Theses and Dissertations (All)

8-15-2013

Accuracy and Confidence in Competency to Stand Trial Evaluations: A Comparison of Clinical and Legal Decision Making

Sarah E. Ryan
Indiana University of Pennsylvania

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

ACCURACY AND CONFIDENCE IN COMPETENCY TO STAND TRIAL EVALUATIONS:

A COMPARISON OF CLINICAL AND LEGAL DECISION MAKING

A Dissertation

Submitted to the School of Graduate Studies and Research

in Partial Fulfillment of the

Requirements for the Degree

Doctor of Psychology

Sarah E. Ryan

Indiana University of Pennsylvania

August 2013

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

We hereby approve the dissertation of

Sarah E. Ryan

Candidate for the degree of Doctor of Psychology

June 25, 2013   __Signature on File_____
David J. LaPorte, Ph.D.
Director of Doctoral Studies, Advisor

June 25, 2013   __Signature on File_____
Margaret Reardon, Ph.D.
Assistant Professor of Psychology

June 25, 2013   __Signature on File_____
Dante Mancini, Ph.D.
Assistant Professor of Psychology

ACCEPTED

_____Signature on File_____   _____
Timothy P. Mack, Ph.D
Dean
School of Graduate Studies and Research

Title: Accuracy and Confidence in Competency to Stand Trial Evaluations: A Comparison of
    Clinical and Legal Decision Making

Author: Sarah E. Ryan

Dissertation Chair: Dr. David J. LaPorte

Dissertation Committee Members: Dr. Margaret Reardon
                                     Dr. Dante Mancini

       This study used a three group (forensic psychologists, clinical psychology graduate students, and lawyers) lens-model design. Protocols containing demographic, psychological, and criminological data taken from actual defendants referred for competency evaluations at an inpatient psychiatric facility were presented to participants. Participants were asked to make decisions of competency to stand trial and to rate their confidence in the decision for each protocol. Participant decisions were compared to decisions of competency to stand trial rendered by the legal system. Results suggest courts relied on a defendant's performance on measures of competency to stand trial, hospitalization history, and criminal offense while the majority of raters primarily relied on a defendant's performance on measures of competency to stand trial when making decisions of competency. There were no substantial differences in accuracy between the different disciplines. As predicted, confidence was unrelated to accuracy. For most raters there were no differences between human models of judgments and linear models of judgments on accuracy rates. When differences were observed, human models tended to outperform linear models of judgments.

ACKNOWLEDGMENTS

I would like to acknowledge the many people who contributed their time, energy, knowledge, and support to make this research project possible. I would like to thank David LaPorte for his guidance, humor, and reassurance during this project and for his encouragement to pursue forensic psychology. I am grateful for having had the opportunity to collaborate with Don Robertson on this project; he initially served as a committee member and later volunteered his time for statistical consultation. I would like to thank Margaret Reardon and Dante Mancini for their guidance, input, and willingness to serve as committee members. I appreciate the support and backing of Danielle Adamsky who made this project possible. I also appreciate the time Laura Gilman spent helping me distribute materials to participants as well as the time the folks at the Applied Research Lab spent providing statistical consultation. I would like to thank my supervisors at Central Regional Hospital for providing support and encouragement as well as the time to work on this document. In addition, this research project received funding through the Indiana University of Pennsylvania Graduate Student Research Grant Award.

This research project would also not have been possible without the support of my family. I would like to thank my mother for offering her support, knowledge of psychological principles, proofreading documents, and mailing materials. I would like to thank Mark Reeder, my rock, for his support throughout this project and his willingness to lend a hand (and an ear). I would like to thank my father for encouraging me to question what is "known" and offering his knowledge of the legal system. In addition, I am grateful for the love and support of Sue, Steve, Mary, Buddy, and the Reeder family. Lastly, I would like to thank Mary "Esther" McCloy for setting the stage for what was to come and George McCloy for his love and encouragement.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER I

INTRODUCTION

Competency to stand trial, also known as fitness to proceed, or adjudicative competence, is a central part of the legal system. Of the various criminal and civil competency evaluations, competency to stand trial evaluations are the most frequently performed, with at least 50,000 to 60,000 evaluations yearly, occurring in as high as 8% of defendants facing felony charges (Mossman et al., 2007; Robinson & Acklin, 2010; Skeem & Golding, 1998). The law requires an individual to be competent before proceeding in the legal process to maintain the integrity, dignity, fairness, and reliability of the legal system. When there is a "bona fide doubt" (p. 842) about the defendant's competency to stand trial, the court is obligated to order a mental health evaluation of the defendant's competency (*Pate v. Robinson*, 1966). This evaluation is conducted to provide the courts with information about a defendant's mental health and the impact of mental health symptoms on relevant legal abilities (Grisso, 1988; Skeem & Golding, 1998).

Although competency is primarily a legal decision that is left for the trier of facts to determine, many states require or encourage the examiner to make a determination or give an opinion to the ultimate issue (Grisso, 1988; Otto, 2006). Even though the trial judge makes the ultimate decision regarding a defendant's competency, if both lawyers accept the examiner's report, a formal hearing may be bypassed (Winick, 1983). Studies comparing judges' opinions with those offered by the examiners who conducted the evaluations suggest judges rely heavily on clinicians' reports to determine competency (Melton et al., 2007; Roesch & Golding, 1980; Skeem & Golding, 1998). Zapf, Hubbard, Cooper, Wheeles, and Ronan (2004) found this agreement rate to be as high as 99.7%, while Hart & Hare (1992) found the agreement rate to be 96.3%. This extreme reliance on the opinions of mental health professionals underscores the

importance of providing courts with reliable and valid evaluations of competence to stand trial (Skeem & Golding, 1998).

So far, the majority of the research on issues of competency to stand trial has focused on (a) defendant characteristics associated with findings of incompetence, (b) performance of defendants on traditional psychological tests, and (c) performance of defendants on specific psychological measures used to assess competency (Pirelli, Gottdiener, & Zapf, 2011). Recent research has focused on a) competency to stand trial in juvenile populations, b) malingering, c) competency restoration, and d) clarifying the components of competency (Fogel, Schiffman, Mumley, Tillbrook, & Grisso, 2013). Research has been conducted on clinician judgment and decision making processes in other areas of forensic assessment, with less emphasis on competency to stand trial assessments. To date, little is known about the validity of these determinations. An understanding of the process by which these decisions are rendered and the validity of these decisions is essential if mental health professionals are to provide expert witness testimony for the court system (Faust, 2003).

The purpose of the present research project is to compare the accuracy and confidence of clinical and legal professionals making decisions of competency to stand trial. This research project will attempt to a) identify the most influential variables for determining competency to stand trial, b) identify the decision making strategies used by clinical (expert and novice) and legal raters (lawyers), c) determine whether linear models of judgments better predict competency status than human models of judgments, d) examine the impact of experience and occupation on judgment accuracy, and e) determine the relationship between confidence and accuracy. Knowledge of the most predictive variables for making determinations of competency and the creation of linear models to represent the decision making strategies of raters can be used

to increase the accuracy and efficiency of these determinations. An increase in the accuracy of these determinations will protect the rights of the defendants referred for evaluations and improving the efficiency may reduce the time and associated cost of these evaluations.

An overview of the competency to stand trial doctrine is provided first, including a review of its origins, historic legal cases, interpretations of the legal statutes, the impact of these legal statutes on modern competency to stand trial evaluations, and a review of competency to stand trial evaluations specific to the state of Pennsylvania. This is followed by a review of the clinical decision making research, its application to forensic psychology, and a model and mathematical formula that has been used to study clinical judgment. Lastly, the current state of the judgment and decision making literature specific to competency to stand trial evaluations will be examined with an emphasis on the reliability and validity of these determinations.

CHAPTER 2

REVIEW OF LITERATURE

Competency to Stand Trial

Origins

Much of the United States law has origins in English common law and the competency to stand trial doctrine is no exception. The competency doctrine can be traced back to at least the 17th century, although some have argued it can be traced as far back as the 14th century, where individuals were required to make a plea prior to the trial (Roesch & Golding, 1980; Winick, 1983). If an individual did not enter a plea, the courts considered whether the individual was "obstinately mute" or whether the individual was "dumb ex visitatione Dei" (by visitation of God; Blackstone, 1765-1769, p. 477). The latter included an individual who was deaf or mute, and was eventually expanded to include those who were diagnosed with a mental illness (Melton et al., 2007). If individuals were suspected to be "obstinately mute," the court would punish them by a process known as "peine forte et dure" (p. S4), a form of torture where heavy weights were placed on the individual (Mossman et al., 2007). Individuals who were considered mute by visitation of God were spared this experience (Melton et al., 2007). The roots of the competence requirement may also be a result of the English common law prohibition against "trials in absentia" (Foote, 1960, p. 834). This prohibition ensured that the defendant would not only be physically present, but also mentally present for the trial (Robey, 1965).

The United States' practice of making sure an incompetent defendant does not proceed to trial is upheld by the sixth and fourteenth amendments and is based on the legal system's desire for accurate, fair, and dignified trials ("Harvard Law Review," 1967). The prevention of an incompetent defendant from proceeding with trial is reflected in the United States law as early as

1847 in the case of *Freeman v. People*. In this case, a New York County Supreme Court ruling stated that the defendant was brought to court "to ascertain whether the prisoner was sufficiently sane to be required to plead to and to be tried upon said indictment" (*Freeman v. People*, 1847, pg. 1). Later, the U.S. Supreme Court would decide that the requirement of competency is "fundamental to an adversary system of justice" (*Drope v. Missouri*, 1975, p. 172).

The sixth amendment guarantees the individual's right to effective counsel, to confront accusers, and to present evidence. An incompetent defendant may have substantial deficits in the ability to do these things and it is this difficulty that could result in an inaccurate (or unfair) trial (Foote, 1960). For example, a defendant's mental illness could potentially prevent the defendant from providing counsel with relevant information necessary to assist in a defense, thus resulting in an inaccurate verdict or outcome (Otto, 2006).

The fourteenth amendment guarantees a fair trial between equally matched adversaries. Some have even said "the trial of a defendant who cannot fulfill this expectation appears inappropriate and irrational" ("Harvard Law Review," 1967, p.458). It seems wrong to prosecute an individual who cannot provide an adequate defense or to prosecute an individual who does not understand the potential consequences of a verdict. In *Godinez v. Moran* (1993), the court confirmed that prosecuting a defendant deemed incompetent to stand trial was a violation of the defendant's right to due process guaranteed by the fourteenth amendment.

The Impact of *Dusky v. U.S.*

Although the United States' emphasis on the importance of a defendant's competency to stand trial is reflected in case history, it was not until Dusky v. U.S. (1960) that the U.S. Supreme Court identified criteria that must be present for a defendant to be deemed competent to stand trial. In this case, Milton Dusky was accused of kidnapping and unlawfully transporting a

juvenile female across state lines. Dusky requested a competency hearing that resulted in the district court judge finding him competent based on the results of a mental status examination that indicated he was oriented to time, person, and place. After his conviction, Dusky appealed his case to the eighth circuit court, arguing that the judge based the determination of competence on insufficient evidence (Dusky v. U.S., 1959). The eighth circuit court then asked the U.S. Supreme Court to establish guidelines for competency. The U.S. Supreme Court, in deciding these guidelines, ruled in favor of Dusky, stating "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'" (Dusky v. U.S., 1960, p. 402).

The U.S. Supreme Court has left the Dusky standard up to interpretation, and as such, researchers have attempted to provide clarification. The Dusky standard, as it has become to be called, has been incorporated into 42 of the 50 states, either through court interpretation or by legislation (Morris, Haroun, & Naimark, 2004). Many states have set guidelines that require the defendant's incompetency to be a result of the presence of a mental disease or defect, although the U.S. Supreme Court did not specify this criterion in Dusky (Grisso, 1988).

Interpreting Dusky

Bonnie (1992), in attempting to provide examiners with clarification, suggested the Dusky standard be separated into two prongs: foundational competence and decisional competence. According to Bonnie, foundational competence refers to the defendant's ability to understand the charges, the roles of the defendant's counsel, the nature and purpose of the legal process, the ability to appreciate the role of being a defendant, and the ability to provide one's

counsel with relevant information related to the case. These are considered the minimum capacities a defendant would need to effectively participate in one's defense.

Decisional competence is exactly what it sounds like: the ability of the defendant to make necessary decisions in a rational manner as they arise throughout the legal process. Bonnie (1992) posits that due to the varying situation of each defendant, the context may largely determine whether a defendant meets the standard of decisional competence. Bonnie also argues that decisional incompetence does not equate to overall incompetence and that the presence of foundational competence is both necessary and sufficient to deem a defendant competent. Clinicians may have difficulty assessing these constructs as foundational competence is easier to access, or measure, than decisional competence (Otto, 2010).

Otto (2006) provides further clarification of the Dusky standard. Otto suggests the term "sufficient" is often interpreted as meaning that some impairment is okay, and "present ability" provides a temporal context for the determination of ability; distinguishing competency evaluations from retrospective or prospective evaluations. The term "ability" is meant to differentiate capacity from willingness. "Reasonable" is similar to the "sufficient" qualifier in that some impairment does not equate to incompetency to stand trial. "Rational" has typically been construed to include reasoning and knowledge that is not tainted or impaired by psychological or other factors. This implies it is possible for an individual to have a factual, but not rational, understanding of courtroom proceedings. For example, an individual may have an understanding of the legal proceedings, roles of courtroom personnel, and the consequences of pending charges, but may irrationally believe the judge is part of a conspiracy to place the individual in jail.

Post-Dusky

Since its inception, aspects of Dusky have been clarified by the U.S. Supreme Court to include guidelines for when to request a competency hearing and factors that may influence competency determinations. For example, in Drope v. Missouri (1975), the U.S. Supreme Court was charged with determining what evidence constitutes reasonable doubt as to the defendant's competency. The U.S. Supreme Court found that "even when a defendant is competent at the commencement of his trial, a trial court must always be alert to circumstances suggesting a change that would render the accused unable to meet the standards of competence to stand trial" (Drope v. Missouri, 1975, pg. 181). The court also decided that a "defendant's irrational behavior, his demeanor at trial, and any prior medical opinion on the competence to stand trial" (p. 180) are all factors to consider when deciding whether a competency evaluation should be ordered by the court (Drope v. Missouri, 1975). However, no threshold is mentioned for when this criterion is met (Mossman et al., 2007).

Competency evaluations are important in both legal and psychological fields. The establishment of competency is essential to the fairness and dignity of the legal process. Trying an incompetent defendant can lead to a mockery of the legal system when the defendant is unable to comport their behavior to courtroom decorum and it seems unjust to prosecute an individual who is unable to assist in their own defense. These sixth and fourteenth amendments are constitutional rights guaranteed to every citizen of the United States. Allowing an incompetent defendant to proceed with trial can result in a violation of the fourteenth amendment, by allowing a trial between unmatched adversaries. On the other hand, barring a competent individual from proceeding to trial on the grounds of erroneous determinations of incompetency can violate the sixth amendment right to a speedy trial. By finding someone

incompetent to stand trial, the examiner is likely signing off on the individual's commitment to an inpatient hospital to undergo treatment aimed at competence restoration (Golding, Roesch, & Shreiber, 1984). This may result in commitment for an unspecified length of time, in a stricter setting, with more stringent requirements for release than those individuals committed civilly (Foote, 1960).

Competency to Stand Trial in Pennsylvania

The most recent census data for incompetency to stand trial defendants comes from a survey by Way, Dvoskin, and Steadman (1991). On a single day in 1986, there were approximately 3,200 defendants being served in inpatient facilities across the United States for incompetency to stand trial. For Pennsylvania, the rate of incompetent to stand trial defendants served in inpatient facilities on any given day was approximately 64. Of these individuals, 67.2% were served in forensic-designated treatment facilities, while 32.8% were served in civil facilities. It is important to note that these rates do not include defendants who were served by outpatient or community facilities or those individuals who were evaluated and found competent to stand trial.

Although most states incorporate some aspect of the Dusky standard, states vary in their definitions, evaluation procedures, and the treatment of incompetent to stand trial defendants. Although originally conducted solely by psychiatrists, a growing number of mental health professionals from different disciplines are being recruited to conduct these evaluations in both inpatient and outpatient settings. As of 1992, forty five states allowed testimony from mental health professionals other than psychiatrists, with Delaware, Iowa, New Jersey, Vermont, and Pennsylvania barring such testimony (Farkas, DeLeon, & Newman, 1997).

According to the Mental Health Procedures Act of 1976, incompetency to stand trial in Pennsylvania is defined as:

> 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 (Definition section, para. 1).

If there is a question as to a defendant's competency, the court may order an evaluation to be completed by a psychiatrist. The Mental Health Procedures Act of 1976 states "the court may allow a psychiatrist retained by the defendant and a psychiatrist retained by the Commonwealth to witness and participate in the examination" (Experts section, para. 6). There is no mention of employing a psychologist or other mental health expert to complete the evaluation.

According to the Mental Health Procedures Act (1976), competency evaluations in Pennsylvania are conducted on an outpatient basis, unless an evaluation at an inpatient facility is authorized. In Pennsylvania, there are two regional forensic psychiatric centers that conduct competence to stand trial evaluations on an inpatient basis: one is located in eastern Pennsylvania and the other is located in western Pennsylvania.

In most states, psychiatrists and psychologists are able to provide expert testimony regarding a defendant's competency to stand trial (Farkas, DeLeon, & Newman, 1997). In Pennsylvania, psychiatrists are the only mental health professionals who provide court testimony about a defendant's competency to stand trial. Pennsylvania House Bill 1405 (2011) proposed to amend the Mental Health Procedures Act of 1976 by adding psychologists as experts the court can retain. This bill was passed in the House and has moved to the Senate (now called Pennsylvania House Bill 21). The use of psychologist expert testimony on matters of

competency to stand trial is practiced in the majority of the United States. Thus, psychologists can be considered qualified to assess competency to stand trial and provide opinions.

Given the number of defendants assessed for competency to stand trial each year, along with a reliance on mental health professionals to make decisions about a defendant's competency to stand trial, the clinician's role in the courtroom is becoming increasingly important. The correct identification of incompetent defendants ensures the protection of individuals' constitutional rights to a fair trial between matched adversaries, where the accused can confront the accusers, present evidence, and assist counsel in their defense (Grisso, 1988; Mossman et al., 2007). However, the misclassification of a competent defendant as incompetent threatens these and other rights guaranteed to the people including the right to a speedy trial and equal protection, as individuals found incompetent are subject to more lenient commitment criteria and more stringent release criteria than their civil counterparts (Foote, 1960). Therefore, it is essential that clinicians form their opinions based on relevant variables to perform reliable and valid evaluations of defendants waiting to stand trial.

The following section will begin with a broad discussion of clinical judgment and decision making, followed by how limitations in decision making apply to forensic psychologists. This will be followed by a discussion of a particular model, the Brunswik Lens Model, used to examine the accuracy and decision making process of judges for a variety of situations.

Clinical Judgment and Decision Making

Judgment and decision making is a topic that has been studied extensively in a variety of disciplines in which judgments and decisions are made. For clinical psychology, research on judgment and decision making has largely focused on how clinicians use information to

formulate judgments as well as the consistency and accuracy of these judgments (Elstein & Bordage, 1979/1988). Through empirical investigation of this construct, researchers have identified limitations that impede individuals from making accurate judgments and models to correct for these errors. The limitations take the form of heuristics (rules that describe how judgments are made) and cognitive biases (mental errors made when making judgments), while corrective strategies include the use of linear models to be used as decision tools when making judgments.

Limitations of Clinical Judgment

When individuals make probabilistic judgments, their decision-making process is facilitated by the use of heuristics (Saks & Kidd, 1980/1986). It is through the use of these heuristics that individuals are able to make quick judgments in everyday decision making tasks. However, heuristics are often simplifications that leave judgments open to bias and error (Tversky & Kahneman, 1974). These heuristics and biases affect individuals involved in a variety of decision making situations. Meehl (1973) identified over fifteen factors that can influence clinicians' decisions, whether when making judgments about diagnosis, identifying appropriate assessment measures to employ, or choosing therapeutic interventions. For example, in competency determinations, clinicians must decide what information to gather, how to gather the information, the importance of the information, and the overall determination. Although Meehl was able to identify numerous factors that influence the decision making process, many of these factors are subsumed under three primary heuristics: representativeness, availability, and adjustment and anchoring.

The representativeness heuristic occurs when a clinician makes a decision about whether an individual meets criteria for Diagnosis X based on how well the individual matches the

clinician's mental schema of previous individuals diagnosed with X. For example, if a clinician has seen many clients with Schizophrenia struggle to maintain adequate hygiene, the clinician may associate poor hygiene with a diagnosis of Schizophrenia. In the future, this clinician may be inclined to diagnose someone as having Schizophrenia based on the individual's hygiene. Another heuristic, known as the availability heuristic, occurs when judgments about the likelihood of an event are made based on how readily other events of a similar nature are recalled. For example, a clinician who sees a large number of individuals with suicidal ideation may believe that suicidal ideation is more prevalent in the general population than base rate information suggests. Lastly, the anchoring and adjustment heuristic includes situations in which individuals form an initial opinion, often based on partial or incomplete information, and fail to adequately adjust their opinion after obtaining more information (Tversky & Kahneman, 1974). For example, during the initial five minutes of meeting a client, the clinician decides the client's likely diagnosis is a mood disorder. As a result of the anchoring and adjustment heuristic, the clinician may fail to adequately adjust the original hypothesis of mood disorder to account for new information learned during the clinical interview. These heuristics and biases may further limit rates of accuracy in decision making situations. This is unfortunate, especially in forensic evaluations, given the high stakes of the outcomes.

If clinical judgment and decision making is a subcategory of the judgment and decision making literature, then examining judgment and decision making in forensic psychology is a more specialized area, with even less research to support it. Borum, Otto, and Golding (1993) described several problems identified in the judgment and decision making literature and applied them to clinicians providing diagnoses or expert testimony in a forensic setting. For example, during testimony, the clinician may recall symptoms the patient did not have solely because the

symptoms fit the schema for inclusion criteria. In addition, information that is unique, salient, or vivid is more easily remembered (i.e., a defendant's legal charges) than less interesting but relevant information.

Not only are individuals' decision processes susceptible to error, but research has shown that introspective accounts of the decision making process are also prone to error. After a review of the relevant literature, Nisbett and Wilson (1977) concluded that a) individuals are unable to provide accurate accounts of their thought processes, b) individuals actually respond to stimuli based on a priori beliefs about the plausible relationship between the stimulus and the response but attribute this to their thought processes, and c) accurate subjective reports are a result of previously held causal theories as opposed accurate introspection. Further, research has demonstrated that individuals tend to be overconfident in their judgments and that this confidence is generally unrelated to accuracy (Jackson, 1986; Saks & Kidd, 1980/1986).

