 2012).

Studies soon demonstrated that the Clinical Scales had poor validity in predicting diagnostic group membership (Hathaway, 1960, as cited in Ben-Porath, 2012). However, patterns or combinations of scores could be used to differentiate diagnoses. Numerical coding systems, or code types, that designated different classes of profiles began to be used to describe normal and abnormal personality characteristics (Ben-Porath, 2012). In addition to code types, several basic characteristics are considered in interpreting an MMPI profile. These include the overall elevation and “phasicity” of the profile, as well as the overall profile slope (Megargee & Bohn, 1979). To further assist in interpretation, Harris and Lingo developed subscales for the Clinical Scales, focusing on grouping together items of specific content domains (1955, as cited in Archer, 2006).

The MMPI-2

Over time, updated norms were needed to represent the changing population in the United States. This included a larger normative sample that was nationally representative and that contained a proportionate representation of ethnic minorities (Greene, 2000). In addition, a number of MMPI items were considered problematic because the language was dated, and item content was no longer relevant, clear, or appropriate for assessing psychopathology and personality. For example, the meaning of a “gay party” had changed considerably over the past 60 years. Various items were also not scored on any of the Clinical, Validity, or Supplementary Scales, and more items were needed to assess contemporary clinical personality concerns (such as drug usage, suicidal ideation, and work-related difficulties; Ben-Porath, 2012). However, researchers

did not want to develop new criterion groups and item derivation procedures (Greene, 2000). The Restandardization Project began in 1982, aiming to resolve these issues. In order to allow for continued use of empirical correlates of code types in interpretation, the goal was to maintain as much continuity with the original MMPI as possible (Ben-Porath, 2012).

The resulting 567-item MMPI-2 was published in 1989. During the redevelopment research, 15,000 individuals were involved in testing to update the normal and clinical groups. The MMPI item pool was modernized by rewriting awkward wording or outdated expressions, and unused items were removed. The item pool was also expanded to assess a broader range of clinical symptoms and problems. Of the 383 items on the basic Validity and Clinical Scales, 372 were retained, 11 were deleted, and 64 items were slightly revised (Archer, 2006). A nationally representative normative sample was developed, including 1,138 men and 1,462 women of diverse ethnic backgrounds and from across five regions of the United States. Several clinical samples were collected to validate diverse patient and nonpatient groups, including inpatient psychiatric cases, alcohol and drug abusers, prison inmates, women judged as being at risk for child abuse, alcohol and drug abusers, and patients with chronic pain. “Normal range” groups also included airline pilot applicants, undergraduate college students, military personnel, and older men (Butcher, 2005).

To provide a more comprehensive assessment of invalidating conditions, three new MMPI-2 validity scales were developed. The Infrequency Back Scale (F_B) was designed to detect changes in the response pattern in the latter part of the test, and it is made up of items that were endorsed infrequently by the normative sample. The Variable

Response Inconsistency (VRIN) aims to measure quasi-random responding by detecting the pattern of responses to nearly identical or opposite pairs. Similarly, the True Response Inconsistency (TRIN) was designed to identify fixed responding (true or false) by considering pairs of responses, all opposite in meaning (Ben-Porath, 2012).

MMPI-2 Content Scales were developed through a series of rational-conceptual and empirical analyses. They facilitate test interpretation by expanding on the content domains represented by the original Clinical Scales and by providing an indication of the examinee's self-presentation. Fifteen Content Scales were produced, each containing a number of items that assessed a unitary personality characteristic (Ben-Porath, 2012). They include Anxiety (ANX), Fears (FRS), Obsessiveness (OBS), Depression (DEP), Health Concerns (HEA), Bizarre Mentation (BIZ), Anger (ANG), Cynicism (CYN), Antisocial Practices (ASP), Type A Behavior (TPA), Low Self-Esteem (LSE), Social Discomfort (SOD), Family Problems (FAM), Work Interference (WRK), and Negative Treatment Indicators (TRT) (Archer, 2006). Finally, the Supplementary Scales were developed to assess various clinical or personality problems that were not focused upon in the original scales. These include the Anxiety (A), Repression (R), Ego Strength (Es), Dominance (Do), Social Responsibility (Re), College Maladjustment and Marital Distress (MDS), Hostility (Ho), Over-controlled Hostility (O-H), MacAndrews Alcoholism Scale- Revised (MAC-R), the Addiction Admission Scale (AAS), and the Addiction Potential Scale (APS) (Archer, 2006).

After the MMPI-2 was published, over 800 journal articles, 70 book chapters, 20 books, and 360 doctoral dissertations on the subject were published just in the following decade. Subsequently, in 2001, a revised MMPI-2 manual was produced. Although there

were a number of changes, no new norms were used and the item composition remained intact. A number of subscales were dropped, and new Validity Scales were also introduced. The Infrequent Psychopathology (F_P) aims to detect over reporting in settings with high base rates of significant psychopathology. The Superlative Self-Presentation (S) is an underreporting measure that adds incremental validity to L and K scales. Finally, the Symptoms Validity Scale (FBS) aims to detect somatic and cognitive malingering (Ben-Porath, 2012).

