A Meta-Analytic Review of Competency to Stand Trial Research

by

Gianni Pirelli, M.A.

A dissertation submitted to the Graduate Faculty in Psychology in partial fulfillment of the requirements for the degree of Doctor of Philosophy,

The City University of New York

2010

This manuscript has been read and accepted for the

Graduate Faculty in Psychology in satisfaction of the

dissertation requirement for the degree of Doctor of Philosophy.

Gianni Pirelli, M.A.

Date

Chair of Examining Committee

William H. Gottdiener, Ph.D.

Date

Executive Officer

Maureen O'Connor JD, Ph.D.

Patricia A. Zapf, Ph.D.

Michele Galietta, Ph.D.

Barry Rosenfeld, Ph.D., ABPP

Robert Bornstein, Ph.D.

Supervisory Committee

THE CITY UNIVERSITY OF NEW YORK

ABSTRACT

A META-ANALYTIC REVIEW OF COMPETENCY TO STAND TRIAL RESEARCH

by

Gianni Pirelli, M.A.

Advisor: Professor William H. Gottdiener

The notion that an individual must be competent to stand trial dates back to 14th century England and has been well documented in English case law. The competency standard in the United States was derived from *Dusky v. United States* (1960). Since the 1960s the number of empirical investigations in this area of research has grown considerably. It is essential for applied empirical research to be conducted to provide policy makers, practitioners, and psycholegal researchers with the most useful and updated information available. A meta-analysis accomplishes this goal. Such research also has important implications for criminal defendants, as they have the most at stake in competency evaluations. The present study is a meta-analysis of competency to stand trial research. Only one meta-analysis, published nearly 20 years ago, has been conducted in this area. Since then over 100 empirical research studies have been conducted and eight competency instruments have been either published or revised. This paper presents the results of a meta-analysis of 68 studies, conducted between 1967 and 2008, comparing competent and incompetent defendants on a number of demographic, psychiatric, and criminological variables. Categorical (e.g., psychiatric diagnosis) and continuous (e.g., scores on traditional and competency assessment measures) variables commonly investigated in competency research were coded and aggregated to generate cumulative effect sizes and moderation was formally tested via meta-*F* and meta-regression analyses. The most robust findings associated with these variables were that defendants diagnosed with a Psychotic Disorder were approximately eight times more likely to be found incompetent than defendants without such a diagnosis and the likelihood of

being found incompetent was approximately double for unemployed defendants as compared to employed defendants as well as those with a previous psychiatric hospitalization versus those without such a history. Comparative data on 12 competency measures, the Wechsler intelligence scales, and the MMPI/MMPI-2 was also explored. The effect sizes associated with the competency measures were larger than those for the traditional measures; however, the findings must be cautiously interpreted because few studies with relatively small sample sizes were included in the analyses. Implications of these findings for research and practice will be discussed.

ACKNOWLEDGEMENTS

This research was supported by a Grant-in-Aid from Division 41 of the American Psychological Association: the American Psychology-Law Society (AP-LS).

I would like to thank my dissertation committee for their professional and personal support. Thanks to: Dr. Bill Gottdiener for being a great mentor who has always pushed me to think, question, and explain myself far beyond a point I would have on my own; Dr. Patty Zapf for so many things…for seeing something in me well before I could actualize it and for never leaving my corner – I am where I am because of you; Dr. Michele Galietta for readily and enthusiastically joining the committee, for being a major advocate of me professionally, and for always finding time to "just talk"; Dr. Nancy Ryba for being on the committee for the first half of the project and for many years of professional and personal guidance – you have helped me much more than you know; and Drs. Barry Rosenfeld and Bob Bornstein for serving as outside readers and, specifically, for their expertise, professionalism, and interest in this study. I would also like to thank Grazyna Kusmierska for serving as the second coder on this project.

Finally, I would like to recognize my family. Thank you to my wife, Elena, who has been my rock for 10 years; to my parents, Mario and Susan, for a lifetime of love, sacrifice and hard work, and for instilling such values in me; and to my sisters, Giulia and Gemma, for leading the way. Also to the Pasquariellos and the Reddins for proving you do not need to be related to be family.