The Brunswik Lens Model

The number of heuristics and biases that impact clinicians' judgments is vast. In the 1950s, Egon Brunswik developed a model to represent the ways in which judgments are made about objects in the environment. This model, known as the Brunswik Lens Model, is based on the idea that individuals make inferences about things they can not readily observe by using observable cues in the environment (Brunswik, 1955; Elstein & Bordage, 1979/1988). For example, physicians make inferences about the presence of medical conditions based on observable symptoms. A pictorial representation of this model is shown in Figure 1. Decades later, this model, and an equation developed after it, continued to find use for studying the validity of judgment and decision making in a variety of disciplines.
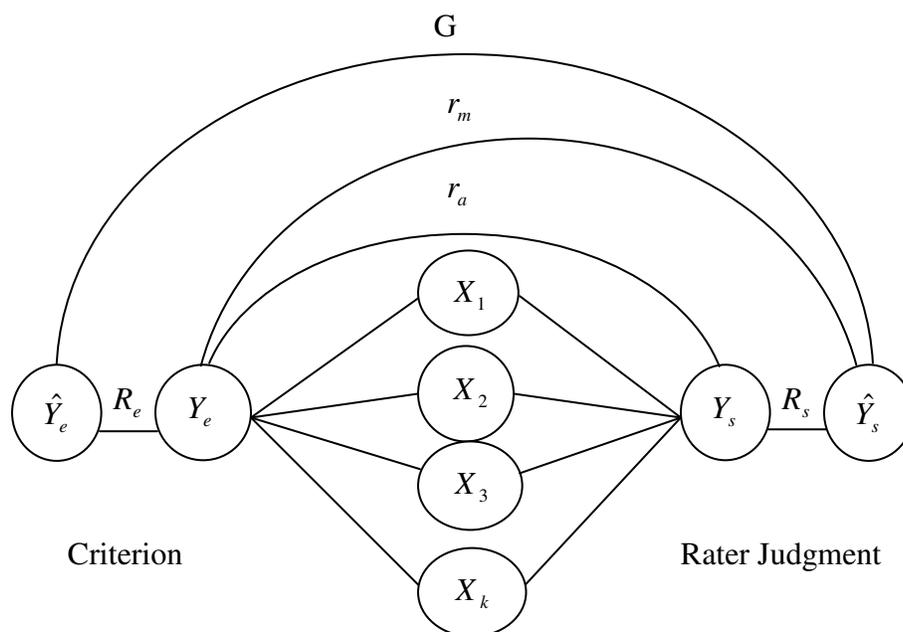
Figure 1. A pictorial representation of Brunswik's Lens Model with relationships shown between the human rater, the criterion, and the data cues. From "The Methodology of Social Judgment Theory," by R. W. Cooksey, 1996, *Thinking & Reasoning*, *2,* p. 161. Copyright 1996 by Psychology Press. Adapted with permission.

Using Figure 1, along with an expanded version of an example provided by Elstein and Bordage (1979/1988) of a physician making a judgment about the presence of illness, the above diagram will be explained. For purposes of this explanation, the terms "judge" and "rater" may be used interchangeably. The right side of the model represents the physician's judgment about the presence of illness (shown as Y_s). The subscript s refers to the human rater or "subject," in this example, the physician. The left side represents the criterion, or whether the illness is present (shown as Y_e). The subscript e refers to the ecology or "environment," another name for the criterion. In this example, the patient's symptoms, represented as $X_{1...k}$, are the data cues the physician uses to make the judgment. Both the physician's judgment and the criterion are related to the data cues, and the extent of the relationship can be determined through analyses of

correlation. The correlation between the symptoms (or data cues) and the physician's judgment is represented by R_s, and is known as the linear predictability of the judge (Goldberg, 1970). The correlation between the symptoms and the criterion is represented by R_e, and is known as the linear predictability of the criterion. The achievement index, or validity coefficient (r_a), demonstrates the correlation between the physician's judgment and the criterion and is a measure of the accuracy of the physician's decision.

If the physician was asked to make judgments about the presence of illness for multiple patients, a linear model of the physician's judgment policy could be created using regression equations. The output of the linear model of the judge is represented as $\hat{Y}_s$, while the output of the linear model of ecology is represented as $\hat{Y}_e$. The correlation between the predicted scores from the linear model of the judge with those from the linear model of the criterion is called linear knowledge and is represented by G. When this linear portion is removed from the criterion and from the judge, the correlation between the residual values of the judge and the residual values of the ecology is called the nonlinear component of judgmental accuracy (referred to as C). Goldberg (1970) defined r_m as the correlation between the predicted scores from the linear model of the judge and the criterion. This provides an estimate of how well the linear model predicts the criterion.

In 1964, Hammond, Hursch, and Todd developed an equation to represent the Lens Model. It was improved by Tucker (1964) and has been used to assess the accuracy of judgments in a variety of disciplines to include child abuse, weather forecasting, and interpersonal learning, to name a few (Goldberg, 1970; Hammond, 2000). Using this equation, one can determine the validity of the judge's judgments as well as the accuracy of using a linear (G) or nonlinear

approach (C) to decision making. The difference between the accuracy of the judge and the linear model of the judge can also be used to determine the usefulness of statistical or human methods.

The term "bootstrapping" has been applied to describe studies using the Lens Model. In bootstrapping, linear models are created from the judgments or decisions made by a judge. The judge's judgments are analyzed using regression analyses and a linear model is created to represent the judge's policy or approach to decision making. The regression analysis helps to identify the data cues used by the judge along with the weight assigned to each cue. The term bootstrapping is used to suggest that the linear model is picking the clinician up by the bootstraps, or helping to improve the clinician's accuracy. This approach rests on the assumption that the clinician is using a valid approach but in an inconsistent or imperfect way (Dawes, 1982). Indeed, linear models have consistently been found to perform at least equal if not better to the judge (Dawes, Faust, & Meehl, 1989).

Research supports that expert clinicians have knowledge beyond that of novice clinicians, yet studies have consistently documented the absence of benefits from experience in tasks of judgmental accuracy (Garb, 1998). So what is going on here? It has been suggested that clinicians are able to effectively identify relevant information, or data cues, to use when forming their judgments, but are unable to integrate the data effectively (Dawes, 1982). Goldberg (1970) provides an example of the contrast between a machine and a clinician. A machine is able make perfectly reliable judgments while clinicians are susceptible to "boredom, fatigue, illness, situational and interpersonal distractions" (p. 423) that may infringe on reliability. This lack of perfect reliability allows room for the presence of error and this decreases accuracy. Another

reason is that clinicians may be using accurate information but in a nonlinear way, or their judgments may be affected by cognitive heuristics and biases (Dawes, 1982).

If the clinician is using a valid approach, bootstrapping will improve upon the clinicians' decision by separating out the unreliability of clinical judgments (Goldberg, 1970). Linear models of judges are able to weigh data cues in a reliable way and assign the most weight to those variables heavily relied upon by the judge. When the judge is using a valid policy, the variables with the most weight can be said to be the best predictors of the construct. Linear models are also able to reliably combine the data cues. As long as the criteria used by clinicians in making their decisions can be identified, a model of the clinician's judgment process can be created and will either outperform or equal the judge because it is reliable, and thus is able to achieve greater accuracy as a result of decrease in error variance.

This is not to say that the clinician's role in the decision making process is unimportant. There are several situations in which an individual may perform better than the linear model. According to Dawes et al. (1989), these situations may be cases in which the judge is using a valid and complex theory or using knowledge of rare events. The problem with using valid theories in the social sciences is that there may be multiple theories to describe a phenomenon. Meehl (1957) illustrates an example of using a rare event to make a more accurate prediction. In Meehl's example, the likelihood of a person attending a movie is estimated by a human judge and an actuarial method. In this case, the actuarial method is likely to do as well or better than the human judge. However, if the human judge learns that the movie-goer has a broken leg, this knowledge would likely render the human judge more accurate. Unfortunately, empirical studies examining the impact of knowledge of rare events on accuracy suggest individuals make too many exceptions based on rare events, leading to decreased accuracy (Dawes et al., 1989).

Dawes et al. (1989) discuss the use of configural methods of data combination and the importance of the human judge. Configural approaches to data combination include "if… then" approaches to decision making that consider complex relationships among available information. Both human judges and actuarial methods are able to use configural approaches to data combination and most information can be quantified and entered into actuarial formulas. The human judge's identification of relevant variables is still a necessary step, but it is by combining these variables with actuarial methods that greater accuracy is achieved (Dawes et al., 1989). As Dawes (1982) stated, "But people *are* important. The statistical model may integrate the information in an optimal manner, but it is always the individual (judge, clinician, subject) who chooses the variables" (p. 394). Thus, Dawes highlights the importance of the human judge's ability to identify relevant information and the statistical model's ability to integrate the information.

The methods allowing clinicians to be bootstrapped led to a controversy in the field regarding the use of clinical versus actuarial judgment. Dawes et al. (1989) summarized three decades worth of research in this area and concluded that actuarial or statistical methods outperform or equal those decisions made by clinicians on nearly every occasion. Of almost 100 studies looking at this phenomenon, statistical models consistently outperformed or equaled the human judges.

The current research project is modeled after two research studies. The first of which was conducted by Gaudette (1992) through the Indiana University of Pennsylvania. Gaudette used the Lens Model and related mathematical equation to determine the accuracy of neuropsychologists. The second study was conducted by researchers Cooper and Werner (1990),

who applied the Lens Model to a forensic setting to determine clinicians' accuracy for predicting dangerousness

In his dissertation, Gaudette (1992) used the Brunswik Lens Model and bootstrapping methods to "bootstrap the clinical neuropsychologist" (p. 2). Gaudette presented experienced and novice neuropsychologists with fifty protocols containing information extrapolated from actual patient reports. The protocols were approximately one page in length and contained demographic information as well as information related to the patient's performance on various neuropsychological assessment measures. Upon reviewing the data, the neuropsychologists were asked to make a determination as to the presence or absence of brain damage as well as the location of the damage. The purpose of the study was to compare the accuracy of linear models of judgment to those of human judgments.

The criteria were determined based on the presence of brain-injury and the location of the injury as determined from the patients' medical records, brain imaging scans, and neurological examinations. Through the use of bootstrapping, a linear regression model was created for each judge based on a subset of information presented in the protocols. This model was then compared to the judgments made by the neuropsychologists to determine differences in accuracy. A Majority Judge was created for both groups (expert and novice) and represented the majority decision for each group. For example, if two of the three expert neuropsychologists on a given protocol decided the person was brain-injured, the decision for that protocol was coded as brain-injured. Linear models for both Majority Judges were created and compared to the decisions rendered by the Majority Judges. Also created was a Composite Judge for each group, created by averaging the decisions rendered by the judges in that group. A linear model of the Composite Judge was created and compared to the Composite Judge. Based on previous research

in this field (Dawes et al., 1989), it was hypothesized that the linear models would outperform the human judges for prediction in both scenarios (presence of damage and localization of damage). It was further hypothesized that no differences would exist in accuracy for judgments made by expert versus novice neuropsychologists.

The results supported the first hypothesis: the linear model performed as well or better than the human judges for making determinations of brain damage and identifying localization of the damage. Regarding the difference in accuracy between expert and novice opinions, the researcher found no significant differences. Also identified were the data cues that neuropsychologists relied on the most when forming decisions (Digit Symbol and Block Design subtests of the Wechsler Adult Intelligence Scale-Revised and Controlled Oral Word Association Test) and those relied on the least (Finger Tapping, Similarities subtest of the Wechsler Adult Intelligence Scale-Revised, and the delayed recall condition for the Visual Reproduction subtest of the Wechsler Memory Scale-Revised).

In this study, clinicians were also provided with information as to the base rate of the occurrence of brain damage patients in the protocol set. For example, the neuropsychologists were notified that 40 protocols were from brain-injured patients while 10 were from normal patients. Clinicians' failure to consider base rates is one of the fallacies identified by both Tversky and Kahneman (1974) and Borum et al. (1993) as a pitfall in judgment and decision making tasks. The base rate refers to how often the phenomenon occurs in the general population. For example, across 59 studies that depicted defendants referred for evaluations of competency to stand trial, the base rate of incompetency was determined to be approximately 27.5% (Pirelli et al., 2011). Thus, regardless of the defendant's presentation, clinicians could

deem the defendant competent in 100% of their decisions and be accurate just over 70% of the time. Interestingly, in the study by Gaudette (1992), no judge performed better than the base rate.

Cooper and Werner (1990) examined the ability for staff to predict violence for newly admitted inmates at a federal correctional facility. The researchers identified eight inmates who had documented violent behavior shortly following their admission to the facility as well as twenty-five inmates who did not display violent behavior after admission. The judges consisted of federally employed psychologists and case managers working for the Bureau of Prisoners. The data presented to the judges were extrapolated from the inmates' records and included demographic and criminological information. Judges were asked to review the 33 protocols and make judgments as to the likelihood that the inmate would act violently during the initial six months at the facility. They were also asked to make a rating of their confidence in their prediction for each protocol. Unlike the judges in the study by Gaudette (1992), these judges were not given base rate information. The purpose of the study was to examine the accuracy and strategies used by correctional staff when making predictions of violence, to identify relevant variables for prediction of violence, to discover whether increasing the number of staff predictions increases the level of accuracy, and whether differences exist among professionals with different training backgrounds. No formal hypotheses were described.

A linear regression model was created for each judge based on the demographic and criminological information and was compared to the judgments made by the correctional staff to determine differences in accuracy. Also created was a Composite Judge for each group as well as an overall Composite Judge by averaging the decisions made by all judges, regardless of the assigned group.

Similar to the results found in Gaudette's (1992) research, Cooper and Werner (1990) discovered that experience did not affect judgment accuracy, as both psychologists and case managers displayed low predictive accuracy. The mean hit rate for psychologists was actually lower (.20) than the case managers' hit rate (.27). By examining the Composite Judges, the researchers discovered the psychologists relied less on the best predictors of violence (legal history) and more on cues that were unrelated to the criterion (ethnicity). The case managers did this but to a lesser extent in that they used at least two cues that were identified as best predictors of violence. These best predictors of violence from the sample set were identified as arrest history, conviction history, age, and place of legal residence. Judges confidence in their predictions was unrelated to their accuracy.

The last section will provide a critical review of the role of forensic psychologists in the courtroom with an emphasis on the current research regarding the reliability and validity of decisions made by forensic psychologists. This will be followed by a review of the factors that have been demonstrated to influence clinical decision making and judgments made by forensic psychologists in competency to stand trial evaluations.

A Critical Look at Forensic Psychology in the Courtroom

Although both mental health professionals and legal professionals may approach decision-making tasks with the same fallibilities, their underlying approaches and conceptualizations of human behavior may be different. Melton et al. (2007) cite "paradigm conflicts" (p. 8), or differences in approach, as reasons behind the tensions that arise between the mental health and legal systems. For example, mental health professionals often adopt a deterministic view of human behavior, believing that an individual's behavior is a product of a multitude of factors in the individual's environment, while legal professionals are likely to

believe that the individual acts out of free will and is thus responsible for their behaviors. Although both the mental health and legal systems have a vested interest in discovering the truth, truth is secondary to the legal system's emphasis on justice. Likewise, although both systems use probabilistic terms, the legal system emphasizes certainty and categorical decisions over probability estimates and dimensional constructs (Melton et al., 2007). These differences in approach have led to some heated debates between mental health and legal professionals regarding the role of psychologists in the legal system.

According to Faust (2003), the ability for clinicians to make accurate predictions in the court of law is the most important task of forensic psychologists. Faust suggests that clinicians difficulties in making accurate decisions is in large part because psychology as a whole often does not meet three conditions necessary for theory-based prediction to be accurate. These three conditions include knowledge of the construct and how it operates, the parameters that affect outcomes, and access to tools that accurately measure these parameters. Faust asserts that although clinicians are aware of the fallibility of using data to make predictions, clinicians often take a deterministic view of the data and would be better served to conceptualize data from a probabilistic view.

Although Faust (2003) identifies four ways to increase the accuracy of clinical judgment and decision making, his main argument boils down to the use of actuarial prediction methods, first discussed by Meehl in the 1950's. Actuarial methods take into consideration the incremental validity of variables, base rate information, are able to avoid conjunction errors, and incorporate inclusion and exclusion criteria. Although Faust recognizes that actuarial methods may not exist for every clinical task, he proposes using them for questions considered in the courtroom.

In their article entitled "Psychiatry and the Presumption of Expertise: Flipping Coins in the Courtroom," Ennis and Litwack (1974) opine for the removal of psychiatrist expert witness testimony in civil commitment hearings and propose that psychiatrists should instead provide descriptive statements that exclude predictions, diagnoses, and judgments. According to the authors, there is an "assumption of expertise" (p. 695) given to psychiatrists in the courtroom that assumes the psychiatrists are able to make reliable and valid conclusions. They posit that in order for psychiatrists to meet the standards for providing expert testimony, conclusions must be not only reliable and valid, but based on professional knowledge with better accuracy than those made by uninformed laypersons.

Impact of Experience and Training

In his book Studying the clinician: Judgment research and psychological assessment, Howard Garb (1998) reviews decades worth of research on the impact of training and experience on reliability and validity for assessments of personality and psychopathology, diagnosis, behavioral predictions, treatment decisions, and neuropsychological assessment. From these studies, several broad conclusions may be drawn: a) experts and novices are able to make judgments with equal accuracy, b) psychologists and graduate students do not differ in the validity of their judgments, however, c) advanced graduate students outperform first-year graduate students, and d) in general, clinicians and graduate students outperform laypersons, but (e) clinicians do not outperform laypersons when making judgments about non-clinical populations.

In the second of two studies, Walters, White, and Greene (1988) examined the abilities of 20 graduate students, 16 psychologists, and 24 expert psychologists to detect malingering in 32 protocols containing MMPI data from an inmate population. Of these participants, 31.7%

performed at a higher than chance level when making determinations of malingering. Actuarial models created to predict malingering from data collected in the first study were compared to the ability of judges in the second study. Although the actuarial models were able to correctly classify more individuals than the human judges, the difference in accuracy was not significant. Also not significant were differences among the judges regarding level of experience.

As an added level to the study, Walters et al. (1988) gave half the participants base rate information and omitted this information from the instructions given to the other participants. Interestingly, this condition was not significant. In other words, the presence or absence of base rate information had no impact on clinicians' decisions. The researchers also examined the difference in the participants' approach to decision making. They found that participants who used linear strategies had a significantly higher mean hit rate (60.2%) than those who relied on configural approaches (54.2%).

Several limitations to this study exist and include problems related to the design and methodology. The largest potential problem lies with the criterion to which the participants' judgments were compared. The researchers identified protocols of malingering to be used in the research by blindly rating each protocol and then determining interrater agreement. It is not known if the malingerers identified a priori by the researchers were in fact malingering.

Grøndahl, Grønnerød, and Sexton (2009) compared decisions made by expert Norwegian forensic psychiatrists and psychologists with those of laypersons through the use of 18 vignettes depicting criminal defendants. The professionals and laypersons were asked to rate the extent to which the vignette depicted a defendant who met legal criteria for Insanity, risk of repeated offense, and need for treatment. Statistically significant differences were found for all three rating conditions when professionals were compared to laypersons. Also noted was an

interaction effect where laypersons ratings were higher for those vignettes depicting defendants with serious crimes and negative attributes while professionals ratings were unaffected by this information.

Less emphasis has been placed on examining differences between mental health professionals compared to other professionals for competence to stand trial evaluations, despite issues raised decades ago by Faust (2003) and Ennis and Litwack (1974). One study made an attempt to compare the decision making strategies of judges, psychiatrists, and laypersons for decisions related to six different areas of forensic assessment, including competence to stand trial (Jackson, 1986). Of three studies that examined differences among evaluators in competency to stand trial decisions, the focus was primarily on differences in the rates of competency decisions, report quality, and clinicians' approach to evaluation.

Jackson (1986) used vignettes depicting defendants with varying social histories, psychiatric assessment information, and criminal charges to identify differences in decisions offered by psychiatrists, laypersons, and district court judges. Also examined was the impact of biasing information on forensic decisions. The vignettes were modeled from actual cases and manipulated information related to the defendant's social history (depicted as positive, negative, or absent this information) psychiatric assessment (positive, negative, or absent), and the seriousness of the offense (always present and depicted as minor or serious). After reading the vignettes, participants were asked to make dichotomous and continuous ratings of the defendant related to 6 areas of forensic assessment: dangerousness, insanity, need for treatment, fitness to stand trial, criminal responsibility, and predictions of recidivism. Following the ratings, participants were asked to rate their confidence in their decisions.

Jackson (1986) found that for decisions of fitness to stand trial, laypersons, psychiatrists, and judges were all significantly influenced by psychiatric assessment information. For laypersons, this included an increased likelihood of finding the defendant fit to stand trial if the vignette contained positive psychiatric information. Decisions of fitness and criminal responsibility were the only decisions significantly affected by the participants' discipline. Specifically, for fitness to stand trial decisions, psychiatrists were more likely than judges to deem a defendant unfit for trial. Otherwise, no differences were noted between decisions rendered by laypersons, psychiatrists, and judges.

Of the various decisions that participants rendered, decisions about a defendant's fitness to stand trial received the highest ratings of confidence for laypersons (80% of laypersons) and the third highest area of confidence for professionals (66% of psychiatrists and judges, third to insanity and criminal responsibility). Also notable was the presence of significant correlations between confidence and the participants' continuous ratings. For example, participants who rated defendants as definitely fit to stand trial expressed higher ratings of confidence in their decisions than participants who rated the defendant's fitness as questionable. Decisions about fitness to stand trial demonstrated the highest correlations between the participants' confidence and the direction of their ratings of the construct.

Petrella and Poythress (1983) examined 120 competency to stand trial reports written by psychiatrists, psychologists, and social workers. These reports were rated for thoroughness and three legal experts (a judge, a lawyer, and a law professor) provided ratings as to the quality of the reports. On measures of thoroughness, psychologists and social workers differed significantly from psychiatrists. Psychologists and social workers wrote lengthier clinical notes and sought

collateral information more often than psychiatrists. No significant differences were found among psychiatrists, psychologists, and social workers for report quality.