The Personality Psychopathology Five (PSY-5) Scales were included in the revised MMPI-2 manual, aiming to provide a dimensional measurement model of Axis II disorders. A series of data-reduction techniques were used to converge upon a model of five underlying factors: Aggressiveness, Psychoticism, Constraint, Negative Emotionality/Neuroticism, and Positive Emotionality/Extraversion. Harkness, McNulty, and Ben-Porath (1995) then developed these scales using what they called replicated rational selection. They had lay judges select MMPI-2 items that seemed pertinent to each of the five dimensions, and items most commonly selected by the lay judges were assigned to the PSY-5 Scales. Next, Harkness and McNulty provided expert review and eliminated certain items that did not conform to the construct they intended to measure. Finally, item analyses were conducted to increase discriminant validity, and items that were highly correlated to different scales were dropped.

The MMPI-2 Restructured Clinical (RC) Scales were added to the MMPI-2 in 2003. The Clinical Scales had been created to identify psychopathological syndromes, which often shared a fundamental factor of distress. As a result, the Clinical Scales contained excessive intercorrelations and substantial heterogeneity (Ben-Porath, 2012;

see Appendix Table B4 for a table of MMPI-2 intercorrelations). The RC Scales sought to remove this common distress, which was coined “demoralization.” This allowed for a more pure measure of the scale components, thereby improving discriminant and convergent validity. As such, the unique “seed” or core components of each Clinical Scale were identified for the RC Scales. For example, after removing the distress associated with Clinical Scale 8 (Schizophrenia), the core component was identified and renamed Aberrant Experiences. The final set of RC Scales included 192 non-overlapping items scored in the following scales: Demoralization (RCd), Somatic Complaints (RC1), Low Positive Emotional (RC2), Cynicism (RC3), Antisocial Behavior (RC4), Ideas of Persecution (RC6), Dysfunctional Negative Emotions (RC7), Aberrant Experiences (RC8), and Hypomanic Activation (RC9; Archer, 2006; see Appendix Tables B1 and B2 for a brief description of Clinical Scales and RC Scales).

The MMPI-2 is currently the most widely used and researched personality assessment tool (Butcher, 2005). Ben-Porath (2012) reported that as of January 2011, there were approximately 2,000 published empirical studies that included the MMPI-2. The vast research literature supports the convergent validity and effectiveness of the scales in detecting a variety of clinical problems. In Archer’s (2006) review, it was concluded that the convergent validity of the Clinical, Content, Supplementary, and PSY-5 Scales and codetypes has been demonstrated across settings, including inpatient, outpatient, forensic, college student, and private practice samples. Additionally, the internal consistencies in the normative sample for the Clinical Scales range from .34 to .87, for the Content Scales from .68 to .86, for the Supplementary Scales from .24 to .90, for the PSY-5 Scales from .65 to .84, and for the RC Scales from .62 to .89 (Butcher et

al., 2001; Tellegen & Ben-Porath, 2008). Although the discriminant validity of the MMPI-2 Clinical Scales has been problematic, the RC scales have greatly improved this. In fact, the RC scales have demonstrated superior discriminant validity compared to the other MMPI-2 scales that were proxies for the RC scales (Tellegen, Ben-Porath, & Sellbom, 2009). The reliability of the RC scales has also been exhibited, with test-retest correlations ranging from .62 to .88 (Butcher, Graham, Ben-Porath, Tellegen, Dahlstrom, & Kaemmer, 2001).

Many research studies also support the use of the Validity Scales in detecting over- and underreporting. For example, in Rogers, Sewell, Martin, and Vitacco's (2003) meta-analysis of the MMPI-2 overreporting scales, it was concluded that the instrument's F, F_B, and F_P scales were effective in detecting malingering. Furthermore, F_P had the largest effect size in differentiating malingerers from non-malingerers. Another meta-analysis found that the L scale was the best indicator of underreporting, and K was noted to be effective as well (Baer & Miller, 2002).

The MMPI-2-RF

The next edition of the MMPI was the MMPI-2 Restructured Form (MMPI-2-RF), created by Tellegen and Ben-Porath in 2008. They sought to make a more efficient form with enhanced construct validity; it contains only 338 items. It is related conceptually and empirically to modern theories of personality psychopathology, and it is comprised of a hierarchical structure with a total of 50 scales. The core components are the RC scales described previously. However, the RC scales were not intended to assess all the information available in the MMPI-2 pool, and so substantive scales were developed to complete the MMPI-2RF.

A series of factor analyses of the RC Scales led to three broad-band dimensions and the construction of Higher-Order Scales (H-O) to assess them: Emotional/Internalizing Dysfunction (EID), which assesses problems associated with mood and affect; Thought Dysfunction (THD), which measures problems associated with disordered thinking; and Behavioral/ Externalizing Dysfunction (BXD), which gauges problems associated with under-controlled behavior. The Specific Problem (SP) scales are the most narrow and are grouped into four domains—Somatic/Cognitive, Internalizing, Externalizing, and Interpersonal. Additional scales on the MMPI-2-RF include the revised personality model (PSY-5R) and Interest scales. Finally, the MMPI-2-RF contains seven revised and shorter Validity Scales: VRIN-r, TRIN-r, F-r, Fp-r, FBS-r, L-r, and K-r. There is also a new Validity Scale, the Infrequent Somatic Responses (Fs), designed to measure infrequent somatic complaints (Tellegen & Ben-Porath, 2008; see Appendix Table B3 for a list and brief description of the MMPI-2-RF Validity Scales).