Warren et al. (2006) examined the process by which clinicians form opinions by reviewing 8,416 competence to stand trial evaluations and Forensic Evaluation Information Forms that contain information about evaluation process (e.g., time spent interviewing, psychological tests used). Differences were uncovered between evaluations prepared by psychiatrists and those prepared by psychologists, although psychologists were overrepresented in the sample, accounting for 84.7% of the evaluations. As a group, psychologists found defendants incompetent more often than psychiatrists (21% versus 9%). Psychologists significantly used more psychological tests (16.9% versus 3%) and neuropsychological tests (5.9% versus 1.1%) than psychiatrists. Psychologists also obtained more sources of information than psychiatrists to include copies of warrants, statements made by the defendant, offense information, criminal history information, and statements from witnesses Also noted was the lack of this information in the literature. Psychiatrists, on the other hand, asked for the referral information and previous medical and psychiatric records more often. Psychologists also logged more hours with the client, spending more time performing the clinical interview, collecting information, testing, and writing the report.

Murrie, Zapf, Boccaccini, Warren, and Henderson (2008) were the first to examine individual clinician variability in competence to stand trial decisions. The researchers completed a retrospective review of more than 7,000 evaluations conducted by clinicians in Virginia and Alabama. The researchers found that some clinicians never found a defendant incompetent, while others found defendants incompetent as often as 62% of the time. The Virginia sample, which contained competency opinions rendered by psychiatrists, psychologists, and social

workers, was used to examine the impact of evaluator profession on competency decisions. Approximately 12% of the variability in the decisions was due to clinician differences. When the evaluators' discipline was accounted for, the amount of variance as a result of evaluator differences decreased from 12% to 9%. Evaluator discipline significantly influenced the findings, with social workers being three and a half times more likely than psychologists to find a defendant incompetent to stand trial and psychologists being two times more likely than psychiatrists to find a defendant incompetent to stand trial. This is an important finding, as the determination of competence should be based on the defendants' psycho-legal abilities, and not on the clinician conducting the evaluation. Although this provides insight into the variability of clinicians' decisions, the researchers were limited in the ability to determine the factors that accounted for this variation. There remained a significant amount of variability due to clinician differences left unaccounted for. For example, clinician training, experience, approach to the evaluation, and referral source were not examined (Murrie et al., 2008).

These have been the only studies to examine differences between mental health disciplines with those of other disciplines for issues of competency to stand trial. No studies have been published that compare the validity of competence to stand trial decisions between experts and novices, psychologists and graduate students, or even clinicians and legal personnel (e.g., judges, lawyers). Despite the superiority of statistical methods to human judgments, and the potential cost saver of their implementation, to date, no studies have attempted to look at the accuracy of forensic psychologists making decisions about competence to stand trial using a Lens Model approach.

Reliability and Validity of Judgments

So far the discussion has focused on the fallibility of clinical judgment, the factors that may affect it, how to study it, and what the literature has shown in terms of the influence of expertise and training. But what do we know about the reliability and validity of clinical decision making in competency to stand trial evaluations? Despite the fallibility of clinical decisions in general, research on examiners' evaluations of competency has largely supported the reliability of these decisions. Agreement among evaluators conducting competency evaluations has ranged between 75% and 100% for categorical determinations, suggesting high rates of reliability (Morris et al., 2004; Poythress & Stock, 1980; Rosenfeld & Ritchie, 1998). However, it seems reliability is not limited to clinicians alone. In a study comparing mental health professionals' determinations with those of lawyers, the agreement rate was 97%. After correcting for chance agreement (by using Kappa), the agreement rate was 93%, suggesting high reliability can be reached even among non forensic mental health evaluators (Golding et al., 1984).

Gowensmith, Murrie, and Boccaccini (2012) retrospectively reviewed 251 cases resulting in 729 competency to stand trial evaluations performed by evaluators in Hawaii between September 2007 to December 2008. In Hawaii, evaluations of competency to stand trial comprise a panel of three evaluators who perform independent evaluations. Of the three evaluators, one must be a state employee and of the private (non-state) evaluators, one must be a psychiatrist. Opinions were compared between evaluators and to the ultimate opinion rendered by the court. For initial evaluations (as opposed to reevaluations), 72.7% of the defendants were deemed competent and 25.4% were deemed incompetent (1.9% were unknown). The courts used a more conservative approach, finding 68.1% of the defendants competent and 30.2% incompetent (1.5% were unknown). All evaluators assigned to a case were in full agreement

70.9% of the time (Kappa = .65). Cases in which there was disagreement contained a nearly equal percentage of disagreement for defendants found competent versus incompetent. That is to say, the defendant's competency status (competent or incompetent) was unrelated to rates of disagreement. Also noteworthy is that disagreement rates were not influenced by profession (psychologist vs. psychiatrist) or setting (state vs. private). When there was disagreement among evaluators, judges agreed with the majority opinion reached by the evaluators 92.5% of the time (Kappa not reported). Interestingly, for cases in which all three evaluators agreed a defendant was incompetent, the court agreed 100% of the time but when all three evaluators found the defendant competent, the court agreement rate was 97.2%.

It is not enough to provide evidence for the reliability of competency to stand trial determinations and there is an inherent problem in determining the validity of these decisions. According to Mossman (2008), no gold standard exists to which to compare clinicians' decisions of competency. Even Ennis and Litwack (1974) note the difficulty in performing assessments of validity and the lack of feedback that clinicians receive regarding their decisions. To assess the validity of clinicians' decisions, it would require courts to let individuals presumed incompetent to stand trial proceed to trial and observe the outcome, a practice that would hardly be considered ethical. However, some have suggested this design in the form of a provisional trial, where if the defendant was convicted, the verdict could be postponed if evidence was available to show that the defendant was incompetent (Roesch, Ogloff, & Golding, 1993). Still others have argued that court decisions, although a potential source of criterion information, are still imperfect criterion measures. This is because judges' decisions about a defendant's competency status are subject to the same limitations in judgment and decision making as those offered by clinicians (Mossman, 2008).

As of approximately ten years ago, little data were published on the validity of competence to stand trial decisions made by examiners, or even on examiner agreement regarding the standards of practice in this area (Grisso, 1996; Melton et al., 2007; Roesch, Hart, & Zapf, 1996). The first published study to evaluate examiner's accuracy of competence to stand trial predictions occurred in 1984 and focused on restoration of competency predictions (Cuneo & Brelje, 1984). For defendants initially found incompetent to stand trial, the ability of clinicians to predict restoration of competency within one year indicated a true positive rate of 77% (correctly predicting restoration), with 23% false positives (incorrectly predicting restoration), 100% true negatives (correctly identifying those who would not be restored), and 0% false negatives (mistakenly identifying those who would not be restored). The research highlighted the examiners' tendency to err on the side of overpredicting competence restoration.

In their meta-analysis, Nicholson and Kugler (1991) issued a call to arms for clinicians to conduct research on the validity of competence to stand trial assessments. However, since the time of their writing, few studies have examined this area of judgment and decision making for competency to stand trial evaluations. Mossman et al. (2010) used latent class modeling (LCM) and receiver operating characteristic (ROC) curves to determine the accuracy of competency to stand trial evaluators. Five board-certified psychiatrists made decisions regarding the competency of 156 defendants depicted in redacted court reports. The reports depicted actual defendants previously referred for competency to stand trial evaluations. Participants were asked to rate the defendant's competency according to the criteria outlined in Dusky as well as to provide separate ratings on 3 specific prongs of Dusky: understanding, appreciation, and reasoning abilities. The average accuracy of participants on ratings of overall Dusky criteria was 0.967. That is to say, for a randomly chosen defendant, 96.7% of the time a participant would

assign a higher score (reflecting a decision of incompetence) to a defendant who truly was incompetent to stand trial. Interestingly, participants' combined ratings of the defendant's understanding, appreciation, and reasoning abilities demonstrated lower accuracy than decisions about whether the defendant met overall Dusky criteria for competency to stand trial.

The study by Mossman et al. (2010) is useful because it provides information about the ability of evaluators to accurately classify defendants when using criteria outlined in Dusky. It also provides support for assessing competency to stand trial using a continuous scale as opposed to categorical determinations. However, the ultimate legal opinion of competence to stand trial still requires a dichotomous decision. Mossman et al. also recognize that they may be measuring agreement among participants (reliability) instead of accuracy (validity). Further, this study does not provide information about the accuracy of the participants' decision making process, such as the variables used by the participants when making these decisions, or information about participants' accuracy when making dichotomous decisions about competency. In addition, the subjects were all board-certified psychiatrists and it is unclear how these results would generalize to other clinicians faced with similar decisions.

Due to the inherent difficulties in assessing the validity of examiners' decisions and the absence of a gold standard, research has largely examined the quality of forensic reports, based on the adherence to guidelines developed by researchers and legal statutes. Many of these studies have found problems with reports, including a failure to explicitly state the criteria used to make determinations, failure to mention confidentiality limitations, inadequate patient history, and little to no connections made between mental health and impairments of psycho-legal abilities (Christy, Douglas, Otto, & Petrila, 2004; Mumley, Tillbrook, & Grisso, 2003; Robinson & Acklin, 2010).

Previously discussed were studies that attempted to assess the reliability and validity of competency to stand trial assessments. Other important areas in the clinical decision making literature include the variables used by clinicians to make judgments, the importance of these variables in making decisions, and the decision making process itself. To date, two meta-analyses have been conducted to examine the variables used by clinicians to make judgments. Other studies have attempted to examine the impact of these variables on clinician decision using vignettes. More recent studies have been devoted to examining the impact of system factors on competency to stand trial decisions.

Factors Involved in Competency to Stand Trial Evaluations

A primary concern is the potential for a determination of competency to be based on extraneous variables, such as irrelevant defendant characteristics. Numerous studies have examined the impact of sociodemographic, criminological, and psychological variables on competency decisions and have provided the field with conflicting information. For example, some studies have reported a relationship between criminal charge and findings of competence (Rosenfeld & Ritchie, 1998), while other studies have attempted to show that the relationship is mediated by diagnosis (Cochrane, Grisso, & Frederick, 2001) and still others have shown no correlation whatsoever (Cooper & Zapf, 2003). Two meta-analyses have attempted to solve these discrepancies through combined analyses of the literature.

Meta-Analyses

A meta-analysis by Nicholson and Kugler (1991) examined 30 studies published between 1967 and 1989 that compared defendants found incompetent with those found competent. Although the authors found that individuals who were older, female, an ethnic minority, or not married were more likely to be deemed incompetent, the correlation was small and accounted for

less than 1% of the variance in the competency decision. Previous legal involvement and hospitalization history were significantly correlated with incompetency. Individuals without a legal history or individuals with a history of previous psychiatric hospitalization were more likely to be found incompetent. In addition, a diagnosis of psychotic disorder was so highly correlated with incompetency that it was the highest correlation in the study. The authors examined psychiatric symptoms and found that disorientation, delusions, hallucinations, disturbed behavior, impairment in memory, and impairment in thought or communication were all significantly correlated with decisions of incompetency. Educational achievement, employment status, type of offense (violent or non-violent), affective disturbance, and impaired judgment were not correlated with the decision. Poor performance on assessment measures designed to assess competency, presence of psychotic disorder, and psychiatric symptoms reflecting serious mental illness were the most strongly correlated to determinations of incompetency. The base rate of incompetency was found to be 30.6% across the studies.

The statistical methods used by Nicholson and Kugler (1991) were criticized by Pirelli et al. (2011) for their reliance on correlation coefficients, despite the low base rate of incompetency. Pirelli et al. argued that the low base rate requires using statistical techniques that are less influenced by rare events, such as odds ratios. In their meta-analysis, Pirelli et al. examined 68 studies published between 1967 and 2008 using odds ratios. They found that the most significant defendant characteristics that predicted incompetence were diagnosis, hospitalization history, and employment status. A diagnosis of a psychotic disorder resulted in an eight fold increase in the likelihood of being found incompetent to stand trial when compared to those without a psychotic disorder diagnosis. Findings of incompetence were double for defendants who had a history of psychiatric hospitalization as well as for those who were

unemployed. No differences existed based on current criminal charge, gender, and prior history of competency evaluations. However, the confidence interval for prior history of competency evaluation was so large that the authors cautioned against making conclusive statements about the seemingly irrelevant nature of it on competency decisions. Ethnicity and marital status were minimally different with non-white being 1.39 more times likely and unmarried 1.43 times more likely to be found incompetent. From these results, it seems clinicians' opinions are impacted by mental health symptoms, which is what one would hope to observe.

Unfortunately, the authors did not assess conditional or joint probabilities so it is not clear how much the variables add to the ability to predict or whether they are interrelated. Specifically, the possibility of multicolinearity among psychotic disorders, marital status, employment status, and hospitalization history should be examined. Although the study by Warren et al. (2006) found no multicolinearity among the demographic, criminological, and clinical variables they reviewed, including diagnosis and hospitalization history, employment status was not a variable they used. On the other hand, a study by Cooper and Zapf (2003) found that multicolinearity existed among all the demographic, criminological, and clinical variables they reviewed, except for diagnosis and employment.

To examine the impact of study characteristics on the findings, Pirelli et al. (2011) examined the influence of six potential moderators: type of publication, source of decision, setting, country, type of participants, and recruitment method. For ethnicity, non-white defendants were more likely to be found incompetent than white defendants in studies that used defendants whose competence was never in question (e.g., inmates) than studies that used defendants referred for competency evaluations and defendants restored to competency. In addition, non-white defendants were more likely to be found incompetent than white defendants

in studies that used prospective methods of data collection than those using archival and retrospective methods. The relationship between gender and competency decision was significantly different depending on type of publication, although this difference was largely due to the small female sample size used in Roesch and Golding's (1980) data. The country of the study also affected this relationship. For example, females were more likely to be found incompetent than males in studies performed in Canada, while studies performed in the United States found no difference. The relationship between decisions of competency and psychiatric diagnosis was affected by the type of participants found in the studies. For example, defendants were more likely to be found incompetent after receiving a diagnosis of psychotic disorder in studies using defendants whose competency was never in question than in studies using defendants referred for competency evaluation, defendants restored to competence, and those studies whose participants were put in the "mixed" category by Pirelli et al. The impact of the six potential moderators could not be examined for variables that had few studies (psychiatric hospital and evaluation history) or variables that had limited variability within moderator groups (employment). Thus, moderator analyses were not conducted for these variables. For other variables, moderator analyses were not conducted because there was no significant heterogeneity.

In 2006, Archer, Buffington-Vollum, Stredny, and Handel surveyed forensic psychologists to identify test use patterns for a variety of forensic evaluations including competency to stand trial evaluations. The frequency of test use was rated on a 7-point Likert scale ranging from never (0) to always (6) with each rating assigned a numerical weight. Weighted scores were calculated for each test by multiplying the number of responses for that rating with the numerical weight assigned to the rating. Measures used during competency to

stand trial evaluations in order of the most frequently used to the least frequently used include: MacArthur Competence Assessment Tool- Criminal Adjudication (MacCAT-CA), Competence Assessment for Standing Trial for Defendants With Mental Retardation (CAST-MR), Evaluation of Competency to Stand Trial- Revised, Grisso's Miranda Rights, Roger's Criminal Responsibility Assessment Scales, Georgia Court Competency Test, MacCAT Treatment, Competency Screening Test, Interdisciplinary Fitness Review Interview (IFI and IFI-Revised), and the Fitness Interview Test (FIT and FIT-Revised). They found that among the 152 respondents, 17.4% reported expertise in conducting competency evaluations, a higher percentage of individuals than in any other area reported among the participants.

One way to identify the variables involved in competency to stand trial evaluations is to use vignettes in lieu of actual defendant information. This methodology is useful because it allows the researchers to make inferences of causation among the variables. Some have challenged the use of this approach, with questions related to the psychometric properties of the vignettes (Schacht, 2005). Other arguments have focused on attacking the ability to generalize these results to actual clinical situations. To this debate, Dawes (1994) adds his belief that when clinicians make judgments based on cues, these judgments are "essential components of more ecologically common evaluations" (p. 97).

To date, few studies have been published that use vignettes to identify causal factors influencing competency to stand trial decisions. In the next section, three of these vignette studies and their limitations will be discussed. The first study examines the variables used to make decisions about competency, the second examines the impact of irrelevant information on these decisions, and the third examines the impact of legal criteria on these decisions.

Vignette Studies

A decade after Cuneo and Brelje's (1984) research on validity of competency restoration, researchers Blashfield, Robbins, and Barnard (1994) asked psychiatrists to make competency determinations based on hypothetical vignettes and found similar results. The vignettes contained two groups of information related to a) the defendant's cognitive status, presence of psychotic features, courtroom behavior, ability to understand the legal system and b) relationship with the lawyer, history of drug and alcohol use, psychiatric history, and criminal history. Like Cuneo and Brelje, the researchers found that clinicians erred toward finding defendants competent. For those case histories set-up to depict an incompetent defendant, roughly one-quarter of participants made a determination of competency, while the reverse did not occur. In addition, no participants found all four hypothetical defendants incompetent while nearly one-fifth found all four defendants competent. Psychiatrists confirmed this tendency through self-report.

The researchers also found that the second group of variables had a significant impact on the competency determination, but not to the extent of the first group variables. Interestingly, for those cases in which the defendant was incompetent related to the first group of variables but competent related to the second group, respondents were equally split on determinations (between 44 and 47% of respondents found defendant competent). This split did not occur when the defendant was competent for the first group of variables and incompetent for the second group (in this case, approximately 94% of respondents found the defendant competent). Although this suggests that there are circumstances in which clinicians do not agree, because the variables were not manipulated independently, it is not known which variables were the most influential in these determinations. Self-report indicated that participants who decided the defendant was incompetent relied primarily on the mental status, ability to understand the legal

system, and courtroom behavior, while those who found the defendant competent used these variables as well as the relationship with the lawyer and the charges to make the decision.

The tendency to include irrelevant information when making competency determinations was documented in a study by Plotnick, Porter, and Bagby (1998). The researchers presented Canadian psychiatrists with eight hypothetical case vignettes that varied in legally relevant variables (ability to understand, assist, and presence of a mental disorder) and legally irrelevant variables (previous hospitalization, current crime, and prior legal history). By manipulating the relevant variables, the researchers created four fitness conditions: fit, unfit, and two that were questionable (the defendant could understand but not communicate or vice versa). The legally irrelevant variables were manipulated to create two conditions within the fitness conditions that biased participants toward fitness or unfitness. For example, *relevant* legal variables would be manipulated to create 2 fitness conditions that portrayed the defendant as fit. Then, for one of these 2 conditions, the *irrelevant* legal variables would be manipulated to bias the respondent toward finding the defendant fit while the other conditions would be biased toward finding the defendant unfit. Participants were asked to categorize the defendant as fit, questionable, or unfit, then to provide a Likert-scale rating of the defendant's fitness (ranging from definitely unfit to definitely fit), and defendant's need for continued hospitalization. Overall, participants were significantly able to classify the vignettes according to the predetermined conditions (75% accuracy for fit condition and 71% accuracy for unfit), with the ambiguous conditions being more difficult for respondents (54% accuracy for the Unable to Communicate condition and 57% accuracy for the Unable to Understand condition). Results suggested an interaction effect between the fit and bias conditions. That is, for participants in the fit conditions, information beyond the psycho-legal criteria influenced their decisions causing their accuracy rate to

decrease (from 80% accuracy in the bias toward fit condition down to 60% accuracy in the bias toward unfit condition).

By examining participant ratings of fitness, the researchers determined that participants followed the legal criteria in that they rated the fit conditions as most fit, followed by the unable to communicate condition, then the unable to understand condition, and lastly the unfit condition was rated as the least fit. It also shed light that psychiatrists viewed the ability to understand as more critical than the ability to effectively communicate as higher ratings of fitness were given to those who were not able to communicate than to those who could not understand. Also examined was the participants' preference for additional information related to the defendant's psychiatric history, current criminal charge, and legal history. Requests were significant for information about defendant's psychiatric history and current charge but not for legal history. These requests were significant for the questionable conditions, where the defendant was either unable to understand or unable to communicate, and not significant for the fit and unfit conditions. The researchers suggest that this implies that psychiatrists may be tempted to use information beyond the legal criteria in ambiguous or unclear cases.

In the study by Morris et al. (2004), the researchers attempted to examine the impact of legal criteria on clinical judgment. Participants consisted of board certified forensic psychologists and psychiatrists who were presented with two vignettes. The first vignette depicted a defendant whose thinking was impaired but whose behavior was intact while the second vignette depicted a defendant whose thinking was intact but ability to control behavior was impaired. Participants were asked to make determinations of competency according to three differently worded legal criteria, one of which emphasized rational understanding, the other rational behavior, and the last was the Dusky standard. Results indicated that the legal standard

did not impact clinicians' decisions. More than three quarters of participants reached identical conclusions regardless of the legal standards they were asked to employ. In addition, participants were about equally split in their decisions for the first vignette, while the defendant in the second vignette was found competent 60 to 70 percent of the time regardless of the legal criteria. Unfortunately, the researchers only reported percentages and therefore no conclusions can be made regarding the degree of difference. In addition, the study sample consisted mostly of psychiatrists.

From these studies, it is clear that information is weighted differently by clinicians, clinicians can be influenced by factors beyond the legally relevant factors, and clinicians may not take the legal criteria into consideration when making decisions about competency to stand trial. The results of the vignette studies suggest that other factors may be playing a role in clinicians' determinations, factors that lay outside the defendant's psycho-legal abilities, such as irrelevant defendant characteristics. This and other research has raised concerns about the inconsistency with which clinicians make determinations for competency to stand trial (Dawes, Palmer, & Jeste, 2008). Another interesting and fairly recent line of research has examined the role of systemic factors on decisions of competency.

Systemic Factors

Previous studies, like those by Murrie et al. (2008), have attempted to examine the impact of clinician factors on determinations of competency but have failed to control for the influence of system factors, such as differences in institutional setting. Few within-state studies have been conducted to control for this variance despite research that has demonstrated the impact of the institutional setting on clinical opinions. For example, a study by Warren, Rosenfeld, Fitch, and Hawk (1997) examined clinical opinions by evaluators in three states with nearly identical legal

standards but different evaluation systems for service provision (inpatient v. outpatient). The results indicated statistically significant differences in rates of incompetency across all three states, from 13% found incompetent in the Virginia sample with up to 29% found incompetent in the Ohio sample. However, only the Virginia sample allowed the researchers to compare differences in clinical opinion based on setting (outpatient v. inpatient) and a statistically significant difference was found, with defendants evaluated in inpatient settings being less likely to be found incompetent than defendants evaluated in outpatient settings (6% vs. 17%).

Otto et al. (1998) hypothesized that clinicians in outpatient forensic facilities would have a lower threshold for making incompetency to stand trial determinations than clinicians in inpatient facilities. The authors examined scores on the MacArthur Competence Assessment Tool- Criminal Adjudication (MacCAT-CA) for defendants found incompetent to stand trial in eight different states. The results showed no differences in MacCAT-CA scores for defendants evaluated incompetent to stand trial across evaluation settings. That is, defendants evaluated by clinicians in private practice and deemed incompetent to stand trial did not have lower scores on the MacCAT-CA than defendants evaluated in inpatient facilities and deemed incompetent to stand trial. However, the researchers did discover that within each state, hospitalized defendants evaluated for competency were more impaired than jail inmates, as evidenced by significantly lower MacCAT-CA scores.

The underlying assumption of this and other research that examines the decision making process of clinicians is that clinicians are accurate in their decisions. This is not to say that this research is unimportant, as the information obtained through this research can inform clinicians about factors outside of the defendant's characteristics that may affect the decision making process. However, before making conclusions about the impact of certain factors on decisions of

competency, it is essential to first confirm that these decisions are accurate. Although it has been

recognized that quantifying clinicians' accuracy in decisions of competency is inherently

difficult, methods do exist for examining these questions. One way to examine this is through the

use of latent class modeling techniques described by Mossman (2010). Another approach is to

use the Brunswik Lens Model in conjunction with outcome information from court documents.

Research on clinical decision making and judgment in the context of forensic evaluations using

the Brunswik Lens Model, with court records as the criterion, can illuminate this discrepancy.

Hypotheses

The correct identification of incompetent defendants ensures the protection of

individuals' constitutional rights to a fair trial between matched adversaries, where the accused

can confront the accusers, present evidence, and assist counsel in their defense (Grisso, 1988;

Mossman et al., 2007). However, the misclassification of a competent defendant as incompetent

threatens these and other rights guaranteed to the people including the right to a speedy trial and

equal protection, as individuals found incompetent are subject to more lenient commitment

criteria and more stringent release criteria than their civil counterparts (Foote, 1960). Therefore,

it is essential that clinicians form their opinions based on relevant variables in order to perform

reliable and valid evaluations of defendants waiting to stand trial.

It is hypothesized that variables identified as having predictive power in the literature will

significantly predict decisions of competency rendered by the legal system. Namely, psychiatric

diagnosis, psychiatric hospitalization history, and employment status will be significant

predictors of the decisions of competency rendered by the courts. Further, it is hypothesized that

these same variables (psychiatric diagnosis, psychiatric hospitalization history, and employment

status) will significantly predict the raters' decisions of competency to stand trial. In addition, it

is hypothesized that linear models of judgments will outperform human models of judgments. It is hypothesized that experience will have no impact on raters' accuracy, with expert raters and novice raters displaying equal levels of accuracy. It is predicted that clinical raters (experts and novices) will outperform legal raters (lawyers) by achieving higher levels of accuracy. Lastly, it is hypothesized that raters' perceived level of confidence will be unrelated to accuracy.

CHAPTER 3

METHODS

Participants

Defendants

Information for the protocols was taken from the medical records of male and female

defendants admitted to a secure inpatient mental health treatment facility for competency

assessment, competency restoration, and/or treatment. Referrals to the facility are made through

the Pennsylvania legal system from 48 counties. The defendants selected were required to have a

forensic discharge summary or forensic summary in their medical record to ensure evaluations of

competency would be available for review. Further, defendants selected for inclusion were

required to have the results of competency-specific assessment instruments (e.g., MacArthur

Competence Assessment Tool for Criminal Adjudication or the Competence Assessment to

Stand Trial for Defendants with Mental Retardation assessment instruments) available in their

medical record. The inclusion of assessment data provided raters with objective information on

which to base their decisions. In addition, the defendants selected were required to have

information regarding their most recent competency status.

A list of defendants evaluated by the psychology department for competency to stand

trial was created using the hospital's transcription data base. To identify defendants evaluated by

the psychology department for competency to stand trial, a search of the hospital's transcription

data base was conducted for the words "MacArthur" and "Competence Assessment." This

ensured that only defendants with the results of competency-specific assessment instruments

would be included in the research project. A list of defendants who met the inclusion criteria was

created and was cross-referenced with the current hospital census to identify defendants who had

not been discharged and thus would not have a discharge or forensic summary available in their medical record. Thirty-two defendants were identified who met criteria for inclusion. Medical records were obtained for each defendant.

The sample of defendants consisted of 17 males and 15 females. Fifty percent were Caucasian and 50% were African American. Approximately 81% were single, approximately 7% were married, and approximately 13% were married but were either divorced or widowed (the marital status of one defendant was unknown). Of the 29 defendants for whom employment information was obtained, approximately 90% were unemployed and approximately 10% were employed. Defendants ranged in age from 20 to 66 years (M = 39.78, SD = 12.77). The number of felonies ranged from 0 to 14; misdemeanors from 0 to 8; and summary offenses from 0 to 8. The number of previous hospitalizations ranged from never having been hospitalized to over 100 hospitalizations, although this number is likely a misrepresentation. All defendants had at least one mental health diagnosis. Nineteen defendants were diagnosed with a psychotic disorder, nine with a mood disorder (five of these were considered severe with psychotic features), 22 with a substance use disorder, 15 with a personality disorder, ten with mental retardation, and six with a diagnosis in the "other" category (e.g., cognitive disorder, posttraumatic stress disorder, impulse control disorder, anorexia nervosa, adjustment disorder, or intermittent explosive disorder). Approximately 84% of the defendants had more than one diagnosis. (This information is summarized in Table 1).

Table 1

Defendant Characteristics

Characteristic	*n*	*%*	Mean	Range
Age	32		39.8	20 - 66
Gender	32			
% Male		53.1		
% Female		46.9		
Ethnicity	32			
% White		50.0		
% African American		50.0		
Marital Status	31			
% Single		80.6		
% Married		6.5		
% Divorced		6.5		
% Widowed		6.5		
Employment	29			
% Employed		10.3		
% Unemployed		89.7		
Hospitalizations	32		10.6	0 - 100
Psychiatric Diagnosis	32			
% Psychotic Disorder		59.4		
% Mood Disorder		28.1		
% Substance Use Disorder		68.8		
% Personality Disorder		46.9		
% Mental Retardation		31.3		
% Other Diagnosis		18.8		
Intelligence Quotient (IQ)	29		73.7	48 - 118
Criminal Offense	32			
Felonies			2.3	0 - 14
Misdemeanors			1.7	0 - 8
Summary Offenses			.8	0 - 8

Defendant performance on different psychological tests was also examined and is summarized in Table 2. Mini Mental Status Examination (MMSE) scores were used to gain an understanding of defendant's mental state and was available for 20 defendants. Scores ranged from 16 to 30 ($M = 25.90$, $SD = 3.85$), suggesting that some defendants were perhaps experiencing problems with orientation and memory. For the 12 defendants without MMSE

information, their mental status came from narrative descriptions of their symptoms which often

included information about orientation. Of these defendants, none were severely impaired and

only one was moderately impaired given his lack of orientation to time. Orientation information

was unable to be determined for two defendants. Estimates of intelligence were available for 29

defendants and ranged from 48 to 118 ($M = 73.66$, $SD = 19.28$).

Despite only ten defendants receiving diagnoses of mental retardation, 21 defendants

were administered the Competence Assessment for Standing Trial for Defendants With Mental

Retardation (CAST-MR) and 11 were administered the MacArthur Competence Assessment

Tool- Criminal Adjudication (MacCAT-CA). The CAST-MR comprises three indices. Basic

Legal Concepts assesses an individual's knowledge of courtroom personnel and legal terms and

scores range from 0 to 25. Skills to Assist Defense uses hypothetical scenarios to examine an

individual's ability to assist their lawyer, with scores ranging from 0 to 15. Understanding Case

Events examines the individuals' ability to describe the events related to the alleged crime and

scores range from 0 to 10. These indices combine to form a Total Score (ranging from 0 to 50),

which provides an overall rating of an individual's competency to stand trial, with scores above

30 indicating competency (Everington & Dunn, 1995). The MacCAT-CA also comprises three

indices but the indices do not combine to provide an overall rating of competency.

Understanding assess the individual's ability to understand courtroom personnel and legal terms

and scores range from 0 to 16. Reasoning examines the individual's ability to reason, make

decisions, and weigh evidence and scores range from 0 to 16. Appreciation examines the

individual's knowledge of his/her own situation related to the legal proceedings with scores

ranging from 0 to 12 (Otto et al., 1998).

Total score on the CAST-MR ranged from 21 to 48 and performance on the three prongs of the MacCAT-CA ranged as follows: Understanding (6 – 16); Reasoning (4 – 16); Appreciation (2 – 12). The ultimate competency to stand trial disposition rendered by the court was available for 25 defendants, with five found incompetent and 20 found competent. For the remaining seven defendants, the psychiatrist's opinion was used in substitution for that of the court's decision. In these cases, three were found incompetent and four were found competent. Combining these decisions, 25% of the defendants were considered incompetent and 75% were considered competent, which closely resembles the base rate estimated by Pirelli et al. (2011).

Table 2

Defendant Test Performance

Test	*n*	Mean	Range
Mental Status			
MMSE	20	25.9	16 - 30
Intelligence			
WASI	5	85.4	66 - 109
WAIS-IV	20	64.3	48 - 100
Shipley	4	105.8	95 - 118
Competency			
CAST-MR			
Section 1	20	21.2	5 - 25
Section 2	20	9.9	5 - 15
Section 3	20	7.2	3 - 10
Total	21	38.4	21 - 48
MacCAT-CA			
Understanding	11	10.8	6 - 16
Reasoning	11	11.6	4 - 16
Appreciation	11	8.8	2 - 12

Note. MMSE = Mini Mental Status Examination; WASI = Wechsler Abbreviated Scale of Intelligence; WAIS-IV = Wechsler Adult Intelligence Scale – Fourth Edition; Shipley = Shipley Institute of Living Scale; CAST-MR = Competence Assessment for Standing Trial for Defendants With Mental Retardation; MacCAT-CA = MacArthur Competence Assessment Tool- Criminal Adjudication.

Raters

Solicitation letters were mailed to 48 graduate students enrolled in the Indiana University of Pennsylvania Clinical Psychology Program (please see Appendix C for a sample letter). Solicitation letters were mailed to 18 members of the Pennsylvania Psychological Association's Forensic and Criminal Justice Committee who held advanced degrees in psychology. As a result of the policies for using the American Psychology-Law and Society (APLS) membership list, solicitation letters were electronically mailed to all 2,000 members of APLS. Solicitation letters were also mailed to 150 individuals with advanced degrees in law (J.D.) who were selected using the Pennsylvania Bar Association (PBA) and the National Association of Criminal Defense Lawyers membership lists. Individuals were randomly selected from the PBA. For the National Association of Criminal Defense Lawyers, the first 75 individuals from the state of Pennsylvania with contact information available were selected.

A solicitation letter, a consent form, and a prepaid mailing envelope were mailed to potential participants (please see Appendix D for the consent form). Potential participants were also provided with a copy of the consent form and were asked to return the signed consent form using the prepaid envelope within a three week time period. For the psychologists solicited through the American Psychology-Law Society, an email containing the body of the letter was sent to potential participants. The consent form and a sample protocol were included as attachments. Interested participants were instructed to reply with their mailing address so the consent form could be mailed along with the prepaid mailing envelope.

A total of 19 psychologists, 12 graduate students, and 9 lawyers agreed to participate in the research project. The initial goal of this research project was to obtain three participants from each group. Participants were randomly selected by blindly selecting three consent forms from

each group of forms (e.g., psychologists, graduate students, or lawyers). These participants were provided with the materials. Given the larger than anticipated response rate, permission was sought from the Indiana University of Pennsylvania Institutional Review Board to increase the sample size in the research project. The sample size was expanded to nine participants per group because that allowed an equal number of participants to be selected from each group. All the lawyers who were recruited participated in the research project and six additional psychologists and six additional graduate students were randomly selected and mailed materials.

The sample of participants consisted of expert raters, novice raters, and legal raters. Expert raters consisted of nine individuals with advanced degrees (Ph.D. or Psy.D.) in clinical psychology and experience conducting competency to stand trial evaluations. Expert raters included psychologists from Pennsylvania, Nevada, California, Oregon, Texas, Indiana, Connecticut, and Puerto Rico. The novice raters consisted of nine doctoral-level clinical psychology graduate students with no prior training or experience in competency to stand trial evaluations, but with experience in psychological evaluation. All graduate students came from the same doctoral program located in Pennsylvania. Legal raters consisted of eight lawyers with advanced degrees (J.D.) in law. All legal raters had law offices in Pennsylvania. Nine lawyers initially consented to participate in the research project. However, one lawyer returned the binder without completing the task. He reported s/he could no longer participate in the research project.

Materials

The hospital's transcription data base was used to create a list of defendants who met criteria for inclusion in this research project. This list was cross-referenced with the hospital census. To ensure the confidentiality of each defendant, data taken from defendants' medical records were entered into a password-protected Excel spreadsheet and stored on a flash drive. As

an added layer of protection, the names of the defendants were replaced in the Excel spreadsheet with the code "Defendant" followed by a number (e.g., Defendant 1, Defendant 2, etc.) and the specific legal charges were replaced with an offense category (e.g., homicide would be noted as "felony"). A list of the defendants' names and the assigned codes (e.g., Defendant 1, Defendant 2, etc.) was stored on a separate password-protected Excel document that was only used during data collection, never left the hospital, and was destroyed after data collection was completed.

The criterion information collected from defendants' court records was stored in a separate password protected file on a flash drive. The criterion information was not included in the protocols. In order to identify the criterion while also protecting defendant confidentiality, the Pennsylvania Judiciary Web Portal was used to identify the case number assigned to each defendant's legal case. This case number was used to access each defendant's public court record instead of using the defendant's name. Public court records were accessed online (using the Pennsylvania Judiciary Web Portal), by mail, by phone, and in person.

Using the data that were entered into the Excel spreadsheet, thirty-two protocols were created. Protocols displayed information related to each defendant's ethnicity, age, gender, marital status, employment status, symptoms, psychiatric diagnosis, hospitalization history (to include number and length of stay), psychological test data (results of mental status examinations, intelligence quotients, and competency-specific assessment results), and category of offense in a one page summary. Protocols were absent of defendant names and other identifying information and included the statement "All information that may reasonably be used to identify someone has been removed" at the top of each protocol. When information about defendants' symptoms contained terms regularly used and understood by mental health professionals, but perhaps not readily understood by other professionals, these terms were further

defined in parentheses (e.g., "…affect [emotional expression] is constricted"). Psychiatric

diagnosis was displayed using a multi-axial system represented by Axis I, Axis II, and Axis III

information. Only medical diagnoses that could impact a defendant's mental health (e.g.,

hypothyroidism or traumatic brain injury) were included on Axis III; all other medical diagnoses

were redacted from the protocols. Psychological test data included: Full Scale Intelligence

Quotient standard scores for the Wechsler Adult Intelligence Scale – Fourth Edition, Wechsler

Abbreviated Scale of Intelligence; Shipley Institute of Living Scale (presented as estimated FSIQ

standard scores); overall score on the MMSE, and scores on the Understanding, Reasoning, and

Appreciation scales of the MacArthur Competence Assessment Tool for Criminal Adjudication

or scores on the Basic Legal Concepts, Skills to Assist Defense, and Understanding Case Events

sections and Total Score achieved on the Competence Assessment to Stand Trial for Defendants

with Mental Retardation. Next to each assessment instrument, the standard score interpretation

was given. For example, for intelligence measures, the following description was used to help

raters make sense of the intelligence quotient: "≥ 110 = high average to very superior intellectual

functioning; 90-109 = average; ≤ 89 = low average to severely impaired." Please see Appendix A

for a sample protocol.

 The 32 protocols and a two-page list of instructions were placed in a three-ring binder for

ease of review. The instructions included information about the base rate of incompetency to

stand trial, the legal criteria for determining competency to stand trial, instructions for

completing the task, and a table depicting offense categories and examples of potential sentences

associated with each offense category (please see Appendix B for the instructions).To maintain

the anonymity of each rater, while also grouping the raters according to experience, the three-

ring binders were marked according to the intended recipient using a piece of tape (e.g., "P" for

psychologist, "G" for graduate student, or "L" for lawyers). Binders were also color-coded (psychologists had blue, graduate students had red, and lawyers had black binders) in the event that the tape fell off.

Solicitation letters were created that asked potential participants for their voluntary participation in the research project. Other materials used include informed consent forms and mailing materials, such as mailing envelopes and stamps.

Procedures

Creation of Protocols

The key variables of interest, or data cues, extracted from defendants' medical records for inclusion in the protocols were chosen because they were identified in the literature on competency to stand trial evaluations as having predictive value or because they represented information available to evaluators when making decisions of competency. The criterion, whether a defendant was ultimately found competent or not competent to stand trial by the court, was identified from the defendant's medical record or through public court records. Documents dated after the date of the competency evaluation were reviewed to determine the court's determination of the defendant's competency.

Ideally, defendants would have a court order in their file declaring the defendant competent or not competent to proceed to trial. Unfortunately, this information was not regularly available so alternate methods were used to determine competency status. Documents used to determine competency status included court orders, sentencing orders, Behavior Clinic clearance status, the signature pages of various plea documents (e.g., Guilty Plea Petition form, Guilty Plea – Explanation of Defendant Rights, Negotiated Guilty Plea – Explanation of Defendant Rights, etc.), Commonwealth Petition for Nolle Prosse, motions for postponements, hearing disposition

reports, and Guideline Sentence Form. If the court's opinion as to the defendant's competency status was not clearly stated in a court order, it was assumed that if the person proceeded to trial, such as by attending a hearing or entering a plea, the court considered the person competent to stand trial. This was also used if the charges were withdrawn and the defendant pled guilty to a lesser charge (as occurs with plea bargains). There were seven cases in which the defendant's competency status was unable to be determined using these methods. One case was ongoing, one case was inactive because the defendant could not be located, one case was nolle prossed due to lack of evidence, in three cases charges were withdrawn for reasons unrelated to the defendant's competency or for unknown reasons, and one case did not have an updated outcome. For these seven cases, the psychiatrist's opinion was used as an estimate of the ultimate opinion by the court. Given the high agreement rate between opinions rendered by evaluators with those rendered by the court (96.3% to 99.7%), it was assumed that the psychiatrist's opinion would accurately reflect that given by the court (Hart & Hare, 1992; Zapf et al., 2004).

Rating Protocols

Using the address provided on the consent form, a binder was mailed to each participant along with a request for delivery confirmation by the United States Postal Service. Also included with the binder was a prepaid mailing envelope. For each protocol, participants were asked to render a categorical decision of competency by marking a line next to the statement "competent" or "not competent." For each protocol, participants were also asked to indicate their confidence in this decision by circling a number on a five-point scale (one being not at all confident and five being extremely confident). After four weeks, the participants were contacted and asked to return the three-ring binder using the prepaid mailing envelope if they had not already done so. All binders were returned. After the binders were collected, a debriefing letter was mailed to each

participant and the binders were removed from the mailing envelopes to prevent future

identification of the participant (please see Appendix E for the debriefing letter). Due to the

small sample size of raters and the attempt to maintain the anonymity of each rater, less

information on individual rater characteristics was collected.

CHAPTER 4

RESULTS

Coding Procedures

Nine psychologists, nine clinical psychology graduate students, and eight lawyers made

decisions of competency to stand trial and rated their confidence in these decisions for 32

protocols that depicted demographic, psychological, and criminological information for actual

defendants. The variables used in the protocols were taken from an archival data source and there

were portions of missing data for several of the variables and outliers were noted for defendants'

hospitalization histories.

Missing data included marital status, employment, and performance on measures of

intelligence and orientation. There were only nine missing data points across the eleven

variables, meaning that 2.5% of the total sample was missing. It was assumed that the missing

data were missing completely at random. Within the variables, 3% of marital status information,

9% of employment information, 6% of orientation information, and 9% of intelligence

information were missing. Rubin, Witkiewitz, St. Andre, and Reilly (2007) compared the impact

of mean substitution, regression, and expectation maximization approaches to addressing missing

data. The authors concluded mean substitution, regression, and expectation maximization

approaches were all suitable procedures to use when 1-5% of the data are missing. Therefore, the

single imputation method of mean substitution was used to address missing data for intelligence

and the mode was used for categorical variables with missing data (employment, marital status,

and orientation).

A new variable entitled "orientation" was created that allowed orientation information as

well as mental status information to be summarized as either "impaired" or "mildly impaired to

unimpaired." Based on their score on the mini mental status examination defendants were

categorized as "impaired" or "mildly impaired to unimpaired" on orientation. For defendants

without mini mental status examination information, the information provided in the medical

records was used to make a determination of their orientation. Using this method, there were

only two defendants for whom orientation could not be determined. For these defendants, the

median (e.g., mildly impaired to unimpaired) was used as a substitution.

A new variable was created to capture marital history. Defendants were originally coded

as single, married, divorced, or widowed. These categories were collapsed to create a new

bivariate variable: single or not single. Defendants who were single, divorced, or widowed were

categorized as single and defendants who were married were categorized as not single.

There existed several outliers for hospitalization history. For example, one defendant

reported to have been hospitalized over 100 times despite being 21-years old. It was assumed this

was either an error in the medical record or an exaggerated estimate by the defendant. Further,

some defendants estimated they had been hospitalized "several" times and an exact number was

not provided in the medical record. To address these issues, Visual Binning (found in SPSS) was

used. With Visual Binning it is possible to transform a continuous variable into approximately

equal groups using cut-off points. This is a transformation technique suggested by Pallant (2007)

to create groups from a continuous variable. Using this transformation technique, the defendants

were split into three approximately equal categories: those who had been hospitalized one time

or less, two to five times, and six or more times. Defendants with "several" previous

hospitalizations were included in the "two to five" category.

New variables were created to capture defendants' diagnosis and criminal offense.

Following methods used by Nicholson and Kugler (1991) to analyze the impact of diagnosis,

defendants were categorized as either having a psychotic disorder or having a non-psychotic disorder. A defendant was identified as having a psychotic disorder if they were diagnosed with a psychotic disorder (e.g., Schizophrenia, Psychotic Disorder not otherwise specified) or if psychosis was part of their diagnosis (e.g., Depression with psychotic features). For criminal offense, if a defendant had at least one felony charge, the criminal offense was categorized as felony. If no felony charges were present the defendant was categorized has having no felony.