Tellegen and Ben-Porath (2008) reported extensive psychometric findings supporting the reliability and validity of the measures. For example, in one study it was concluded that F-r, Fp-r, F-s, and FBS-r are all effective at detecting various threats to protocol validity, including feigned psychiatric, somatic, and neurocognitive symptoms, in the context of civil and criminal forensic litigation (Wygant, Selbom, Ben-Porath, Stafford, Freeman, & Heilbronner, 2008). Tellegen and Ben-Porath also cited a study by Selbom and Bagby, where the L-r and K-r were found to be effective at identifying simulating test takers as well as test takers in real-life settings expected to under-report in the context of their evaluation. Further, they reported that the two measures complement each other, providing incremental validity. Finally, in a meta-analysis by Rogers, Sewell,

Martin, and Vitacco (2003) the Fp Scale was found to be the strongest MMPI-2 over-reporting indicator of psychopathology.

As mentioned earlier, the RC scales that form the core of the MMPI-2-RF have demonstrated good psychometric properties. Further, comprehensive reliability and validity analyses have revealed that compared to the Clinical Scales, “the considerably shorter RC scales were found to be about equally or more reliable, to be less saturated overall with Demoralization, to be less highly intercorrelated, and to achieve comparable to improved convergent validities and substantially improved discriminant validities” (p. 8, Tellegen, Ben-Porath, McNulty, Arbisi, Graham, & Kaemmer, 2003, as cited in Tellegen & Ben-Porath, 2008). MMPI-2-RF as well as recent empirical research have provided positive convergent and divergent validity findings for the scales, particularly in differential diagnosis. For example, Selbom, Toomey, Wygant, Kucharski, and Duncan (2011) examined the MMPI-2-RF responses of various diagnostic groups. Multivariate and logistic regression analyses of their test results revealed differences among groups that were largely consistent with each group’s unique diagnostic characteristics. Other research has evinced good convergent validity for the MMPI-2-RF Specific Problems and Interest scales, reflected in large effect size correlations with criterion measure scores (Forbey, Lee, & Handel, 2010).

Lanyon and Thomas (2013) explored the validity of the three H-O scales among various relevant groups, including psychiatric inpatients, incarcerated prisoners, and pretrial criminal defendants. The comparison group was the MMPI-2-RF normative group. The results demonstrated construct validity, as there were medium to large effect sizes for the correlations between each scale and its relevant participant group. The

Thought Dysfunction scale showed a large effect size for criminal defendants, whereas the Externalizing/ Internalizing Dysfunction scale demonstrated a large effect size for psychiatric inpatients. The Behavioral/ Externalizing Dysfunction scale demonstrated the largest effect size for prison inmates, followed by criminal defendants.

Functional Utility of the MMPI-2/ MMPI-2-RF With Forensic Populations

The MMPI-2 has been used extensively in forensic settings; in fact, various surveys have concluded that the MMPI-2 is administered more frequently than any other test in forensic settings, and it is considered appropriate for most forensic questions (Sellbom & Ben-Porath, 2006). It is used in generating diagnostic impressions, treatment planning, and to inform evaluative decisions. Although it is still relatively new, the MMPI-2-RF is also gaining popularity in forensic evaluations. Similar to the MMPI-2, the MMPI-2-RF meets the criteria for the *Daubert* (1993) standard of scientific techniques admissible in court. An abundance of empirical correlates are available among criminal pre-trial defendant samples, and Ben-Porath (2013) reported that as of December 2012, 160 peer-reviewed journal articles had already been published on the MMPI-2-RF. The potential rate of error is known, as data on reliability and standard error of measurement scale scores are reported in the Technical manual. Finally, there are standard interpretive guidelines enhancing cross-interpreter reliability (Ben-Porath, 2013).

Scores on the MMPI-2-RF have been correlated with many indicators for large samples of men and women undergoing forensic assessments related to competency to stand trial evaluations, as the test is intended to be used with these populations (Tellegen & Ben-Porath, 2008). Empirical correlates reported in the manual are based on clinician

surveys and rating forms, detailed chart data, and self-report measures. They include information such as juvenile conduct problems, substance abuse, psychotic symptoms, violence, abuse, mood instability, and criminal history, among other data. However, such information is not divided among incompetent and competent defendants. Ben-Porath (2013) also cited 16 studies investigating the utility of MMPI-2-RF over-reporting indicators with forensic samples, as well as studies with civil litigants, individuals undergoing parental competency evaluations, and the assessment of psychopathy.

Although the MMPI-2-RF does not provide direct answers to psycholegal questions, it can provide relevant information such as an individual's psychological functioning (or dysfunction), personality characteristics, and behavioral tendencies that are at issue in forensic evaluations. Moreover, how individuals approach the assessment, and whether they attempt to over- or under-report psychological problems is particularly relevant (Ben-Porath, 2013). Various studies have noted that malingering is quite common in forensic evaluations. For example, Ardolf, Denney, and Houston (2007) estimated that 54% of individuals evaluated for neuropsychological deficits in forensic populations fall in the probably malingering range or higher. As discussed prior, research supports the utility of the MMPI-2-RF in detecting such disingenuous response patterns.