A variable entitled Competence Test Performance was created to capture defendants' performance on measures used to assess competence to stand trial. The CAST-MR and the MacCAT-CA do not use the same scale so performance on these measures was categorized nominally using an interpretive strategy. The CAST-MR yields a total score and Everington and Dunn (1995) suggest using a cut-off score of 30 to identify a competent defendant. Thus, defendants falling below 30 were classified as severely impaired and those achieving total scores of 30 or higher were considered unimpaired. The MacCAT-CA does not yield a total score but yields scores for three sections. Defendants may be found unimpaired, mildly impaired, or severely impaired on any section. If a defendant was found severely impaired on any section, they were categorized as severely impaired. If they were not severely impaired on any section but were mildly impaired on one section they were categorized as mildly impaired. If there was no impairment on any section they were categorized as unimpaired. Only two defendants were classified as "mildly impaired" so the variables were recoded into two categories: unimpaired and impaired. If the defendant was not impaired on any section the defendant was categorized as unimpaired. If the defendant was impaired on any section (severe or mild), they were categorized as impaired. This yielded eight impaired defendants (25%) and 24 unimpaired defendants (75%). These changes are summarized in Table 3.

Table 3

Updated Defendant Characteristics for Analyses

Characteristic	*n*	*%*
Marital Status	32	
% Single		93.8
% Not Single		6.3
Employment	32	
% Employed		9.4
% Unemployed		90.6
Hospitalizations	32	
% Zero to One		34.4
% Two to Five		40.6
% Six or More		25.0
Psychiatric Diagnosis	32	
% Psychotic Disorder		75.0
% Non-Psychotic Disorder		25.0
Orientation	32	
% Mildly Impaired/Unimpaired		68.8
% Impaired		31.3
Criminal Offense	32	
% Felony		71.9
% Not Felony		28.1
Competency Test Performance	32	
% Unimpaired		75.0
% Impaired		25.0

Lens Model Analyses

It was hypothesized that variables identified as having predictive power in the literature, such as psychiatric diagnosis, psychiatric hospitalization history, and employment status, would significantly predict the decision of competency rendered by the legal system (the criterion). To investigate this hypothesis, standard multiple regression was used as a descriptive method to assess the influence of the eleven data cues on the decisions rendered by the courts. The intercorrelations among the eleven data cues along with the correlations between the data cues and the criterion are presented in Table 4. Substantial multicolinearity was not observed among the eleven data cues. Once these eleven data cues were entered into the regression, the total variance accounted for by the model was 76%. Data cues that reached statistical significance included competency test performance (beta = .61, p = .000), criminal offense (beta = .46, p = .001), and hospitalization history (beta = .39, p = .003). Marital status, although it did not reach statistical significance, also appeared to influence the decision (beta = .26, p = .07). Standardized beta coefficients are presented in Table 5.

The accuracy rate for each protocol was calculated. Results indicated there were three protocols for which no rater was accurate and one protocol for which only one rater was accurate. In a separate set of exploratory analyses these four protocols were controlled for and the regression analysis for the court decision was rerun. Results indicated an improvement in the total variance accounted for from 76% to 79%. Competency test performance (beta = .63, p = .000), criminal offense (beta = .46, p = .001), and hospitalization history (beta = .37, p = .008) continued to achieve statistical significance, with the addition of marital status (beta = .45, p = .04). Orientation (beta = .30, p = .06) and gender (beta = -.27, p = .07) approached statistical significance.

Table 4

Pearson Correlation Coefficients Between Predictor Variables and Between Predictor Variables and the Criterion

Variables	1	2	3	4	5	6	7	8	9	10	11	12
1. Age	-	-.06	.01	-.08	-.08	-.14	-.17	.33	.44**	.20	-.17	.14
2. Gender		-	-.06	-.02	.30	.42	.40*	.04	-.09	.03	.18	-.11
3. Ethnicity			-	-.26	.11	.22	.29	.14	-.03	-.07	.00	-.14
4. Marital Status				-	.36*	.31	-.15	-.10	-.19	-.16	.15	.15
5. Employment					-	.21	.31	-.01	-.05	.04	.19	-.06
6. Hospitalizations						-	.21	.20	.06	.05	.04	.40*
7. Diagnosis							-	.08	-.02	-.04	.17	.00
8. Orientation								-	.31	-.03	-.08	.08
9. Intelligence Quotient (IQ)									-	.13	-.21	.16
10. Criminal Offense										-	-.12	.36*
11. Competency Test											-	.50**
12. Criterion												-

*p ≤ .05, ** p ≤ .01

Table 5

Standardized Beta Values for the Criterion

Variables	Beta Coefficients
Age	.12
Gender	-.18
Ethnicity	-.05
Marital Status	.26
Employment	-.22
Hospitalizations	.39**
Diagnosis	.17
Orientation	.15
Intelligence Quotient	.13
Criminal Offense	.46**
Competency Test	.61**

** $p \leq .01$

It was hypothesized that these same variables (psychiatric diagnosis, psychiatric hospitalization history, and employment status) would significantly predict decisions of competency rendered by the raters. Standard multiple regression analyses were run as a descriptive method to examine this hypothesis for each rater and majority rater. An Expert Majority Rater, Novice Majority Rater, and Legal Majority Rater was created by using the competency decision rendered by the majority of the raters for each group. For example, the Expert Majority Rater's decision for protocol 1 was based on the majority decision of all the psychologist raters. If five of the nine psychologists found the defendant depicted in protocol 1 to be competent, then the Expert Majority Rater's decision for protocol 1 was entered as competent. This same method was applied to create the Novice Majority Rater and the Legal Majority Rater.

For the Expert Majority Rater, the total variance accounted for by the model was 43% and competency test performance was the only cue to reach statistical significance. For the Novice Majority Rater, the total variance accounted for by the model was 50% and competency

test performance was the only statistically significant cue. For the Legal Majority Rater, the total variance accounted for by the model was 42% and competency test performance was the only statistically significant cue.

In a separate set of exploratory analyses, the four least accurate protocols were controlled for and the regression analyses for the majority raters were rerun. Results again indicated an improvement in the total variance accounted for. For the Expert Majority Rater, the total variance accounted for increased substantially from 43% to 72%. Competency test performance remained significant in the model (beta = .51, p = .002) but also significant was criminal offense (beta = .64, p = .001), marital status (beta = .53, p = .03), and orientation (beta = .40, p = .03). Gender approached statistical significance (beta = -.31, p = .07). For the Novice Majority Rater, the total variance accounted for also increased substantially from 50% to 76%. Competency test performance remained significant in the model (beta = .57, p = .000) but also significant was criminal offense (beta = .58, p = .001), marital status (beta = .50, p = .03), gender (beta = -.34, p = .03), and hospitalization history (beta = .32, p = .03). For the Legal Majority Rater, the total variance accounted for again increased substantially from 42% to 72%. Competency test performance remained significant in the model (beta = .49, p = .003) but also significant was criminal offense (beta = .63, p = .001), marital status (beta = .50, p = .04), and orientation (beta = .40, p = .03). Hospitalization history (beta = .28, p = .06) and gender (beta = -.29, p = .08) approached statistical significance.

Results of analyses including all 32 protocols indicated that competency test performance was the only cue that reached statistical significance for six psychologists. Of these psychologists hospitalization history approached significance (p = .06) for one rater. Of the remaining psychologists, ethnicity was the only other cue to reach statistical significance. Seven of the graduate students relied on competency test performance (the only statistically significant cue),

with hospitalization history nearing statistical significance (p = .06) for one rater. Similar to the

psychologists, ethnicity was the only other cue to reach statistical significance. More variability

was observed in the sample of lawyers. Half of the lawyers did not rely on one cue over another;

however, for two of these lawyers competency test performance neared significance (p = .06).

Only two lawyers relied on competency test performance, with this cue reaching statistical

significance. For another lawyer, hospitalization was the only statistically significant cue but

competency test performance neared statistical significance (p = .06). Diagnosis and competency

test performance were statistically significant for one lawyer, with intelligence (p = .06), marital

status (p = .07), and ethnicity (p = .08) nearing significance. Standardized beta coefficients for

each rater and majority rater are presented in Tables 6 through 8.

Table 6

Standardized Beta Values for Psychologists

	Psychologist Raters									
Variables	1	2	3	4	5	7	8	9	Majority	
Age	-.01	-.01	-.10	-.13	-.01	.21	.01	.07	-.01	
Gender	-.09	-.02	.08	.19	-.02	.00	-.08	-.12	-.02	
Ethnicity	-.04	-.07	.03	-.50**	-.07	-.07	-.01	-.04	-.07	
Marital Status	-.09	-.11	-.04	-.35	-.11	-.22	-.01	-.10	-.11	
Employment	.20	.17	.12	.25	.17	.20	.19	.20	.17	
Hospitalizations	.18	.18	.07	.04	.18	.31	.05	.20	.18	
Diagnosis	.25	.21	.19	.06	.21	.14	.37	.29	.21	
Orientation	.11	.03	-.03	.27	.03	-.04	.17	.02	.03	
Intelligence Quotient	.24	.22	.27	.00	.22	.09	-.10	.00	.22	
Criminal Offense	.11	.21	.27	.04	.21	.10	.26	.26	.21	
Competency Test	.42*	.47*	.41*	.24	.47*	.57**	.23	.40*	.47*	

$* p \leq .05, ** p \leq .01$

Table 7

Standardized Beta Values for Graduate Students

	Graduate Student Raters									
Variables	1	2	3	4	5	6	7	8	9	Majority
Age	.12	-.06	-.16	-.27	.21	.26	.17	.13	.19	.11
Gender	.01	.16	-.03	-.06	.00	.09	.05	-.07	-.14	-.06
Ethnicity	-.09	-.21	-.42*	.02	-.07	.04	.11	-.14	.04	.01
Marital Status	-.19	-.22	-.37	-.30	-.22	-.15	-.15	-.16	-.12	-.15
Employment	.20	.20	.39	.16	.20	.24	.15	.23	.22	.19
Hospitalizations	.14	.22	.16	.11	.31	.24	.22	.19	.29	.27
Diagnosis	.13	.08	.04	.20	.14	.19	.09	.30	.25	.19
Orientation	.08	.17	.16	.17	-.04	-.02	-.09	.08	.03	-.09
Intelligence Quotient	.15	.14	-.03	.23	.09	.29	.20	.08	.21	.20
Criminal Offense	.10	.11	-.12	-.19	.10	-.10	.11	-.01	.04	.16
Competency Test	.42*	.38*	.24	.46**	.57**	.28	.59**	.49**	.48**	.53**

*$p \leq .05$, ** $p \leq .01$

Table 8

Standardized Beta Values for Lawyers

				Lawyer Raters					
Variables	1	2	3	4	5	6	7	8	Majority
Age	.11	-.24	.26	.00	.10	-.01	-.08	.13	-.03
Gender	-.06	.12	-.14	.00	.00	.02	.01	-.07	.01
Ethnicity	.01	-.03	.14	-.03	-.08	-.14	-.10	-.30	-.06
Marital Status	-.15	-.25	-.08	-.32	.09	-.10	-.12	-.33	-.15
Employment	.19	-.01	.24	.24	.08	.18	.21	.09	.19
Hospitalizations	.27	.08	.41*	-.07	.27	.03	.19	.13	.23
Diagnosis	.19	.18	.33	-.07	.11	.34	.20	.48**	.17
Orientation	-.09	.14	.06	.02	.16	.11	.17	.18	.01
Intelligence Quotient	.20	.23	-.03	.36	.10	-.27	.25	-.33	.23
Criminal Offense	.16	.25	.00	-.23	.03	.27	.05	.26	.17
Competency Test	.53**	.29	.34	.46*	.39	.34	.30	.36*	.45*

$* p \le .05, ** p \le .01$

It was also hypothesized that linear models of judgments would outperform human models of judgments. According to Goldberg (1970), the lens model can be used to compare human models to linear models by examining the difference between the achievement indices, or the validity coefficient of the rater (r_a) and the validity coefficient of the linear model of the rater (r_m). The validity coefficient of the human model is subtracted from the validity coefficient of the linear model (e.g., r_m - r_a) to provide a measure of the difference between the two coefficients represented as Δ (Goldberg, 1970). Negative numbers indicate the human model outperformed the linear model by achieving a higher validity coefficient. The lens model for each rater and majority rater is shown in Table 9.

Most raters showed little to no difference in validity coefficients between the human model and the linear model. Raters who showed a benefit of the human model tended to rely on nonlinear approaches to decision making. In some cases these raters did rely heavily on linear approaches to decision making but they relied just as heavily on nonlinear approaches. Psychologist raters 4 and 7 showed some benefit of the human model. This is not surprising as psychologist rater 4 primarily showed a nonlinear approach to decision making (C = .40) and psychologist rater 7 showed a high nonlinear approach (C = .75). Graduate student raters 3, 4, and 5 also showed some benefit of the human model over the linear model, with raters 3 and 4 displaying primarily nonlinear approaches to decision making (C = .48 and .51 respectively) and rater 5 showing a high nonlinear approach (C = .74). Lawyer rater 6 showed some improvement when using the human model over the linear model and also demonstrated primarily a nonlinear approach to decision making (C = .55). Lawyer rater 5 was the only rater to show some benefit when using the linear model over the human model. For majority raters there were virtually no differences observed between the human model and the linear model. Figure 2 displays a pictorial representation of the lens model using indices from the Expert Majority Rater.

Table 9

Brunswik Lens Model for Rater Judgments of Competency to Stand Trial

| Raters	r_a	r_m	Δ	R_e	R_s	G	C
Criterion	-	-	-	.87	-	-	-
Psychologists							
Rater 1	.55**	.56**	0.01	-	.63	.65**	.51**
Rater 2	.60**	.61**	0.06	-	.65	.70**	.53**
Rater 3	.55**	.52**	-0.03	-	.63	.60**	.58**
Rater 4	.28	.20	-0.08	-	.70	.23	.40*
Rater 5	.60**	.61**	0.01	-	.65	.70**	.53**
Rater 7	.70**	.61**	-0.09	-	.74	.71**	.75**
Rater 8	.43**	.38*	-0.05	-	.60	.44**	.52**
Rater 9	.55**	.60**	0.05	-	.62	.69**	.46**
Majority Rater	.60**	.61**	0.01	-	.65	.70**	.53**
Graduate Students							
Rater 1	.55**	.55**	0.00	-	.58	.64**	.56**
Rater 2	.55**	.49**	-0.06	-	.64	.56**	.61**
Rater 3	.30	.18	-0.12	-	.62	.21	.48**
Rater 4	.33	.22	-0.11	-	.75	.25	.51**
Rater 5	.70**	.61**	-0.09	-	.74	.71**	.74**
Rater 6	.39*	.39*	0.00	-	.63	.45**	.37*
Rater 7	.55**	.58**	0.03	-	.72	.67**	.37*
Rater 8	.51**	.51**	0.00	-	.68	.59**	.45**
Rater 9	.60**	.58**	-0.02	-	.70	.66**	.55**
Majority Rater	.65**	.62**	-0.03	-	.70	.72**	.59**
Lawyers							
Rater 1	.65**	.62**	-0.03	-	.70	.72**	.59**
Rater 2	.40*	.38*	-0.02	-	.63	.43**	.44*
Rater 3	.51**	.48**	-0.03	-	.70	.55**	.49**
Rater 4	.22	.18	-0.04	-	.64	.21	.27
Rater 5	.57**	.68**	0.11	-	.55	.78**	.47**
Rater 6	.43**	.34	-0.09	-	.70	.39*	.55**
Rater 7	.45**	.47**	0.02	-	.58	.54**	.44*
Rater 8	.40*	.39*	-0.01	-	.78	.45**	.31
Majority Rater	.59**	.59**	0.00	-	.65	.68**	.55**

*p ≤ .05, ** p ≤ .01
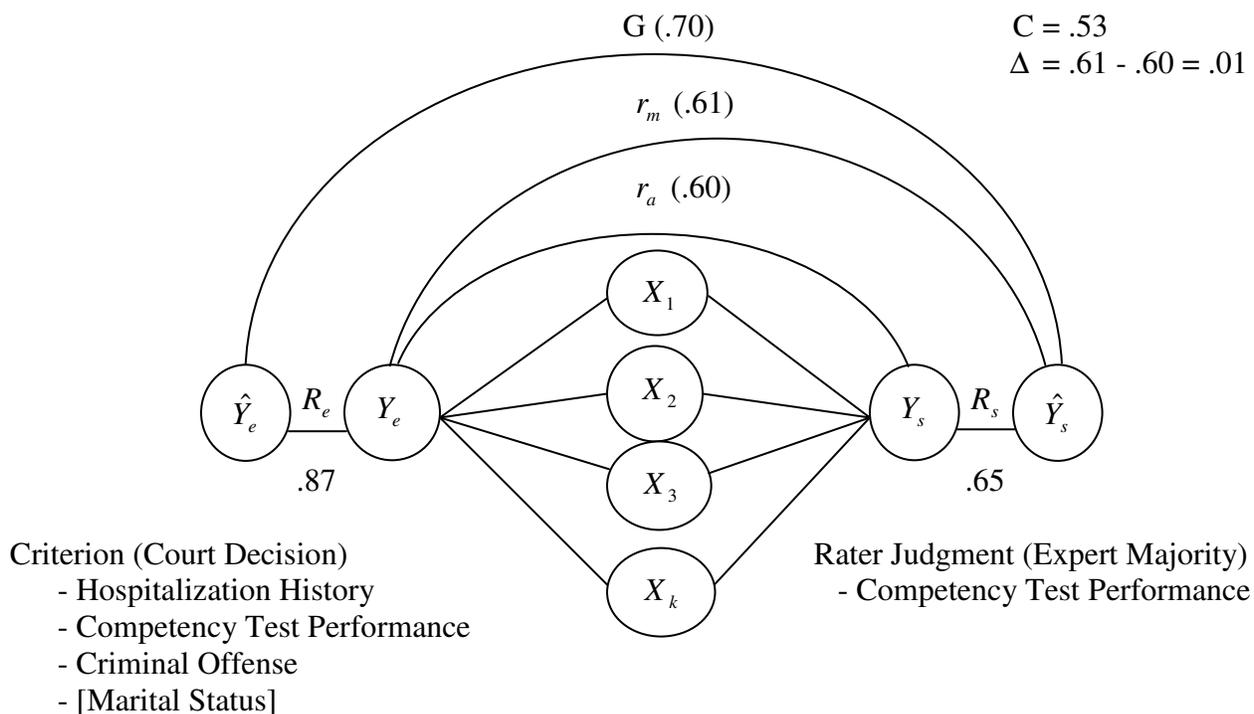
Figure 2. A pictorial representation of Brunswik's Lens Model for the Expert Majority Rater.

Rater Accuracy

It was hypothesized that experience would have no impact on raters' accuracy, with expert raters and novice raters displaying equal levels of accuracy. Further, it was predicted that clinical raters (experts and novices) would outperform legal raters (lawyers) by achieving higher levels of accuracy.

Rates of achievement were examined using the achievement index (or validity coefficient). Also examined was the hit rate, or the number of correctly classified protocols. The least accurate rater of all the raters across disciplines was a psychologist (rater 6) who correctly identified only 12 protocols. The probability that a rater would get 12 correct by chance is about 14 in 100 ($p = .14$). It is possible that this rater blindly rated or simply guessed the classification for each protocol; however, the low likelihood of achieving 12 correct by chance suggests the

rater misunderstood the directions or knew the correct classification but intentionally chose the incorrect classification. In addition, this rater only classified 31 protocols. This rater was considered an outlier and was excluded from the data analysis and from the creation of the Expert Majority Rater.

Results of statistical analyses indicate the highest achievement indices were obtained by psychologist rater 7 and graduate student rater 5 (r_a = .70). The most accurate lawyer was rater 1 (r_a = .65). Achievement indices (r_a) for psychologists ranged from .28 to .70. Achievement indices for graduate students ranged from .30 to .70. Achievement indices for lawyers ranged from .22 to .65. On average, there were minimal differences in the hit rates between the three groups. The psychologists' average hit rate was 78% (range = 69% to 88%), the graduate students' average hit rate was 77% (range = 59% to 88%), and the lawyers' average hit rate was 75% (range = 66% to 84%). Achievement indices and hit rates for all raters are presented in Table 10. Figure 3 displays the hit rates for each discipline as a bar graph.

The highest achieving majority rater was the Novice Majority Rater (r_a = .65), followed by the Legal Majority Rater (r_a = .61), and the Expert Majority Rater (r_a = .60). The Novice Majority Rater obtained a hit rate of 84%, followed by the Expert Majority rater with a hit rate of 81%, and the Legal Majority Rater with a hit rate of 78%. Although differences in achievement and hit rates among the majority raters were observed, these differences were negligible. Majority achievement indices, hit rates, and confidence ratings are presented in Table 10. See Figure 3 for a graphical representation of majority rater hit rates.

Table 10

Participant Achievement and Confidence by Discipline

Participant	CST [b]	IST [c]	Hit Rate [d]	r_a	Mean	Range
			Achievement		*Confidence* [a]	
Psychologists						
Rater 1	18	7	25	.55**	4.0	3 - 5
Rater 2	19	7	26	.60**	3.6	2 - 5
Rater 3	18	7	25	.55**	4.6	2 - 5
Rater 4	19	4	23	.28	3.6	3 - 4
Rater 5	19	7	26	.60**	4.1	3 - 5
Rater 7	21	7	28	**.70****	**4.1**	**1 - 5**
Rater 8	15	7	22	.43**	3.3	1 - 5
Rater 9	18	7	25	.55**	4.3	1 - 5
Majority Rater	19	7	26	.60**		
Graduate Students						
Rater 1	20	6	26	.55**	3.0	1 - 5
Rater 2	20	6	26	.55**	3.4	1 - 5
Rater 3	17	5	22	.30	3.6	1 - 5
Rater 4	12	7	19	.33	3.1	2 - 5
Rater 5	21	7	28	**.70****	**4.0**	**2 - 5**
Rater 6	19	5	24	.39*	3.3	2 - 5
Rater 7	20	6	26	.55**	3.3	2 - 5
Rater 8	17	7	24	.51**	3.3	2 - 5
Rater 9	19	7	26	.60**	4.0	1 - 5
Majority Rater	20	7	27	.65**		
Lawyers						
Rater 1	20	7	27	**.65****	**3.5**	**2 - 5**
Rater 2	17	6	23	.40*	3.4	2 - 5
Rater 3	17	7	24	.51**	3.8	2 - 5
Rater 4	20	3	23	.22	3.9	2 - 5
Rater 5	22	5	27	.57**	3.3	1 - 5
Rater 6	15	7	22	.43**	3.9	2 - 5
Rater 7	18	6	24	.45**	3.4	1 - 5
Rater 8	14	7	21	.40*	4.3	2 - 5
Majority Rater	19	6	25	.59**		

Note. The highest achievement score within each discipline is in boldface.
[a] Confidence was measured from 1 (not at all confident) to 5 (extremely confident). [b] Competent to Stand Trial: the number of competent protocols that were accurately classified out of 24 competent protocols. [c] Incompetent to Stand Trial: the number of incompetent protocols that were accurately classified out of 8 incompetent protocols. [d] Overall Hit Rate: the overall number of protocols that were accurately classified out of a total of 32 protocols.
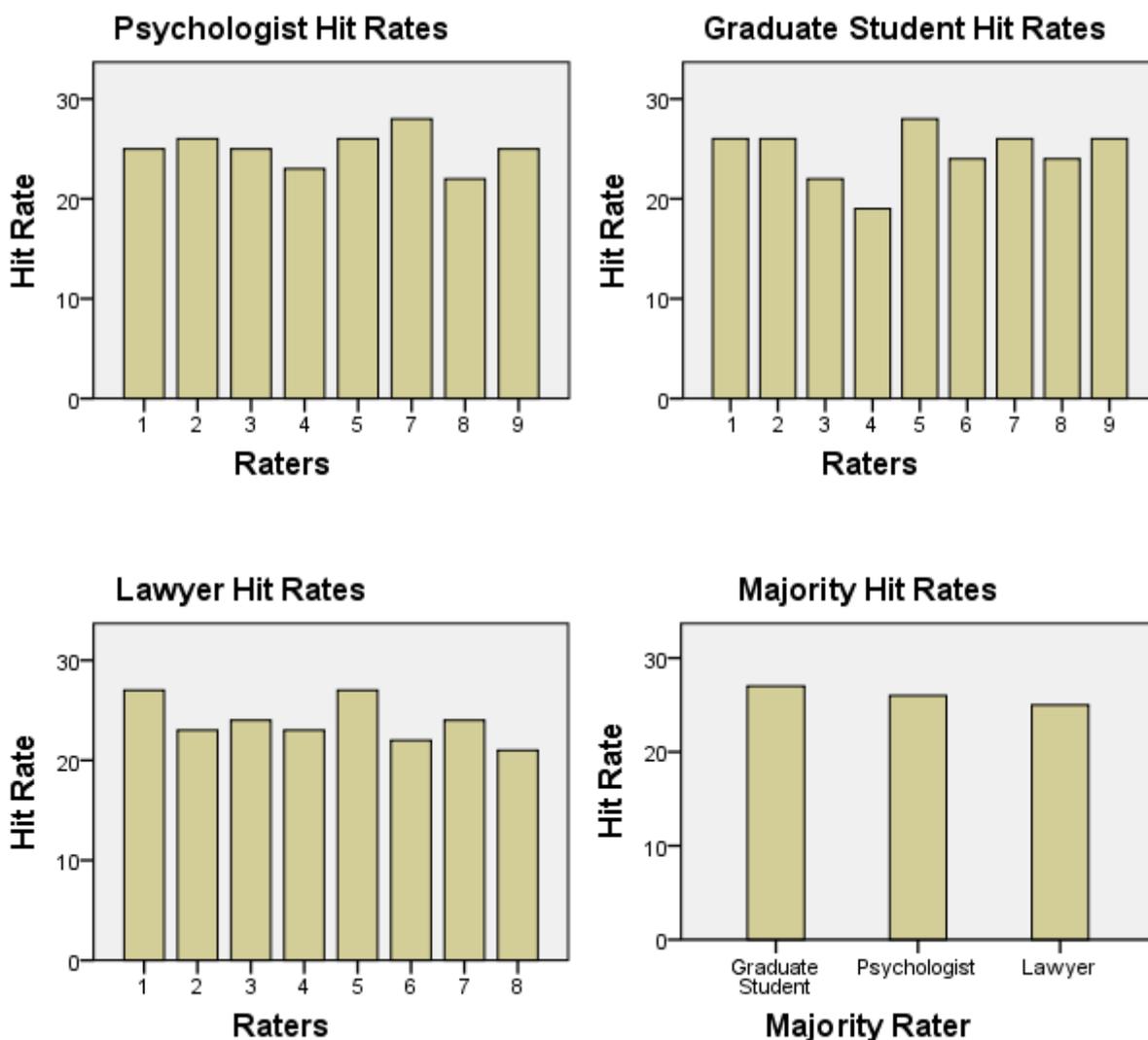* $p \leq .05$. ** $p \leq .01$.

Figure 3. Rater achievement represented within each discipline and between each discipline.

Exploratory analyses were conducted to examine the potential impact of order effects on rates of achievement. To examine whether order effects existed, accuracy rates for the first 16 protocols were compared to the accuracy rates of the last 16 protocols for each rater. Differences in accuracy rates for the first set of protocols compared to the second set of protocols were negligible for psychologists, graduate students, and lawyers. Only one rater, a graduate student, showed a large difference in accuracy for the two sets of protocols. This rater accurately

classified only six protocols in the first set but was able to accurately classify 13 protocols in the second set. Of the four protocols that raters had difficulty classifying (protocols for which no rater was accurate or for which only one rater was accurate), two of these protocols were in the first set and two of these protocols were in the second set. Thus, these four protocols did not impact the current set of exploratory analyses. Lastly, the protocols were divided into thirds and examined to determine whether similar cases were found within the three groups or whether certain protocols in the first group may have been more easily categorized and thus stood out to raters as a template to which to compare subsequent protocols. No major differences in the ease at which a decision could be rendered were observed for the first set of protocols; however several protocols at the end of the second and third set may have been more easily categorized. These protocols occurred at the end of the set and thus were unlikely to have influenced the raters' earlier decisions.

Rater Confidence

When mean confidence ratings were averaged within each discipline, psychologists obtained the highest average confidence rating (4.0), followed by lawyers (3.7) and graduate students (3.4). The most accurate psychologist and most accurate lawyer were not the most confident in their respective groups. The most accurate psychologist obtained the third highest confidence average (4.1), tied with another psychologist. The least accurate psychologist obtained the lowest confidence rating. The most accurate lawyer obtained the fifth highest confidence average (3.5) and the least accurate lawyer obtained a higher confidence average (3.9). For the graduate student group, the most accurate graduate student did obtain the highest confidence average (4.0) and that score was tied with that of the second most accurate graduate

student within that group. Similar to the lawyers, the least accurate graduate student did not

obtain the lowest confidence average. Mean confidence ratings are depicted in Table 10.

Relationship Between Accuracy and Confidence

It was hypothesized that raters' perceived level of confidence would be unrelated to

accuracy. The relationship between achievement and confidence was explored using correlations

and probabilities of accuracy for the highest and lowest confidence ratings. The greatest

relationship between confidence and accuracy was obtained by graduate student rater 8 who

showed a moderate positive correlation between confidence and accuracy, $r = .44$, $n = 32$, $p =$

.01, with higher levels of confidence associated with greater accuracy. Moderate positive

correlations showing greater confidence associated with greater accuracy were also obtained for

psychologist rater 3, $r = .42$, $n = 32$, $p = .02$, and lawyer rater 7, $r = .38$, $n = 32$, $p = .03$.

Probabilities of accuracy for the highest and lowest confidence ratings were also

explored. Theoretically, a protocol for which a rater provided a confidence rating of 5 (extremely

confident) would have a higher probability of accuracy than a protocol for which a rater provided

a confidence rating of 1 (not at all confident). To identify this probability, protocols rated with a

confidence of 5 (or the highest confidence rating assigned to any protocol) were identified and

the accuracy rate was determined by dividing the number of protocols rated as a 5 that were

accurate by the total number of protocols rated with a 5. This was determined for each rater.

For psychologists, the average probability of accuracy for protocols rated a 5 was 76%.

This is to say that on average, for protocols in which psychologists provided the highest

confidence ratings, the psychologists were accurate 76% of the time (range: 56% to 83%). This

means that on average, when the psychologists provided the highest confidence rating they were

inaccurate 24% of the time. For graduate students, the average probability of accuracy for

protocols rated a 5 was 88% (range: 67% to 100%) and the probability of inaccuracy was only

12%. For lawyers, the average probability of accuracy for protocols rated a 5 was 78% (range:

60% to 100%) and the probability of inaccuracy was 22%.

This procedure was also used to determine the probability of accuracy for protocols

given the least amount of confidence. For psychologists, the average probability of accuracy for

protocols rated a 1 was 51% (range: 0% to 100%). This means that on average, when the

psychologists provided the lowest confidence rating they were inaccurate approximately half the

time. For graduate students the average probability of accuracy for protocols rated a 1 was 73%

(range: 0% to 100%) and the probability of inaccuracy was only 27%. For lawyers, the average

probability of accuracy for protocols rated a 1 was 65% (range: 0% to 100%) and the probability

of inaccuracy was 35%. This information is presented in Figure 4.

This technique was used to examine raters with the greatest correlations between levels

of confidence and accuracy. For psychologist 3, confidence ratings of 5 were associated with

accuracy 83% of the time and confidence ratings of 1 were never associated with accuracy.

When graduate student 8 assigned a confidence rating of 5, an accurate decision was reached

100% of the time, compared to only 50% of the time for confidence ratings of 1. When lawyer 7

assigned a confidence rating of 5, an accurate decision was reached 71% of the time, compared

to only 40% of the time for confidence ratings of 1. This information is presented in Figure 5.
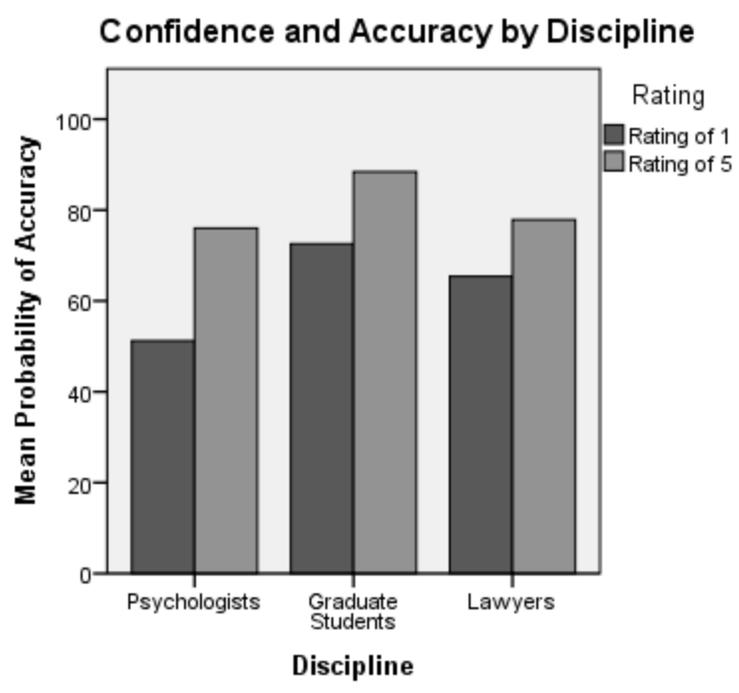
Figure 4. Probabilities of accuracy for confidence ratings of 1 (not at all confident) and 5 (extremely confident) separated by discipline.
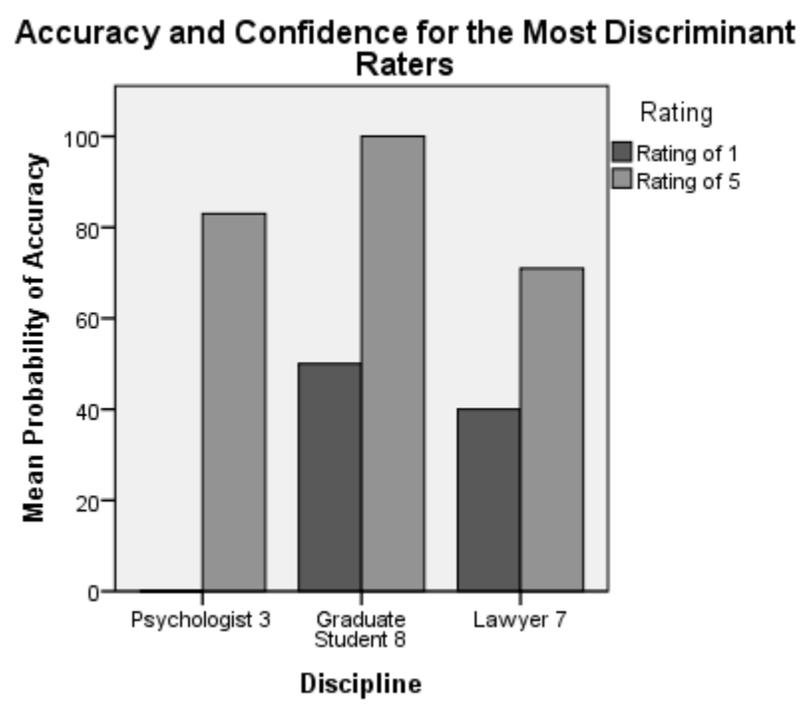

Figure 5. Probabilities of accuracy for confidence ratings of 1 (not at all confident) and 5 (extremely confident) for the most discriminate raters.

CHAPTER 5

DISCUSSION

Competency to stand trial evaluations are the most frequently performed forensic

competency evaluations (Mossman et al., 2007; Robinson & Acklin, 2010; Skeem & Golding,

1998) and research suggests judges rely heavily on expert opinions to form their decisions

(Melton et al., 2007; Roesch & Golding, 1980; Skeem & Golding, 1998). Despite this, less

emphasis has been placed on determining the decision making models used by clinicians when

performing these evaluations or examining the validity of these evaluations, partly because no

gold standard exists by which to compare clinicians' decisions (Mossman, 2008). To date, no

research studies have used a lens model approach to examine the decision making and accuracy

of psychologists performing evaluations of competency to stand trial. An understanding of the

process by which these decisions are rendered and the validity of these decisions is essential if

mental health professionals are to provide expert witness testimony for the court system (Faust,

2003). The purpose of this research project was to compare the decision-making, accuracy, and

confidence of clinical professionals with varying levels of expertise and legal professionals

making decisions of competency to stand trial using Brunswik's Lens Model.

Court Decision Making

As a result of the absence of a gold standard to which to compare raters' decisions of

competency, the ultimate decision rendered by the legal system was used as the criterion. Ideally,

competency test performance and diagnosis would be the two cues that influenced the court's

decision because a defendant must have a diagnosis to be considered incompetent and

competency measures attempt to capture a defendant's competence. Based on past research it

was predicted that psychiatric diagnosis, psychiatric hospitalization history, and employment

status would significantly predict decisions of competency rendered by the legal system. Results indicated hospitalization history was the only variable found to influence the decisions that was predicted. The ultimate opinion was primarily influenced by a defendant's performance on measures of competency to stand trial, criminal offense, and hospitalization history.

The influence of competency to stand trial measures on the court's decision is what one would hope to observe and means court judges considered relevant information when forming their decision. In this sample, there was a positive relationship between competency measure performance and the court's decision, such that increased performance on competency measures was associated with findings of competence. This is consistent with previous research that showed defendants found competent obtained significantly higher scores on measures of competency to stand trial than defendants found incompetent (Pirelli et al., 2011).

The influence of the defendant's criminal offense on the determination of competency is not surprising and there are many possible reasons for this. Courts may consider the seriousness of the charge and the defendant's ability to go to trial when considering whether the court will pursue charges. Some research suggests a relationship between criminal charge and findings of competence and some scholars have argued that the seriousness of the charge is a factor that should be considered in competency to stand trial evaluations (Buchanan, 2006; Rosenfeld & Ritchie, 1998). For example, if evaluators considered the seriousness of the charge and the complexity of the defendant's case when making decisions of competency to stand trial, then evaluators may be able to have a lower threshold for competency for defendants facing probation violation charges than defendants facing murder charges. There may also be a tendency for courts to err on the side of caution when making decisions of competency and this may be particularly so for felony cases when the cost of an incompetent defendant going to trial or

entering into a plea bargain has more serious consequences (Gowensmith et al., 2012). The relationship between the court's decision and the defendant's criminal offense was positive and suggested that as criminal offense moved from felony offenses to non felony offenses court judges moved from findings of incompetence to competence.

Lastly, there was a positive association observed between hospitalization history and the court's decision of competency, with greater number of hospitalizations associated with raters moving from findings of incompetence to competence. The association between hospitalization history and competency determination was predicted based on findings in the literature; however, the direction of the relationship observed in the current sample is opposite of what has been found in the literature. According to a meta-analysis, defendants with a previous psychiatric hospitalization were more likely to be found incompetent than defendants without this history, although only five studies were examined (Pirelli et al., 2011). Hospitalization history is not directly tied to the competency criteria but research by Nicholson and Kugler (1991) showed that it may correlate with competency decisions. Further, the direction of the relationship between hospitalization history and competency status was similar to that found between competency and diagnosis or psychiatric symptoms (Nicholson & Kugler, 1991). Thus, hospitalization history may be a way of judging the severity of an individual's symptoms as well as severity of diagnosis. Borum et al. (1993) suggest information that is unique or salient is more easily remembered. It may be that hospitalization history, especially extensive hospitalization histories, were inadvertently remembered and thus used in making the decision of competency. Although the direction of the relationship is opposite what was predicted, it is possible that the defendants in this sample with greater psychiatric hospitalization histories had prior interactions with the legal system and prior histories of being found competent to stand trial. This would possibly

allow these defendants to have more legal knowledge on which to draw despite their psychological symptoms.

Rater Decision Making

Based on past research it was predicted that psychiatric diagnosis, psychiatric hospitalization history, and employment status would significantly predict decisions of competency rendered by the raters. For the majority of psychologists and graduate students, defendants' competency test performance was the primary factor influencing decisions of competency. This was also true for the Expert Majority, Novice Majority, and Legal Majority raters. Greater performance on competency measures was associated with a change in determinations of incompetence to competence. While the Lawyer Majority Rater showed a primary reliance on competency test performance, individual lawyers varied more in their decision-making with some raters also relying on diagnosis or on hospitalization history alone.

The raters' emphasis on defendant's competency test performance is encouraging in that most raters did not seem to be influenced by extraneous or biasing information (such as the defendant's ethnicity), but this is also concerning because the creators of these measures stress the importance of using a multi-modal approach to making decisions of competency and not relying solely on a defendant's performance on psychological testing. It may be that in the absence of other information (such as first-person observation of the defendant and defendant's mental state), raters relied on the most objective information available to them that contained the least amount of error. It is not unreasonable to assume that standardized methods of examining a defendant's competency (e.g., psychological measures) are more reliable than narrative descriptions of symptoms and abilities; however, results of other standardized tests were not largely relied on by raters to make their decisions (e.g., intelligence measures). It is also possible

that this finding is a result of the methodology used in this research project that required raters to make a decision of competency for 32 protocols. It may be that raters relied on the defendant's performance on competency measures as a shortcut to quickly make decisions of competency for each protocol and avoid reading other information in the protocol.

Interestingly, the only rater that relied on diagnosis and competency test performance to inform their decisions, the two cues that are the most relevant to the legal statute, did not achieve high rates of accuracy and instead performed fairly poor compared to other raters. It seems this rater used both linear and nonlinear approaches to decision making. Had this lawyer relied primarily on a linear approach, the amount of error variance would have decreased, allowing for greater accuracy; however, given that the courts' decisions were primarily impacted by competency test performance, criminal offense, and hospitalization history, a greater reliance on diagnosis may not have improved accuracy despite its relevance to the statute.

The majority of raters did not rely on diagnosis to form their opinions. One possible reason for the lack of association between diagnosis and rater decisions could be that a defendant's diagnosis is not as relevant to their competency to stand trial as their specific psychological symptoms. For example, a defendant who is actively hearing voices may still be able to demonstrate legal knowledge, an appreciation of his/her own legal situation, and an ability to work with his/her lawyer, so long as the voices are not commenting about legal strategies or about courtroom personnel (e.g., the judge, the lawyers, etc). Similarly, a defendant who is delusional may still be considered competent to stand trial if the delusion is unrelated to their legal situation. If the delusion is related to their legal situation, this could have devastating consequences and impair the defendant's right to a fair trial. The Colin Ferguson trial is an example of such an event. Colin Ferguson was accused of murdering six passengers on a New

York train. Based on Mr. Ferguson's decision to proceed pro se, his unwillingness to consider an insanity defense, and his insistence that he was framed for the crime as part of an elaborate conspiracy, his lawyers requested an evaluation of his competency to stand trial. Despite urgings from his defense lawyers that the court find Mr. Ferguson incompetent, the judge found him competent. Mr. Ferguson proceeded to trial, represented himself, and his delusional beliefs and psychiatric symptoms intruded into his defense strategy during the trial. As such, his trial was considered by many to be a "mockery" of the judicial system (Peil, 1997, "Kuby and Other Observers Insisted," para. 7).

In this research project diagnosis was examined by assessing whether or not a defendant was diagnosed with a psychotic disorder. Research has suggested an increased likelihood of findings of incompetence for defendants with psychotic disorders (Pirelli et al., 2011). Individuals without a psychotic disorder may also be found incompetent to stand trial but this is typically a result of intellectual impairment. It is notable that only 44% of the sample used by Pirelli et al. (2011) had diagnoses of psychotic disorder and 6% had diagnoses of mental retardation. Seventy-five percent of the current sample had diagnoses of psychotic disorder and 31% had diagnoses of mental retardation. The higher rates of defendants diagnosed with psychotic disorder and mental retardation in this sample could also have influenced the impact of diagnosis on rater decisions.

One possible reason for the lack of observation of the influence of employment on decision making in this sample could be that the majority of the defendants in this sample (90.6%) were unemployed compared to 65% of the defendants in the literature review on which this prediction was based (Pirelli et al., 2011). This lack of variability in the sample could have

influenced the lack of relationship observed between rater decisions and defendants' employment status.

The study by Pirelli et al. (2011) on which these predictions were based found an association between findings of competence and greater performance on competency measures; however, Pirelli et al. examined this association separately. If competency measure information had not been given to the raters in this research project there may have been greater reliance on other information, such as diagnosis, hospitalization history, and employment status.

Disagreement Among Court Judges and Psychiatrists

There were three protocols for which no rater was accurate and one protocol for which only one rater was accurate. In each of these cases the decision rendered in the defendant's medical record by the treating psychiatrist conflicted with that of the court's decision. Given that the information presented to the raters was based on information rendered in the patient's medical record (with symptoms described by the psychiatrist), it is not surprising that the raters, when following this information, did not match the decision rendered by the court.