MMPI in Classification and Prediction Among Forensic Populations

Given the heterogeneity of forensic populations, researchers have used the MMPI to classify offenders since its inception. Meaningful categorization can allow for increased understanding regarding differences in behaviors, outcomes, and risk assessments, for example. One such classification system that was particularly successful was created by Megargee and colleagues. They used hierarchical analyses to develop 10

MMPI “types” to distinguish criminals (Megargee & Bohn, 1979). This was done with a sample of 1344 males incarcerated at a Tallahassee federal medium security prison. In addition to administering the MMPI, the prisoners completed a battery of other psychological tests as well as a structured clinical interview that was tape-recorded and scored by independent raters. Characteristics of a typical member of each group were established for each of the 10 types, which differed in their family and social history, lifestyle and personality patterns, prison adjustment, and recidivism.

Zager (1988) reviewed the classification system’s reliability, discussing five studies examining the inter-rater reliability, seven examining the test-retest reliability, and three examining the replicability. Overall, strong reliability of the system was exhibited across studies. Successful replication of the Megargee system has also been extended to include forensic psychiatric patients (Wrobel, Wrobel, & McIntosh, 1988). Regarding predictive validity, the classification system proved to be effective in distinguishing potentially violent and non-violent prisoners. When the Tallahassee Federal Correctional Institution used it to assign inmates their living arrangements, serious assaults dropped by 46% (Bohn, 1979). Zager’s review also concluded that the system has demonstrated value in predicting the success of the different groups in work release programs and for predicting institutional adjustment among some state offenders.

In a different study, researchers attempted to utilize MMPI profiles to meaningfully differentiate criminals by offense type. Investigators compared murder or attempted murder of a family member or girlfriend, murder or attempted murder of an unrelated victim, rape, arson, child molesting, and property. Each offense group consisted of 25 men, and they had all been remanded by the courts to a maximum-security

psychiatric hospital for pretrial assessment. The researchers hypothesized that scores on a supplemental scale, Overcontrolled Hostility, would be highest in the family murder group. This hypothesis was not supported; rather, the profiles between groups were notably similar. Elevated MMPI mean scores on scales 2, 4, 7, and 8 indicated that there were relatively high levels of psychopathology across all of the groups. The researchers then compared the sample's profiles with data from six other studies that included forensic assessment cases, personality disordered individuals, and drug abusers. They found that the 4-8-2 profile prominent in their study was also fairly common in the peer-reviewed literature. It was noted, however, that the profiles of psychotic patients were less similar (Quinsey, Arnold, & Pruesse, 1980). In a more recent example, a team of investigators used a replicable cluster analytic method to establish an MMPI-2 based categorical classification system with chronically ill forensic psychiatric outpatients. Augmented by Rorschach inkblot data scored using Exner's Comprehensive System, the result was a seven-cluster solution. The cluster types demonstrated meaningful differences across several characteristics and test results (Nieberding et al., 2003).

Various teams have sought to derive MMPI profile subgroups among sex offenders. In one such study, researchers used a multivariate clustering procedure for classification. Three profile subgroups emerged: one with significant elevations on scales F and 8, another with elevations on scales 4 and 9, and a third with elevations on scales 1, 2, 3, and 4. Although child and adult offenders were distributed evenly among them, behavioral indices were associated with the various subgroups (Anderson, Kuncze, & Rich, 1979). Another team found five MMPI profile subgroups among incarcerated male rapists. The subgroups ranged in psychopathology from slight to extreme, and they also

differed by circumstances of the crime as well as responses to the Multiphasic Sex Inventory (Kalichman, Szymanowski, McKee, Taylor, & Craig, 1989). Finally, significant differences in MMPI scores were found among sex offender groups (sexual abuse, rape, and sodomy) and inpatients without a history of sexual offending; sex offender groups demonstrated significantly more psychopathology. Differences were also found among sex offender groups with single-scale elevations and 2-point code types. Generally, increased psychopathology was associated with increased sexual deviancy (Herkov, Gynther, Thomas, & Myers, 1996).

The MMPI has also been used to predict treatment participation and outcomes. Kalichman, Shealy, and Craig (1990) found a significant relationship between MMPI profile subgroup membership and treatment participation among men convicted of rape, with several MMPI scales correlating with clinicians' ratings of the men's participation. In a different sample of sex offenders offered treatment at a medium-security state prison, logistic regression analysis revealed that lower VRIN and Mf scores as well as higher L scores significantly predicted treatment refusal (Clegg, Fremouw, Horacek, Cole, & Schwartz, 2011).

The MMPI-2 RC scales have also been shown to predict treatment failure and recidivism. Selbom, Ben-Porath, Baum, Erez, and Gregory (2008) studied a sample of offenders enrolled in a batterers' intervention program. Using regression analyses, RC scales contributed to predicting treatment failure and recidivism when added to historical variables such as criminal history, substance abuse problems, mental health treatment, anger problems, and amount of partner violence. Relative risk analyses also demonstrated that elevated scores on RC4 and RC9 increased risk for negative outcomes. Additionally,

The MMPI-2 RF was used to predict Drug Court treatment completion in a sample of individuals identified as being at high risk for failure. Higher scores on Behavior/Externalizing Dysfunction, Antisocial Behavior, Aberrant Experiences, Juvenile Conduct Problems, Aggression, and Disconstraint-Revised scales were associated with increased risk for failure to complete treatment. Zero-order correlations and relative risk analyses provided evidence for the MMPI-2-RF's predictive utility (Mattson, Powers, Halfaker, Akesson, & Ben-Porath, 2012).