Court judges do not typically disagree with experts' opinions of competency to stand trial so it seems unusual that the court would disagree with the psychologist for as many as four out of 32 defendants. Gowensmith et al. (2012) recently found that disagreement between experts and judges often occurred when the expert found the defendant competent but the court found the defendant incompetent. That was not the case for this sample. Of the four defendants, only one was a case in which the court found the defendant incompetent after the expert found the defendant competent. In this case the defendant was facing 15 charges, 14 of which were felony charges. In the other three cases the court deemed the defendant competent after the expert found the defendant incompetent. In the three cases in which the court deemed the defendant competent

after the expert found the defendant incompetent, all three defendants had lower level charges (misdemeanors and summary offenses). It may be that the nature of the charges these three defendants were facing were considered less serious and the level of involvement required by the defendant in the case was minimal (e.g., perhaps the defendant was not expected to enter a plea, testify, etc).

The discrepancy could be a result of many reasons, such as the county in which the hearing occurred, the individual judge assigned to the case, the defendant's presentation in court, pressure by an overcrowded legal system to move defendants through the system by settling cases with plea agreements, the presence of an opposing expert, and the nature of the charge. Unfortunately, it was not possible to determine the reasons for the discrepancy for these four protocols. It was unknown whether these defendants were seen by the same judge or even in the same county as that information had been destroyed by the time of this writing to protect defendant confidentiality It is not possible to know which of the many possible factors or combinations of factors contributed to the disagreement between the court judges and psychiatrists on these four protocols but it is possible to control for these four protocols and examine the impact on the raters' decision making strategies.

For all majority raters competency test performance was originally the only variable that reached statistical significance in the model. After controlling for these four protocols competency test performance remained significant for all majority raters while other cues reached statistical significance. For the Expert Majority Rater criminal offense, marital status, and orientation also reached statistical significance. For the Novice Majority Rater, criminal offense, marital status, gender, and hospitalization history reached significance. For the Legal Majority Rater, criminal offense, marital status, and orientation reached statistical significance.

The decision making of the majority rater is a reflection of the general decision making strategies employed by individual raters. It may be that the four protocols that showed disagreement between the court and experts' opinion were "noise" that was affecting the ability to best capture the raters' decision making strategies. If this view is taken, then raters not only relied on results of defendants' performance on competency to stand trial measures, but also on defendants' criminal offense, marital status, orientation, gender, and hospitalization history. Some of these variables match with what was predicted based on the literature (e.g., marital status and hospitalization history), while others were not predicted (e.g., competency measure performance, criminal offense, orientation, and gender).

As previously mentioned, raters may have emphasized standardized methods of examining a defendant's abilities (e.g., performance on psychological measures) over narrative descriptions. If this were the case, one would expect to find performance on measures of competency to stand trial and orientation (which included defendants' performance on the mini mental status examination) to influence raters' opinions. Once the four unusual protocols were accounted for, both competency measure performance and orientation reached statistical significance, which is what one would expect if raters were relying more on psychological measures than narrative descriptions.

Majority raters showed a significant relationship between competency decision and gender, with male defendants more likely to be found incompetent. Other research has found a relationship between decisions of competency to stand trial and gender but the direction of the relationship was different, with women more likely to be found incompetent than men; however, this relationship was small (Nicholson & Kugler, 1991). A more recent meta-analysis did not find a relationship between gender and competency status (Pirelli et al., 2011). It has been

suggested that this relationship may be moderated by other variables, such as the prevalence of certain diagnoses by gender (Nicholson & Kugler, 1991). Most of the research on competency to stand trial has focused on male or predominantly male samples and more research is needed to examine characteristics and differences between male and female defendants (Kois, Pearson, Chauhan, Goni, & Saraydarian, 2012). This study was comparatively unusual in that approximately half the sample was male and half was female.

Linear Models vs. Human Models

Research on clinical judgment and decision making provides evidence that cognitive limitations in the form of heuristics and biases impact the accuracy of human judgments. To correct for these limitations, linear models can be used to cut down on the unreliability (and bias) found in human judgments (Dawes, 1982; Goldberg, 1970). The tendency to use extraneous information when forming opinions, such as using information about a defendant's marital status or gender would be corrected for by using a linear model. In fact, studies have shown that linear models are generally equal to or better at making predictions than human models (Dawes et al., 1989). Therefore it was predicted that linear models would outperform human models by attaining higher achievement rates.

Contrary to what was predicted, for most individual raters and for all the majority raters there were little to no differences in achievement between human models and linear models. For six raters the human model outperformed the linear model. Only one rater showed improvement of the linear model compared to the human model. This rater did not rely on a standard set of cues and thus it is not surprising that the linear model would outperform this rater given the linear model's ability to improve reliability and decrease error variance. Despite having employed a primarily linear approach to decision making, this rater did not achieve the highest

accuracy rates. This could be because the rater did not also use nonlinear strategies to decision making that were valid for this sample and helped improve accuracy. This reflects the importance of the human component of decision making when using actuarial methods, particularly if raters are able to use nonlinear approaches in a valid way. Humans must first identify the best cues to use before the cues are adopted into a linear model. Once linear models are using the best cues identified by humans, they may outperform humans because humans are unreliable in their decision making (as a result of fatigue, environmental distractions, etc.) or because they are not using the cues in a linear way (Dawes, 1982; Goldberg, 1970). However, if there is the presence of "systematic nonlinear variance" (p. 145), as indicated by high values of C, then this means raters are using accurate strategies apart form what is captured by the model (Hammond et al., 1964). This suggests raters may have been relying on their clinical judgment, their intuition, or other means of making decisions that what was captured by the model. Raters that followed primarily nonlinear approaches achieved lower levels of accuracy, which is not surprising given the higher possibility of error and unreliability associated with this approach. These results suggest that when raters use cues in a linear way along with valid nonlinear approaches, linear models may offer little improvement.

Similar to findings by Walters et al. (1988), raters who primarily used linear approaches had a higher average hit rate (78%) than raters who primarily used nonlinear approaches (68%); however, the most accurate psychologist, graduate student, and also relied heavily on nonlinear approaches. Linear models may not have outperformed human models in this sample because raters were already using linear approaches in their decision making and the linear model could not improve upon this. Further, raters also used valid nonlinear approaches to decision making

that the linear model could not account for. Thus, when raters use cues in a linear way and also validly use cues in a nonlinear way, linear models may offer little improvement.

Rater Accuracy

Research does not support the benefit of experience for judgment accuracy (Garb, 1998). It was predicted that experience would be unrelated to accuracy. Other research by Garb (1998) does suggest a general benefit for clinical raters over non-clinical raters when making decisions about clinical populations. Based on this research it was predicted that clinical raters would outperform legal raters. As predicted, there were virtually no differences between the achievement rates obtained by psychologists and graduate students, suggesting no benefit of experience. Contrary to the prediction that clinical raters would achieve higher rates of accuracy than legal raters, virtually no differences between clinical and legal raters were observed. Psychologists, graduate students, and lawyers were among the three most accurate raters and among the three least accurate raters. In addition the average hit rates obtained by the different disciplines differed by only 3%.

The lack of differences in accuracy between clinical and non-clinical raters was also observed by Golding et al. (1984) and Cooper & Werner (1990) for decisions of competency to stand trial and violence risk. While clinical populations may be involved in these predictions, non-clinical populations may also be present and the clinician will have to decide which of the defendants before them have mental health diagnosis, and of the defendants who reach criteria for a diagnosis, which of these defendants are capable to stand trial. It may be that clinicians are better able to identify clinical populations but may not be better at applying this knowledge to areas outside of psychology, such as the legal system. The lack of differences observed in this research project could also be a result of raters' experiences with mental health populations. The

amount of mental health experience possessed by each rater was unknown. It may be that lawyers who had experience or interest in working with mental health populations self-selected into this research project. If this occurred, their knowledge of mental health information could have increased their accuracy above what would be found in a sample of lawyers with little to no experience or interest in mental health populations.

Results also suggest most raters were able to achieve higher rates of accuracy than would have been achieved by using the base rate alone. The rate of incompetency for this sample was 25%, which closely resembles the base rate of incompetency (27.5%) determined by Pirelli et al. (2011). Raters using the base rate would deem all defendants competent and achieve an accuracy rate of 75%. No rater used this approach. Two psychologists, two graduate students, and four lawyers performed lower than what would be predicted by using the base rate. This is dissimilar to the results found by Gaudette (1992), in which all raters performed lower than the base rate.

Rater Confidence

Given past research on confidence and accuracy, it was hypothesized that confidence would be unrelated to accuracy (Cooper & Werner, 1990; Jackson, 1986; Saks & Kidd, 1980/1986). Not surprisingly, graduate students had the lowest average confidence ratings even though virtually no differences in accuracy were observed between the disciplines. This may reflect the tendency of graduate students to underestimate their knowledge given their status as trainees. Graduate students were more discriminate in their ratings. When a graduate student assigned the highest confidence rating to a protocol, they were accurate on average 88% of the time, compared to 78% for lawyers and 76% for psychologists. Research has shown a tendency for raters to be overconfident in their decisions and psychologists and lawyers as a group were more overconfident than graduate students (Jackson, 1986; Saks & Kidd, 1980/1986).

Despite being overconfident, psychologists were better at identifying protocols that posed a challenge and adjusted their confidence accordingly. On average, when a psychologist provided a protocol with the lowest confidence rating, they were inaccurate about half the time, while lawyers were inaccurate 35% of the time and graduate students were inaccurate only 27% of the time. This suggests graduate students and lawyers achieved higher accuracy rates than psychologists for protocols for which they were not confident. Thus, graduate students and lawyers underestimated their knowledge for some of the protocols.

The greatest relationship between confidence and accuracy showed a moderate positive correlation and was achieved by a graduate student. When this rater assigned confidence ratings of 5 they were accurate 100% of the time and when they assigned confidence ratings of 1 they were inaccurate 50% of the time.

Limitations and Future Directions

There are several limitations to this study, the first and perhaps most significant limitation being the lack of a gold standard for which to compare the rater's decisions of competency (Mossman, 2008). Judges, like raters, are humans and thus are subjected to the same cognitive traps, biases, and errors as the raters in this study. However, judges do make final decisions of competency so their decision is the best criterion to which decisions could be compared. Further, when comparing the ecological side of the lens model to the human side, the judges' decisions represent decisions based on all available information, while the raters' decisions were based on a snapshot of information provided to them in the protocols. This strategy is imperfect but it is the best option at the current time (Mossman, 2008).

A second limitation to this study includes the process by which decisions were rendered for defendants' competency to stand trial. It is unclear whether judges' decisions were based on

information obtained from one expert or whether there was an opposing expert. The timing of the

evaluation and the court's determination of competency was also unclear in some cases. A

defendant may be evaluated for their competency to stand trial but may go months before going

before the judge to receive a formal determination of competency. Further, there were several

defendants in this sample for which the ultimate decision from the court could not be determined

because the judge's decision was not available in the court records. In these situations, the

psychiatrist's opinion was used as a substitute opinion. This was not considered a large limitation

because research consistently suggests there are high agreement rates between experts' opinions

and those rendered by the court but it is a limitation worth noting (Hart & Hare, 1992; Zapf et al.,

2004). Further, competency may exist on a continuum but this was not investigated in this study

because the legal system requires categorical decisions of competency. There may have been a

greater association observed between competency and accuracy had a dimensional approach to

rating the protocols been used. Melton et al. (2007) suggests the legal system emphasizes

categorical decisions over dimensional constructs. Thus, it is unlikely a dimensional approach

will be taken in the future, at least without establishing useable cut-off scores.

There were also limitations as a result of the sample of defendants and participants.

Given the lack of consistent data in defendants' medical records, the sample size from which

protocols were created was lower than anticipated and portions of missing data had to be

addressed. The limited number of protocols was not considered to be a substantial limitation of

this study as other research using the lens model to explore decision making has used a similar

number of protocols (Cooper & Werner, 1990).

As a result of the low number of participants in this study, information was not collected

to address outside influences on raters' decisions, such as demographic information and personal

history (e.g., years practicing, degree, etc). Within discipline variability was not a focus of this research project and as such it was not explored. As a result, the practice setting in which the psychologists conducted their evaluations (e.g., inpatient vs. outpatient) was not known. It was also not known whether the lawyers in this study primarily consisted of defense lawyers or lawyers who served in both defense and prosecutorial roles. If defense lawyers were oversampled, there may have been a propensity for defense lawyers to err on the side of finding a defendant not competent to stand trial; however, lawyers in this sample did not show greater rates of incompetency findings than psychologists or graduate students. Similar to psychologists who may pursue specialty areas in training, lawyers may also pursue specialty areas and it is not known whether the lawyers in this sample had experience working with defendants with mental health issues on a regular basis. In addition, this study did not employ a true control group in that it did not sample individuals with no knowledge of either psychological or legal principles (e.g., engineers, artists, etc).

Another factor that was not examined that could have impacted the ultimate court decision is that the evaluations were all conducted on an inpatient basis and thus may represent defendants with more severe symptoms, increased hospitalization histories, and less access to resources than outpatient defendants. In fact, Warren et al (1997) found that defendants evaluated in inpatient settings were less likely to be found incompetent than defendants evaluated in outpatient settings. Otto et al. (1998) found no differences in MacCAT-CA scores for defendants deemed incompetent to stand trial across evaluation settings but did find hospitalized defendants to have lower MacCAT-CA scores. The differences between these two groups in Pennsylvania, if any, are unknown at the present time. Further, the criminal histories of

the defendants in this sample were unknown. Defendants with lengthier criminal histories may be more likely to be found competent given previously acquired knowledge of the legal system.

Faust (2003) suggests clinicians should rely on actuarial prediction methods to inform their decisions. According to Ennis and Litwack (1974), in order for mental health professionals to provide expert testimony to the courts, their conclusions should be reliable, valid, based on professional knowledge, and more accurate than conclusions rendered by lay persons. The results of this study are encouraging in that most raters seemed to be using linear approaches to decision making and seemed to rely on objective information (test results) over subjective or biasing information (e.g., second hand descriptions of symptoms, irrelevant information). Results from previous research provides support for the reliability of clinicians making decisions of competency to stand trial and results of this research project provide support for the accuracy of clinician's decisions and reliance on professional knowledge (psychological test data) when forming their opinions (Morris et al., 2004; Poythress & Stock, 1980; Rosenfeld & Ritchie, 1998). Despite this, the current research project does not provide overwhelming support for the use of expert witnesses in the courtroom because there were not substantial differences between clinical and legal raters; however, the findings may have been different had a true control group been used. The results suggest the majority of raters and court judges relied on the use of psychological test data when making their decisions, and psychologists are the only discipline qualified to administer and interpret these measures. This suggests there is need for psychologists in the legal system to help the courts make decisions of competency to stand trial, even if it is merely to relay the results of psychological test information. It is important to not rely solely on this information when forming decisions and the decision maker should also considered factors such as psychological symptoms, intellectual abilities, and diagnoses. More research is needed to

explore the validity of clinician's decisions of competency to stand trial on a broader scale and to explore the differential validity between clinical and non-clinical raters before the clinician's place as expert to the courts can be secured.

REFERENCES

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

Blackstone, W. (1765-1769). *Commentaries on the laws of England*. Oxford, England: Clarendon Press.

Blashfield, R. K., Robbins, L., & Barnard, G. W., (1994). An analogue study of the factors influencing competency decisions. *Bulletin of the American Academy of Psychiatry & the Law, 22,* 587-594. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/7718931

Bonnie, R. J. (1992). The competence of criminal defendants: A theoretical reformulation. *Behavioral Sciences and the Law, 10,* 291-316. doi:10.1002/bsl.2370100303

Borum, R., Otto, R., & Golding, S. (1993). Improving clinical judgment and decision making in forensic evaluation. *Journal of Psychiatry & Law*, *21,* 35-76.

Brunswik, E. (1955). Representative design and probabilistic theory in a functional psychology. *Psychological Review, 62,* 193-217. doi:10.1037/h0047470

Buchanan, A. (2006). Competency to stand trial and the seriousness of the charge. *Journal of the American Academy of Psychiatry and the Law, 34,* 458-465.

Christy, A., Douglas, K. S., Otto, R. K., & Petrila, J. (2004). Juveniles evaluated incompetent to proceed: Characteristics and quality of mental health professionals' evaluations. *Professional Psychology: Research and Practice*, *35,* 380-388. doi:10.1037/0735-7028.35.4.380

Cochrane, R. E., Grisso, T., & Frederick, R. I. (2001). The relationship between criminal

charges, diagnoses, and psycholegal opinions among federal pretrial defendants. *Behavioral

Sciences & the Law*, *19,* 565-582. doi:10.1002/bsl.454

Cooksey, R. W. (1996). The methodology of social judgement theory. *Thinking & Reasoning*, *2*,

141-174. doi:10.1080/135467896394483

Cooper, R. P., & Werner, P. D. (1990). Predicting violence in newly admitted inmates: A lens

model analysis of staff decision making. *Criminal Justice and Behavior*, *17,* 431-447.

doi:10.1177/0093854890017004004

Cooper, V. G., & Zapf, P. A. (2003). Predictor variables in competency to stand trial decisions.

Law and Human Behavior, *27,* 423-436. doi:10.1023/A:1024089117535

Cuneo, D. J. & Brelje, T. B. (1984). Predicting probability of attaining fitness to stand trial.

Psychological Reports, 55, 35-39. doi:10.2466/pr0.1984.55.1.35

Dawes, R. M. (1982). The robust beauty of improper linear models in decision-making. In D.

Kahneman, P. Slovic, & A. Tversky (Eds.), *Judgment under uncertainty: Heuristics and

biases* (pp. 391-407). New York, NY: Cambridge University Press.

Dawes, R. M. (1994). *House of cards: Psychology and psychotherapy built on myth*. New York,

NY: Free Press.

Dawes, R. M., Faust, D., & Meehl, P. E. (1989). Clinical versus actuarial judgment. *Science*,

243, 1668-1674. doi:10.1126/science.2648573

Dawes, S. E., Palmer, B. W., & Jeste, D. V. (2008). Adjudicative competence. *Current Opinion

in Psychiatry*, *21*, 490-494. doi:10.1097/YCO.0b013e328308b2ee

Drope v. Missouri, 420 U.S. 162 (1975).

Dusky v. U.S., 271 F.2d 385 (U.S. App. 1959).

Dusky v. U.S., 362 U.S. 402 (1960).

Elstein, A. S., & Bordage, G. (1988). Psychology of clinical reasoning. In J. Dowie & A. S. Elstein (Eds.), *Professional judgment: A reader in clinical decision making* (pp. 109-129). New York, NY: Cambridge University Press. (Reprinted from *Health psychology – A handbook,* pp. 333-349, by G. Stone, F. Cohen, & N. Adler, Eds., 1979, San Francisco, CA: Jossey-Bass)

Ennis, B. J., & Litwack, T. R. (1974). Psychiatry and the presumption of expertise: Flipping coins in the courtroom. *California Law Review, 62,* 693-752. doi:10.2307/3479746

Everington, C., & Dunn, C. (1995). A second validation study of the competence assessment for standing trial for defendants with mental retardation (CAST-MR). *Criminal Justice and Behavior, 22*, 44-59. doi:10.1177/0093854895022001004

Farkas, G. M., DeLeon, P. H., & Newman, R. (1997). Sanity examiner certification: An evolving national agenda. *Professional Psychology: Research and Practice, 28,* 73-76. doi:10.1037/07357028.28.1.73

Faust, D. (2003). Holistic thinking is not the whole story: Alternative or adjunct approaches for increasing the accuracy of legal evaluations. *Assessment*, *10,* 428-441. doi:10.1177/1073191103259534

Fogel, M. H., Schiffman, W., Mumley, D., Tillbrook, C., & Grisso, T. (2013). Ten year research update (2001–2010): Evaluations for competence to stand trial (adjudicative competence). *Behavioral Sciences and the Law.* Advance online publication. doi:10.1002/bsl.2051

Foote, C. (1960). A comment on pre-trial commitment of criminal defendants. *The University of Pennsylvania Law Review, 108,* 832-846. doi:10.2307/3310455

Freeman v. People, 4 Denio 9 (Sup. Ct., N.Y. Co. 1847).

Garb, H. N. (1998). *Studying the clinician: Judgment research and psychological assessment*. Washington, DC: American Psychological Association. doi:10.1037/10299-000

Gaudette, M. D. (1992). *Clinical decision-making in neuropsychology: Bootstrapping the neuropsychologist utilizing Brunswik's Lens Model* (Doctoral dissertation). Retrieved from ProQuest Dissertations and Theses database. (Order No. 9222328)

Godinez v. Moran, 509 U.S. 389 (1993).

Goldberg, L. R. (1970). Man versus model of man: A rationale, plus some evidence, for a method of improving on clinical inferences. *Psychological Bulletin*, *73,* 422-432. doi:10.1037/h0029230

Golding, S. L., Roesch, R., & Shreiber, J. (1984). Assessment and conceptualization of competency to stand trial: Preliminary data on the interdisciplinary fitness interview. *Law and Human Behavior, 8,* 321-334. doi:10.1007/BF01044699

Gowensmith, W. N., Murrie, D. C., & Boccaccini, M. T. (2012). Field reliability of competence to stand trial opinions: How often do evaluators agree, and what do judges decide when evaluators disagree? *Law and Human Behavior, 36,* 130-139. doi:10.1037/h0093958

Grisso, T. (1988). *Competency to stand trial evaluations: A manual for practice*. Sarasota, FL: Professional Resource Exchange.

Grisso, T. (1996). Pretrial clinical evaluations in criminal cases: Past trends and future directions. *Criminal Justice and Behavior, 23,* 90-106. doi:10.1177/0093854896023001007

Grøndahl, P., Grønnerød, C., & Sexton, J. (2009). A comparative case vignette study of decision making in forensic psychiatric cases. *The International Journal of Forensic Mental Health*, *8,* 263-270. doi:10.1080/14999011003635639

Hammond, K. R. (2000). *Judgments under stress*. New York, NY: Oxford University Press.

Hammond, K. R., Hursch, C. J., & Todd, F. J. (1964). Analyzing the components of clinical inference. *Psychological Review*, *71,* 438-456. doi:10.1037/h0040736

Hart, S. D., & Hare, R. D. (1992). Predicting fitness to stand trial: The relative power of demographic, criminal, and clinical variables. *Forensic Reports*, *5,* 53-65.