MMPI and Competency Assessment

As mentioned previously, the MMPI/ MMPI-2 is the most frequently utilized multi-scale instrument in forensic evaluations (Sellbom & Ben-Porath, 2006), and it is the primary instrument used by psychologists and psychiatrists when evaluating competency to stand trial (Borum & Grisso, 1995). Its widespread use is likely rooted in its standardization, empirically established reliability and validity, and the fact that when used properly, it meets the admissibility standards established by the U.S. Supreme Court (Otto, 2002). This is not surprising given the strong relationship between incompetency decisions and psychotic diagnoses and symptoms that has been demonstrated in peer-reviewed literature.

Despite its widespread use in competency evaluations, MMPI, MMPI-2, or MMPI-2-RF data comparing incompetent and competent defendants are fairly limited. In the most recent meta-analysis comparing such populations, ten studies were presented comparing incompetent and competent defendants' MMPI and/or MMPI-2 validity and clinical scale scores. However, only two of these studies (Maxson & Neuringer, 1970; Sachsenmaier, 1991, as cited in Pirelli, Gottdiener, & Zapf, 2011) presented comparative

data of defendants on three scales for which effect sizes were calculable. In both studies, incompetent defendants produced higher scores across scales F, 6, and 8. The standardized mean difference effect sizes for the three scales were 0.33, 0.39, and 0.33, respectively, which are all considered small to medium.

A few other studies that were not included in the meta-analyses reported divergent conclusions about MMPI scales. One research team observed a significant difference between competent and incompetent defendant scores only on Scale 9, with higher elevations among incompetent defendants (Rogers, Gillis, McMMain, & Dickens, 1988). In contrast, Johnson et al. (1990) observed that incompetent defendants scored significantly lower on scale 9 than those judged to be competent. These researchers also found that incompetent defendants scored significantly higher on Scale 2 than competent defendants. Johnson's sample, however, was limited to 111 defendants found competent to stand trial and only nine judged to be incompetent.

One study examined the relation between MMPI-2 Psychoticism scale scores and psycholegal abilities as measured by the MacCAT-CA. Data were collected from three groups of felony defendants: defendants admitted to forensic psychiatric units after being adjudicated incompetent to proceed; defendants in jail receiving treatment for mental health problems but who were presumed competent; and, randomly selected jail inmates who were presumed competent. MMPI-2 Psychoticism scores differed significantly between groups: hospitalized incompetent defendants scored highest with a mean T score of 75; jail inmates receiving mental health services followed with a mean T score of 70; and, unscreened jail inmates had the lowest T score, which averaged 63. As predicted, the findings also revealed that MacCAT-CA measures correlated negatively with the

psychoticism scale. The incompetent defendants scored significantly lower on all three MacCAT-CA psycholegal ability measures: Understanding, Reasoning, and Appreciation (Otto et al., 1998).

Somewhat more research has been conducted to address the possibility of malingering among defendants referred for competency to stand trial evaluations, including the MMPI-2 validity scales, in detecting feigned psychopathology, somatic symptoms, and cognitive symptoms. As mentioned previously, malingering has been shown to be fairly common among defendants whose competency is questionable. After all, successfully feigning symptoms may result in delaying the onset of the trial and receiving hospitalization rather than incarceration (Wygant, Selbom, Ben-Porath, Stafford, Freeman, & Heilbronner, 2007).

Wygant et al. (2007) demonstrated that MMPI-2 FBS and Fp scales correlated with cognitive symptom validity test failure among criminal defendants evaluated for competency to stand trial, criminal responsibility, and drug dependence. Thirty-four of the 127 defendants failed symptom validity tests as determined by their scores on the Test of Memory Malingering or Word Memory Test, and concurrently had elevations on the FBS and Fp over-reporting measures as well as the RC1 somatic complaining scale. Another study reported a much higher rate of malingering; among 55 men undergoing pretrial psychological evaluations, 24 defendants were classified as feigning. This was determined by utilizing the Structured Inventory of Reported Symptoms, and the MMPI-2 F and Fb scales successfully differentiated feigners from honest responders as well. In fact, the MMPI-2 Fb scale best predicted feigning in the sample, and no other variable added significant incremental predictive power (Lewis, Simcox, Berry, 2002).

Finally, in a study utilizing the MMPI-2-RF, 125 men referred for competency to stand trial, criminal responsibility, or aid-in-sentencing evaluations were examined. In this sample, twenty-five men were categorized as probably malingering. The F-r and Fp-r differentiated malingering and non-malingering groups with very large effect sizes. The FBS-r and Fs scales were also associated with respectable effect sizes in differentiating malingering and non-malingering groups, which suggested that criminal forensic populations tend to feign symptoms across psychopathology, neurocognitive, and somatic domains (Sellbom et al., 2010).

Data are sorely lacking regarding the utility of the MMPI-2-RF in differentiating competent and incompetent defendants. However, research has demonstrated that a number of scales are strong, unique predictors in differentiating between pairs of diagnostic groups. For example, differences were revealed among groups of patients with bipolar disorder, major depressive disorder, and schizophrenia. The pattern of mean MMPI-2-RF scale differences among patient groups was generally consistent with the prominent features of each diagnostic group. Furthermore, the H-O scales Emotional/Internalizing Dysfunction and Thought Dysfunction were most useful in differentiating between groups (Selbom, Bagby, Kushner, Quilty, & Ayearst, 2012). Given the significant differences between diagnoses among competent and incompetent defendants that have been found in peer-reviewed literature (Pirelli, Gottdiener, & Zapf, 2011), it is likely that the pattern of mean MMPI-2-RF scale distinctions would extend to defendants.

Having objective information regarding someone's personality structure, psychopathology, and test-taking attitude is highly valuable when making a decision as

grave as whether that person is competent to stand trial. The MMPI-2 and MMPI-2-RF relay such information while offering sound reliability and validity statistics, and their ubiquitous use in forensic settings follows. While findings exist regarding the utility of the MMPI in various forensic domains, and studies have also demonstrated common characteristics among individuals who are incompetent to stand trial, there is currently a gap in the literature regarding the MMPI's utility in competency to stand trial evaluations. This is particularly surprising given the great frequency with which competency to stand trials evaluations are ordered by courts and the frequency of use of the MMPI with forensic populations. Furthermore, some degree of ambiguity is inherent in the current construct of competency, and there is no standard protocol for competency evaluations. Ensuring that such decisions are determined with as much relevant, impartial data as possible is essential in order to uphold fairness and accuracy within the legal system.

The literature that has been discussed leaves much to be explored. MMPI/ MMPI-2 scales F, 6, and 8 have been demonstrated to be significantly higher among defendants who are incompetent to stand trial. Other research has demonstrated that Validity Scales effectively differentiate malingerers, although further findings have been limited to single studies or were inconclusive. The lack of data regarding the MMPI-2 and, even more so, the MMPI-2-RF, leaves a great deal of potential for discovering valuable information regarding these tests' clinical utility. This is particularly true given the fact that the MMPI-2 is the most commonly utilized instrument in competency to stand trial evaluations, and it has been used to successfully distinguish diagnostic groups as well as various forensic populations. The present study will build upon the literature that has

been presented by investigating the MMPI-2/ MMPI-2-RF's ability to distinguish competent and incompetent groups and by determining which test, if any, better distinguishes groups.

Hypotheses

The first and second hypotheses to be investigated are whether the MMPI-2 and MMPI-2-RF protocols differ among competent and incompetent to stand trial individuals. First, it is hypothesized that incompetent defendants will score significantly higher on MMPI-2 scales F (Infrequency), 6 (Paranoia), and 8 (Schizophrenia), as this would be consistent with previous research (Pirelli, Gottdiener, & Zapf, 2011). Second, it is similarly hypothesized that incompetent defendants will score higher on MMPI-2-RF Scales F-r (Infrequent Responses), RC6 (Ideas of Persecution), and RC8 (Aberrant Experiences), given that these scales were derived from the significant MMPI-2 scales. Additionally, it is hypothesized that the incompetent defendants will score significantly higher on the MMPI-2-RF higher order Thought Dysfunction (THD) scale. This has not yet been explored, but it would be congruent with the established conclusion that disordered thinking is more prevalent among incompetent defendants (Rosenfield & Wall, 1998). All of these hypotheses would be analogous with the substantial body of literature demonstrating a higher rate of psychotic diagnoses (Pirelli, Gottdiener, & Zapf, 2011; Stafford & Wygant, 2005; Johnson et al., 1990) and symptoms including disorientation, delusions, hallucinations, and paranoia among incompetent versus competent defendants (Rosenfield & Wall, 1998). While other studies have suggested the significance of some additional scales, such findings were not robust or were inconclusive.

The overall significance of the MMPI-2 and the MMPI-2-RF scales will then be compared. The third hypothesis is that the MMPI-2-RF will demonstrate greater effectiveness in discriminating competent and incompetent defendants' profiles. This is likely because the MMPI-2-RF RC scales have a common demoralization factor removed, allowing for the unique core components of each Clinical Scale to be more prominent (Archer, 2006). Additionally, the added predictive power of the THD scale will likely contribute to the MMPI-2-RF's efficacy. It is expected that the correlations between incompetent defendants and the MMPI-2-RF Scales F-r, R6, R8, and THD will create a predictive model that distinguishes incompetent and competent defendants.

CHAPTER 3

METHODS

Participants

Patients

This is an archival study, and MMPI-2 protocols were taken from the medical records of male and female patients, ages 18 and older, admitted to Saint Elizabeths Hospital (SEH) in Washington, D.C. Half of these protocols are from individuals who were court-ordered to receive competency treatment and evaluation. Their records contained psychological evaluations that include an MMPI-2 protocol, and the first 30 encountered were selected. These protocols along with corresponding demographic data were collected and comprise the “incompetent to stand trial” group. The other half of the protocols are from patients who were admitted for court-ordered involuntary treatment after being found Not Guilty by Reason of Insanity (NGRI). These individuals are presumed to be competent, because competency is essential in order for the court hearing to occur. It should be noted that some NGRI individuals may have been incompetent at one time, in which case they likely underwent competency restoration treatment. This group must also have psychological evaluations that include MMPI-2 protocols. The first 30 encountered comprise the “competent to stand trial” group. The 30 records comprising each group were chosen from a random sample of hospital records. The sample was limited to a total of 60 records due to the limited amount of funding available to cover the cost of re-scoring the profiles (see Procedure section).

Setting

Saint Elizabeths Hospital is the District of Columbia's public psychiatric facility for individuals with serious and persistent mental illness who need intensive inpatient care. It houses approximately 400 males and females who are either forensic (criminal) or civilly committed to the hospital for treatment. Forensic patients are those who are mandated by the court to receive competency restoration treatment and evaluation or are adjudicated be criminally insane (NGRI). Civil patients are those who are admitted due to an acute need for psychiatric care.