Harvard Law Review (1967). Incompetency to stand trial. *Harvard Law Review*, *81,* 454-473. doi:10.2307/1339465

H.B. 1405, 196th Gen. Assem., Reg. Sess. (Pa. 2012).

Jackson, M. W. (1986). Psychiatric decision-making for the courts: Judges, psychiatrists, lay people? *International Journal of Law and Psychiatry, 9,* 507-520. doi:10.1016/0160-2527(86)90075-0

Kois, L., Pearson, J., Chauhan, P., & Saraydarian, L. (2012). Competency to stand trial among female inpatients. *Law and Human Behavior*. Advance online publication. doi:10.1037/lhb0000014

Meehl, P. E. (1957). When shall we use our heads instead of the formula? *Journal of Counseling Psychology, 4,* 268-273. doi:10.1037/h0047554

Meehl, P. E. (1973). *Psychodiagnosis: Selected papers*. Oxford, England: University of Minnesota Press.

Melton, G. B., Petrila, J., Poythress, N. G., & Slobogin, C. (with Lyons, P. & Otto, R.). (2007). *Psychological evaluations for the courts: A handbook for mental health professionals and lawyers* (3rd ed.). New York, NY: Guilford Press.

Mental Health Procedures Act of 1976, 50 Pa. Code § 7402 (Westlaw 2012 effective 1996).

Morris, G. H., Haroun, A. M., & Naimark, D. (2004). Assessing competency competently: Toward a rational standard for competent-to-stand-trial assessments. *Journal of the American Academy of Psychiatry and the Law, 32,* 231-245. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/15515910

Mossman, D. (2008). Conceptualizing and characterizing accuracy in assessments of competence to stand trial. *Journal of the American Academy of Psychiatry and the Law, 36,* 340-351. Retrieved from http://www.jaapl.org/cgi/content/full/36/3/340

Mossman, D., Bowen, M. D., Vanness, D. J., Bienenfeld, D., Correll, T., Kay, J., Klykylo, W. M., & Lehrer, D. S. (2010). Quantifying the accuracy of forensic examiners in the absence of a "gold standard." *Law and Human Behavior, 34,* 402-417. doi:10.1007/s10979-009-9197-5

Mossman, D., Noffsinger, S. G., Ash, P., Friersion, R. L., Gerbasi, J., Hackett, M., … Zonana, H. V. (2007). AAPL practice guideline for the forensic psychiatric evaluation of competence to stand trial [Supplemental material]. *Journal of the American Academy of Psychiatry and the Law*, *35,* S3–S72. Retrieved from http://www.peterash.com/pdf/AAPL_competence_guideline.pdf

Mumley, D. L., Tillbrook, C. E., & Grisso, T. (2003). Five year research update (1996-2000): Evaluations for competence to stand trial (adjudicative competence). *Behavioral Sciences & the Law*, *21*, 329-350. doi:10.1002/bsl.534

Murrie, D. C., Boccaccini, M. T., Zapf, P. A., Warren, J. I., & Henderson, C. E. (2008). Clinician variation in findings of competence to stand trial. *Psychology, Public Policy, and Law, 14,* 177-193. doi:10.1037/a0013578

Nicholson, R. A., & Kugler, K. E. (1991). Competent and incompetent criminal defendants: A quantitative review of comparative research. *Psychological Bulletin*, *109,* 355-370. doi:10.1037/0033-2909.109.3.355

Nisbett, R. E., & Wilson, T. D. (1977). Telling more than we can know: Verbal reports on mental processes. *Psychological Review, 84,* 231-259. doi:10.1037/0033-295X.84.3.231

Otto, R. K. (2006). Competency to stand trial. *Applied Psychology in Criminal Justice*, *2,* 82-113.

Otto, R. K. (2010). *Evaluation of competence to stand trial*. [Workshop]. Philadelphia, PA.

Otto, R. K., Poythress, N. G., Nicholson, R. A., Edens, J. F., Monahan, J., Bonnie, R. J., … Eisenberg, M. (1998). Psychometric properties of the MacArthur Competence Assessment Tool– Criminal Adjudication. *Psychological Assessment, 10,* 435-443. doi:10.1037/1040-3590.10.4.435

Pallant, J. (2007). *SPSS survival manual: A step by step guide to data analysis using SPSS for windows* (3rd ed.). Berkshire, England: Open University Press.

Pate v. Robinson, 383 U.S. 375 (1966).

Peil, M. (1997). *The Long-Island railroad murder - "Black rage" or "So crazy he thinks he's sane?"* Retrieved from http://www.law.cornell.edu/background/unabom/lirr.html

Petrella, R. C., & Poythress, N. G. (1983). The quality of forensic evaluations: An interdisciplinary study. *Journal of Consulting and Clinical Psychology*, *51,* 76-85. doi:10.1037/0022-006X.51.1.76

Pirelli, G., Gottdiener, W. H., & Zapf, P. A. (2011). A meta-analytic review of competency to stand trial research. *Psychology, Public Policy, and Law, 17,* 1-53. doi:10.1037/a0021713

Plotnick, S., Porter, J., & Bagby, M. (1998). Is there bias in the evaluation of fitness to stand trial? *International Journal of Law and Psychiatry, 21,* 291-304. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/9717092

Poythress, N. G., & Stock, H. (1980). Competency to stand trial: A historical review and some new data. *Journal of psychiatry and law, 8,* 131-146. Retrieved from http://heinonline.org/HOL/LandingPage?collection=journals&handle=hein.journals /jpsych8&div=19&id=&page=

Robey, A. (1965). Criteria for competency to stand trial: A checklist for psychiatrists. *The American Journal of Psychiatry, 122,* 616-623. doi:10.1176/appi.ajp.122.6.616

Robinson, R., & Acklin, M. W. (2010). Fitness in paradise: Quality of forensic reports submitted to the Hawaii judiciary. *International Journal of Law and Psychiatry, 33,* 131-137. doi:10.1016/j.ijlp.2010.03.001

Roesch, R., & Golding, S. L. (1980). *Competency to stand trial*. Urbana: University of Illinois Press.

Roesch, R., Hart, S. D., & Zapf, P. A. (1996). Conceptualizing and assessing competency to stand trial: Implications and applications of the MacArthur treatment competence model. *Psychology, Public Policy, and Law, 2,* 96-113. doi:10.1037/1076-8971.2.1.96

Roesch, R., Ogloff, J. R., & Golding, S. L. (1993). Competency to stand trial: Legal and clinical issues. *Applied & Preventive Psychology*, *2,* 43-51. doi:10.1016/S0962-1849(05)80160-X

Rosenfeld, B., & Ritchie, K. (1998). Competence to stand trial: Clinician reliability and the role of offense severity. *Journal of Forensic Science. 43*, 151-157. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/9456535

Rubin, L.H., Witkiewitz, K., St. Andre, J., & Reilly, S. (2007). Methods for handling missing data in the behavioral neurosciences: Don't throw the baby rat out with the bath water. *The Journal of Undergraduate Neuroscience Education*. *5*, A71-A77. Retrieved from http://www.funjournal.org/

Saks, M. J. & Kidd, R. F. (1986). Human information processing and adjudication: Trial by heuristics. In H. R. Arkes & K. R. Hammond (Eds.), *Judgment and decision making: An interdisciplinary reader* (pp. 213-242). New York, NY: Cambridge University Press. (Reprinted from *Law and Society Review,* 1980, 15, 123-160)

Schacht, T. E. (2005). [Letter to the Editor]. *Journal of the American Academy of Psychiatry and the Law*, *33,* p. 134.

Skeem, J. L., & Golding, S. L. (1998). Community examiners' evaluations of competence to stand trial: Common problems and suggestions for improvement. *Professional Psychology: Research and Practice*, *29,* 357-367. doi:10.1037/0735-7028.29.4.357

Tucker, L. R. (1964). A suggested alternative formulation in the developments by Hursch, Hammond, and Hursch, and by Hammond, Hursch, and Todd. *Psychological Review*, *71,* 528–530. doi:10.1037/h0047061

Tversky, A., & Kahneman, D. (1974). Judgment under uncertainty: Heuristics and biases. *Science, 185,* 1124-1131. doi:10.1126/science.185.4157.1124

Walters, G. D., White, T. W., & Greene, R. L. (1988). Use of the MMPI to identify malingering and exaggeration of psychiatric symptomatology in male prison inmates. *Journal of Consulting and Clinical Psychology*, *56,* 111-117. doi:10.1037/0022-006X.56.1.111

Warren, J. I., Murrie, D. C., Stejskal, W., Colwell, L. H., Morris, J., Chauhan, P., & Dietz, P. (2006). Opinion formation in evaluating the adjudicative competence and restorability of criminal defendants: A review of 8,000 evaluations. *Behavioral Sciences & the Law*, *24,* 113-132. doi:10.1002/bsl.699

Warren, J. I., Rosenfeld, B., Fitch, W. L., & Hawk, G. (1997). Forensic mental health clinical evaluation: An analysis of interstate and intersystemic differences. *Law and Human Behavior, 21,* 377-390. doi:10.1023/A:1024855118855

Way, B. B., Dvoskin, J. A., & Steadman, H. J. (1991). Forensic psychiatric inpatients served in the United States: Regional and system differences. *Bulletin of the American Academy of Psychiatry and the Law, 19,* 405-412. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/1786420

Winick, B. J. (1983). Incompetency to stand trial. In J. Monahan & H.J. Steadman (Eds.), *Mentally disordered offenders: Perspectives from law and social science* (pp. 3-38). New York, NY: Plenum Press.

Zapf, P. A., Hubbard, K. L., Cooper, V. G., Wheeles, M. C., & Ronan, K. A. (2004). Have the courts abdicated their responsibility for determination of competency to stand trial to clinicians? *Journal of Forensic Psychology Practice, 4,* 27-44. doi:10.1300/J158v04n01_02

Appendix A: Sample Protocol

Defendant 01
All information that may reasonably be used to identify someone has been removed

Demographic Information:_____

Symptoms:
_____
_____
_____
_____
_____
_____

Psychiatric Diagnoses: Axis I _____
                 Axis II _____
                 Axis III _____

Full Scale Intelligence Quotient: _____

> **≥ 110 = high average to very superior intellectual functioning;**
> **90-109 = average; ≤ 89 = low average to severely impaired**

Mini-Mental Status Examination: _____

> **27-30 = intact orientation and cognitive functioning;**
> **11-20 = moderately impaired; 0-10 = severely impaired**

Competence Assessment to Stand Trial for Defendants with Mental Retardation:

Section 1 Basic Legal Concepts: ___
Section 2 Skills to Assist Defense: ___
Section 3 Understanding Case Events: ___
Total Score: ___

> **During the construction of this instrument, individuals with mental**
> **retardation who were deemed competent to stand trial achieved the**
> **following mean scores on this test:**
>   **Section 1:** 18.3   **Section 2:** 10.7 **Section 3:** 8   **Total Score:** 37

Psychiatric Hospitalization History:
_____
_____

Category of Criminal Offense:
_____

Decision

1) Based on the above information, I believe the defendant is (please check one option):
     _____ Competent   _____ Not Competent

2) Using the following scale, please indicate how confident you feel in your decision by circling a number in the following boxes:
*The scale ranges from 1 (not at all confident) to 5 (extremely confident)

1	2	3	4	5

Appendix B: Instructions

1. For each protocol, please make a decision as to the defendant's competency to stand trial. After reviewing the defendant's information, please indicate your decision by placing a check next to the option that best reflects your decision.

2. For this study, the term "*competence*" is defined using the standard set-forth by the U.S. Supreme Court in *Dusky* v. *U.S.* (1960) as "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)."

3. For this study, the term "*incompetence*" is defined using the standard set-forth by the Mental Health Procedures Act (1976), under section 7402 of the Pennsylvania Code as "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."

4. After making a decision regarding the defendant's competence, please rate your confidence in this decision by circling the corresponding number in the box provided.

5. For the purposes of this study, it is important that you make a decision for each protocol, even if you are unsure of your decision.

6. The base rate of incompetency has been determined to be approximately 27.5% (Pirelli, Gottdiener, & Zapf, 2011).

7. The information presented in the protocols was taken from the medical records of actual defendants referred for competency to stand trial evaluations in Pennsylvania. In Pennsylvania, psychiatrists are the only mental health professionals who may provide information to the court about a defendant's competency to stand trial. As such, information presented in the protocols was taken from reports generated by psychiatrists. When information from evaluations performed by members of the Psychology Department was available, this information was also included in the protocols.

8. The protocols may differ in the amount and depth of information provided. The term "N/A" was used to designate information that was not available. This is not a manipulation of the study, but a reflection of the information available to the original evaluator at the time of the defendant's psychiatric evaluation.

9. Duplication of the material in this binder is strictly prohibited. Do not make copies of the protocols. The information presented in the protocols was obtained through records reviews of actual defendants with completed competency to stand trial evaluations. The information presented in this binder is for this study's research purposes only.

Instructions (continued)

10. The defendant's original legal charge was replaced by the category of offense. The following
    table depicts the offense categories used in the protocols:

Offense Category	Possible Sentence and Examples
Summary Offense	Often receive a fine and up to 90 days in jail Examples: disorderly conduct, harassment
Misdemeanor	Often imprisoned less than 1 year in jail Examples: petty theft, trespassing
Felony	Often imprisoned more than 1 year in a state or federal prison facility Examples: rape, arson, murder

11. You may use outside information obtained from books, journal articles, or other sources of
    information other than what is provided in this binder. However, please make your decision
    without consulting other individuals.

12. You will have four weeks to complete this task. After four weeks, the Student Researcher will
    contact you to collect the binder. You will be asked to return the binder using the pre-paid
    mailing envelope included with the materials. The amount of time you choose to spend on this
    task in one sitting is up to you. You may choose to classify all the defendants at once or to
    take breaks in-between.

13. Whether the defendants depicted in the following protocols were ultimately found competent
    or incompetent to stand trial by the legal system will be determined by the Student Researcher
    through review of defendants' medical or court records.

References

Dusky v. U.S., 362 U.S. 402 (1960)

Pirelli, G., Gottdiener, W. H., & Zapf, P. A. (2011). A meta-analytic review of competency to stand

    trial research. *Psychology, Public Policy, and Law, 17,* 1-53. doi:10.1037/a0021713

Appendix C: Participant Letter

DATE

Participant Name
12 Willow Lane
City, State 12345

Dear Participant Name,

As a psychologist, you have unique knowledge about the experience of working with defendants whose competency to stand trial is in question. **If you have experience conducting competency to stand trial evaluations and an advanced degree (Ph.D. or Psy.D.) in psychology, I am writing to request your participation in a study that examines how individuals make decisions during competency to stand trial evaluations.** Your participation will allow for a comparison of the decision making process of individuals with different areas or levels of expertise. This includes a comparison of clinical raters (forensic psychologists and psychology graduate students) with decisions made by raters with legal training (such as attorneys). With your help, I will be able to examine the accuracy and confidence of both clinical and legal professionals making decisions of competency. This study will explore the impact of experience on accuracy, the relationship between confidence and accuracy, and the accuracy of linear models of judgments compared to human judgments. Also examined will be the decision making strategies and influential variables for competency to stand trial evaluations.

If you choose to participate in this study, you will be asked to review one-page protocols containing information obtained from actual defendants referred to a state hospital in Pennsylvania for evaluations of competency to stand trial. In Pennsylvania, psychiatrists are the only mental health professionals who may provide information to the court about a defendant's competency. As such, the information presented in the protocols was taken from reports generated by psychiatrists. When these reports referenced evaluations performed by members of the Psychology Department, this information was also included in the protocols. For each protocol, you will be asked to make a decision about the defendant's competency and rate your confidence in this decision. Whether the defendants depicted in the protocols were ultimately found competent or incompetent to stand trial by the legal system will be determined by the Student Researcher after reviewing the defendants' medical or court records.

You will have four weeks to complete the task and it is expected to take no more than two hours of your time. Once you are finished with the task, you will be asked to return the materials using a pre-paid mailing envelope. A sample protocol has been included in this letter to provide you with an example of your task. It is hoped that information obtained through this study will improve the accuracy and efficiency of competency to stand trial evaluations. As defendants with mental health continue to access the legal system, it is increasingly important to provide the courts with accurate information.

I realize that as a practicing psychologist your time is precious. Your status as a professional involved in the competency to stand trial process makes your contribution to this study all the more valuable. This research is only possible with the participation of individuals like you. If you are willing to participate in this study, please sign the enclosed Informed Consent Form, include your mailing address to ensure delivery of the materials, and return it using the pre-paid mailing envelope. Please keep the extra unsigned copy for your records. If you choose not to participate in the study, please return the unsigned copies using the pre-paid mailing envelope. Please respond to this inquiry in the next three weeks so additional participants may be contacted if needed.

Sincerely,

Sarah Ryan, M.A.   David J. LaPorte, Ph.D.
Student Researcher   Faculty Sponsor

Appendix D: Informed Consent Form

Accuracy and Confidence in Competency to Stand Trial Evaluations:
A Comparison of Clinical and Legal Decision Making

You have been selected to participate in a study conducted through the Department of Psychology at Indiana University of Pennsylvania (IUP) by Sarah Ryan, M.A (Student Researcher) and David J. LaPorte, Ph.D. (Faculty Sponsor). The present study seeks to examine the accuracy and confidence of clinical and legal professionals making decisions of competency to stand trial. This study will attempt to explore the impact of experience on accuracy, the relationship between confidence and accuracy, and the accuracy of linear models of judgments compared to human judgments. The study also seeks to identify decision making strategies and variables of influence when making determinations of competency to stand trial.

Participation in this study is **voluntary** and there are no adverse effects associated with a decision to not participate. If you choose to participate in this study, you will be one of nine raters. Raters will either hold an advanced degree in clinical psychology (Ph.D. or Psy.D.), an advanced degree in law (J.D.), or will be enrolled in a clinical psychology graduate program. You will be asked to make determinations of competency for 32 defendants and rate your confidence in each decision. The defendant's information will be presented in the form of one-page protocols. Information presented in the protocols was taken from actual defendants referred for competency to stand trial evaluations in Pennsylvania. In Pennsylvania, psychiatrists are the only mental health professionals who may provide information to the court about a defendant's competency. As such, the information presented in the protocols was taken from reports generated by psychiatrists. When information from evaluations performed by members of the Psychology Department was available, this information was also included in the protocols. Whether the defendants depicted in the protocols were ultimately found competent or incompetent to stand trial by the legal system will be determined by the Student Researcher after reviewing the defendants' medical or court records.

These protocols and instructions for completing the task will be given to you in a three-ring binder. Participation is expected to take less than two hours of your time. You will have four weeks to complete the task, by which time you will be contacted by the Student Researcher, who will retrieve the binder. After information has been gathered from all participants, you will receive a letter in the mail providing a detailed description of the study's purpose, as well as contact information if you wish to receive the results of the study. It is hoped that information obtained through this study will improve the accuracy and efficiency of competency to stand trial evaluations.

Your responses to this study are anonymous. Your name will not be associated with your answers. This study poses no known risks to participants. You may withdraw from the study at any time without negative consequences. To withdraw from the study, contact the Student Researcher, who will destroy all data you have provided. The results of this study may be published in professional journals and presentations. All publications will exclude any information that will make it possible to identify you as a subject. If you are willing to participate in this study, please sign the statement below and return it using the pre-paid mailing envelope. Please keep the extra unsigned copy for your records. If you choose not to participate in the study, please return the unsigned copies using the pre-paid mailing envelope. Please contact Sarah Ryan with further questions.

Student Researcher: Faculty Sponsor:
Sarah Ryan, M.A.   David J. LaPorte, Ph.D.
Clinical Psychology Doctoral Candidate   Director of Doctoral Studies
Department of Psychology   Department of Psychology
Uhler Hall, 1020 Oakland Avenue   Uhler Hall, 1020 Oakland Avenue
Indiana, PA 15705   Indiana, PA 15705
Phone: 302/853-0088 Email: s.e.ryan2@iup.edu   Phone: 724/357-4524 Email: laporte@iup.edu

This project has been approved by the Indiana University of Pennsylvania Institutional Review Board for the Protection of Human Subjects (Phone: 724/357-7730).

Informed Consent Form (continued)

Accuracy and Confidence in Competency to Stand Trial Evaluations:
A Comparison of Clinical and Legal Decision Making

VOLUNTARY CONSENT FORM:

I have read and understand the information on the form and I consent to volunteer to be a subject in this study. I understand that my responses are completely anonymous and that I have the right to withdraw at any time. I have received an unsigned copy of this informed Consent Form to keep in my possession.

Name (PLEASE PRINT): _____

Signature: _____

Date: _____

Phone number and location where you can be reached: _____

_____

Best days and times to reach you: _____

I certify that I have explained to the above individual the nature and purpose, the potential benefits, and possible risks associated with participating in this research study, and have answered any questions that have been raised.

_____   _____
Date   Investigator's Signature

Appendix E: Debriefing Letter

DATE

Participant Name
12 Willow Lane
City, State 12345

Dear Participant Name,

Thank you for your participation in this study. This study sought to examine the accuracy and confidence of clinical and legal professionals making decisions of competency to stand trial. The decision-making strategies used by participants and the variables that influenced these decisions will be examined along with the impact of experience on accuracy, the relationship between confidence and accuracy, and the accuracy of linear models of judgments compared to human judgments.

Although competency to stand trial is primarily a legal decision, research suggests the agreement rate between the opinions of judges and evaluators is as high as 99.7% (Zapf, Hubbard, Cooper, Wheeles, & Ronan, 2004). This extreme reliance on the opinions of mental health professionals underscores the importance of providing the courts with reliable and valid evaluations of competency (Skeem & Golding, 1998). The correct identification of incompetent defendants ensures the protection of individuals' right to a fair trial between matched adversaries (Grisso, 1988; Mossman et al., 2007). Therefore, it is essential that clinicians perform reliable and valid evaluations of defendants. An understanding of the process by which these decisions are rendered and the validity of these decisions is essential if mental health professionals are to provide expert witness testimony for the court system (Faust, 2003). Further, identifying the variables that contribute to these decisions could improve the accuracy and efficiency of conducting competence to stand trial evaluations. This could aid in decreasing the amount of time, and thus cost, spent performing these evaluations in the future. As defendants with mental health continue to access the legal system, it is increasingly important to provide the courts with the most accurate and reliable information.

If you are interested in obtaining the results of this study, please contact Sarah Ryan at s.e.ryan2@iup.edu.

Sincerely,

Sarah Ryan, M.A.   David J. LaPorte, Ph.D.
Student Researcher   Faculty Sponsor