Procedure

The psychology records of patients admitted to SEH were reviewed, and those with completed MMPI-2 protocols were selected. The first 30 reviewed who were from individuals admitted for competency to stand trial evaluation as well as the first 30 who were admitted after being found NGRI and court-ordered for treatment were selected. Data collected included demographic variables (age, gender, race, marital status, employment status) and psychological variables (psychiatric diagnosis, history of previous hospitalization, length of current hospitalization, and psychological test data). These data were chosen because research has suggested that such variables have predictive value in competency to stand trial determinations (Pirelli, Gottdiener, & Zapf, 2011). If demographic characteristics varied significantly across competent and incompetent groups and there was not a clinical rationale to account for such differences, attempts were made to gather additional data to homogenize groups.

After selection, the principal investigator made physical copies of the MMPI-2 protocols (i.e. multiple choice responses and scale configurations) and redacted them on

site to de-identify participants. Each participant's record was given a unique identifier number by the principal investigator, and electronic data were securely maintained in a password protected format on the investigator's computer. Such data were redacted of all identifying information to ensure confidentiality and anonymity. MMPI-2 raw data were re-scored using the Q Local software to produce MMPI-2-RF score reports at the Center for Applied Psychology at Indiana University of Pennsylvania. The Graduate Student Research Grant awarded by the School of Graduate Studies and Research at Indiana University of Pennsylvania covered the cost of re-scoring the profiles. The other data were coded using Microsoft Excel and subjected to analysis using Statistical Package for the Social Sciences (SPSS). This project was approved by both the Indiana University of Pennsylvania and the Saint Elizabeths Hospital Institutional Review Boards for the Protection of Human Subjects prior to data collection.

Planned Analyses

The current state of the literature regarding the MMPI-2 profiles of competent and incompetent individuals is sparse and conflicting. Although hypotheses about how these scales relate to competency have been presented, the literature is insufficient to focus solely on those few scales and neglect the possibility of significance among other scales. The fact that there is virtually no research on the MMPI-2-RF profiles of these two groups further complicates the matter. As a result, data analysis will proceed in three phases that allow for both exploration of the data and testing of research hypotheses. During the first phase, descriptive statistics will be computed in order to inform interpretation of subsequent analyses. The second phase will involve testing the differences between competent and incompetent patients using MANOVA to identify any

reliable differences between the two groups. In the third phase, a series of regression analyses will be carried out in order to describe the data using explanatory models that (hopefully) have predictive power and may be of practical use to those in applied competency settings.

During phase 1, descriptive statistics of MMPI-2 and MMPI-2-RF profiles among competent vs. incompetent individuals will be presented. The means, standard deviations, and ranges of the MMPI-2 10 Clinical Scales and 6 Validity Scales, as well as the MMPI-2-RF 9 Restructured Clinical scales, 8 Validity scales, and Higher-Order Thought Dysfunction scale will be calculated and presented in tabular form for competent and incompetent patients. Graphs containing this information will also be presented to allow for a visual examination of scale elevations, ranges of scores, and other patterns that unite or distinguish these two clinical groups. To this author's knowledge, there are only a handful of studies examining the full range of MMPI-2 Clinical and Validity scales in this population (Pirelli, Gottdiener, & Zapf, 2011; Nicholson & Kugler, 1991; Johnson et al., 1990), and there are no efforts examining the Restructured Clinical and other MMPI-2-RF scales. Accordingly, describing the types of profiles generated by these two different clinical groups is an important endeavor. Following these descriptive analyses, hypothesis testing will begin.

The first and second hypotheses to be investigated are whether the MMPI-2 and MMPI-2-RF protocols differ among competent and incompetent to stand trial individuals. MMPI-2 scales F (Infrequency), 6 (Paranoia), and 8 (Schizophrenia) are hypothesized to be higher among incompetent individuals, and for the MMPI-2-RF, scales F-r, RC6 (Ideas of Persecution), RC8 (Aberrant Experiences), and THD (Thought Disorder) are

hypothesized to be higher among incompetent individuals. Multivariate analysis of variance (MANOVA) tests will be used and competency status will be the independent variable in both analyses. For the MMPI-2, the three primary Validity scales (F, L, and K) as well as the nine Clinical scales will be the dependent variables. For the MMPI-2-RF, the three primary Validity scales (F-r, L-r, and K-r), the nine Restructured Clinical scales, and the Thought Disorder scale will be the dependent variables. This is grossly consistent with Bray and Maxwell's (1985, as cited in Meyers, Gamst, & Guarino, 2006) recommendation to use a maximum of approximately 10 conceptually related variables as dependent variables when conducting such analyses. Main effects from these analyses will be further examined using a series of univariate ANOVAs with Bonferroni adjustments to constrain the familywise error rate to 0.10.

MANOVAs are appropriate in situations where researchers wish to examine the extent to which multiple dependent variables differentiate several distinct groups (Meyers, Gamst, & Guarino, 2006). This analysis is optimal when correlations between dependent variables are moderate (Tabachnik & Fidell, 2001, as cited in Meyers, Gamst, & Guarino, 2006), and the intercorrelations of the RC scales range from $r = 0.15$ to $r = 0.70$ (see Appendix Table B4 for a table of intercorrelations of RC and Clinical Scales). Although *a-priori* power analyses are not reported for the MANOVA analyses, this range of intercorrelations can be expected to increase power (Tabachnik & Fidell, 2001, as cited in Meyers, Gamst, & Guarino, 2006). In many ways, this analysis complements those that follow, by testing whether linear combinations of conceptually related scales reliably differentiate competent and incompetent defendants.

Numerous methodological considerations informed the selection of this particular analytic strategy. First, there is little literature available to suggest which MMPI-2 and MMPI-2-RF scales are likely to differ between the two clinical groups. Limiting the analyses to testing only those scales that have been shown to differ would run the risk of missing potentially important effects. Fully exploring all of these effects, however, introduces a number of other concerns. For the MMPI-2-RF, 13 different variables have to be analyzed, making hypothesis testing problematic. In the univariate domain, constraining the experiment-wise error rate at $\alpha = 0.05$ would require setting the alpha level at $\alpha = 0.0038$ for each of the 13 tests. This would result in levels of power far below Cohen's (1977) recommendation of 0.80 and would signify a less-than-chance possibility of uncovering actual differences that may exist between groups in the present of even medium effect sizes (Faul, Erdfelder, Lang, & Buchner, 2007). Although this remains a problem in the current analysis, using multivariate statistics capitalizes on scale intercorrelations to maximize power and represents the most reasonable compromise between statistical sensibility and comprehensive data analysis.

Testing hypothesis 3 (i.e., the MMPI-2-RF will be superior to the MMPI-2 in predicting competency status) is more straightforward. A series of hierarchical binomial logistic regressions will be used in these analyses. Because this hypothesis is concerned with the difference between tests (MMPI-2 vs. MMPI-2-RF), the profiles of competent and incompetent patients will be pooled for these analyses, yielding increased power to test regression coefficients. The series of regression analyses will complement the more comprehensive tests above, by examining those variables that maximally distinguish the groups with greater power.

Two regression equations will be built, one for the MMPI-2 and one for the MMPI-2-RF. Competency status will serve as the criterion variable to be predicted by a weighted linear composite comprised of various clinical and validity scales. Simple variable entry will be employed. The primary goal of the present project is to provide useful pragmatic information regarding the prediction of competency and this entry strategy is sensible given this goal.

Once the optimal models are computed, they will be compared qualitatively in terms of face validity. Likelihood ratios, model summary, pseudo R^2 and the Hosmer and Lemeshow Test statistics will also be used to quantitatively examine the difference between the two competing models. Although it would be advantageous to do so, the two models will not be tested for statistically significant differences for a variety of reasons. The typical scenario in which one compares models involves testing nested models, as is the case when determining what variables significantly increase the power of the model, or whether the model is better than no model at all. In this case, 2-log-likelihood ratio can be computed for the null and alternate models. This statistic is distributed approximately as a Chi-Square with $df = df_2 - df_1$ where df_2 = the degrees of freedom for the alternate model and df_1 = the degrees of freedom for the null model (Wilks, 1938). This assumption does not hold in cases where the models are not nested (Vuong, 1989). Testing the difference between the MMPI-2-RF model and the MMPI-2 model would be a non-nested comparison. Several competing methods are appropriate in such circumstances, although they typically involve the use of simulation, complex structural equation modeling, or bootstrapping (e.g., Cheung & Chan, 2004; Kapetianos & Weeks, 2003). These approaches exceed the resources available in the present project.

CHAPTER 4

RESULTS

Participants were drawn from a population of individuals who underwent psychological evaluations at Saint Elizabeths Hospital, a District of Columbia Department of Mental Health federally funded hospital. Archival records available within the section of psychology were reviewed in order to determine whether participants met selection criteria, and all suitable protocols were considered for inclusion. Eligible records contained a completed MMPI-2 protocol. They also contained documentation stating that they were court-ordered to receive restoration treatment as a result of being adjudicated incompetent to stand trial (comprising the incompetent group) or stating that the participant was admitted for court-ordered involuntary treatment after being found Not Guilty by Reason of Insanity (NGRI; comprising the competent group). The individuals found NGRI were presumed to be competent, because competency is essential in order for the court hearing to occur. Records from the 60 individuals were collected and were evenly distributed between competent and incompetent status. One file could not be analyzed due to an administrative issue.

Demographic characteristics of the 59 remaining profiles are reported in Table 1. Independent samples t-tests of continuous demographic variables indicated significant group differences between competent and incompetent individuals with respect to age $t(57) = 3.413, p = 0.001$, length of hospitalization $t(57) = 5.039, p < 0.001$, and number of prior hospitalizations $t(55) = 2.297, p = 0.025$. Incompetent individuals were younger, had a briefer period of hospitalization, and fewer previous hospitalizations. Pearson Chi-square values revealed no significant differences with respect to gender $\chi^2(1, N = 59) =$

0.137, $p = 0.711$, race $X^2(2, N = 59) = 3.033$, $p = 0.223$, marital status $X^2(3, N = 59) = 6.010$, $p = 0.111$, or employment status $(2, N = 59) = 0.449$, $p = 0.799$.

The preponderance of individuals were male, African American, single, and unemployed. The average age of individuals was 43 years, although individuals who were incompetent to stand trial tended to be older. This is likely related to the groups' divergent lengths of hospitalizations, which will be addressed below. Additionally, only one competent individual was married, while seven incompetent individuals were married. Again, this may be due to the fact that many competent individuals were hospitalized for much longer periods of time.