DEDICATION

For Elena**.**

"Although the strength of these hands will undoubtedly be tested in the years to come, when joined, they are unbreakable."

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF APPENDICES

A Meta-Analytic Review of Competency to Stand Trial Research

The principal in Western jurisprudence that an individual must be competent to stand trial[1] has its roots in English common law dating back to the time of Edward I in the 14th century (Roesch & Golding, 1980) and is well documented in English case law and legal commentary (Blackstone, 1783; *Frith's Case*, 1790). The concept of competency may have stemmed from defendants who remained mute in lieu of making a plea in which case the English courts sought to determine whether their muteness was a function of "malice" or "by visitation of God" (Melton et al., 2007). The right to be competent to stand trial in American courts can be traced back to the early 19th century (*United States v. Lawrence,* 1835) and has been recognized as both a constitutional guarantee and essential to ensuring the integrity of our criminal justice system (*Drope v. Missouri,* 1975; *Youtsey v. United States,* 1899).

The current legal standard for competency to stand trial in the United States was set forth in *Dusky v. United States* (1960). In *Dusky*, the United States Supreme Court held:

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

The *Dusky* holding has been criticized for both its brevity and ambiguity by mental health professionals and legal scholars alike. Despite these concerns, the *Dusky* standard, or some variation of it, has been adopted by every state in the United States (Favole, 1983).

The presence of a mental illness (e.g., psychosis) or disability (e.g., mental retardation) relates to the threshold issue and is a minimal requirement that must be met before a defendant can be found incompetent to stand trial. In addition, a linkage between such an illness or

disability and deficits in trial-related abilities (e.g., comprehension of legal proceedings; ability to assist in one's defense) must be established (Skeem & Golding, 1998). If a defendant is found incompetent it must be determined whether his or her competence is restorable, and many states require evaluators' reports to include the likelihood of restorability (Golding, 1992). Some defendants are unlikely to have their competency restored; particularly those with permanent brain damage, severe developmental disabilities, and/or treatment resistant psychosis (Mossman, 2007; Skeem, Golding, & Emke-Francis, 2004). Commenting on the competency restoration literature, Golding (1992) concluded that prior psychiatric history, poor premorbid social functioning, negative symptoms, age of onset, insidious onset, and history of response to treatment were the most promising predictors of competency restorability or lack thereof; however, he posited that the best predictor of restorability may be initial responsiveness to treatment.

Competency to stand trial evaluations have been regarded as "the most significant mental health inquiry pursued in the system of criminal law" (Stone, 1975, p. 200) with the number conducted throughout the Unites States each year estimated to be around 60,000 (Bonnie & Grisso, 2000). Over two decades ago Winick (1985) estimated that over $185 million is spent in the United States annually for competency evaluations and related treatment (i.e., competency restoration). He has since suggested that this number may be closer to double or triple his initial estimate (Winick, 1996). Now that another decade has passed it is likely that this number is larger still.

In addition to monetary expenses, there are a number of costs associated with competency evaluations should they be conducted poorly. There is the potential of violating a defendant's *due process* rights by allowing an incompetent defendant to stand trial; or of

violating a defendant's civil rights by temporarily committing him or her to a forensic psychiatric facility for the purposes of competency restoration (i.e., pharmacotherapy) when he or she is actually competent. These concerns are particularly salient when placed within the context of the base rates of incompetency. Base rates from competency referrals and ultimate decisions of competency vary between and within jurisdictions and settings (Murrie, Boccaccini, Zapf, Warren, & Henderson, 2008; Nicholson & Kugler, 1991), but the modal jurisdictional estimate of incompetency for referred defendants is thought to be 20% (Roesch, Zapf, Golding, & Skeem, 1999). Such a low base rate has major implications for the use of screening measures used to identify clearly competent defendants thereby avoiding the costs, time, and resources required to conduct full competency evaluations. These considerations have been recognized by psychologists for decades and hundreds of articles and numerous books have been published since the 1960s aimed at developing and refining practice standards in the competency arena (e.g., Ackerman, 1999; Bonnie, 1992, 1993; Goldstein, 2003, 2007; Grisso, 1986, 2003; Heilbrun, 2001; Heilbrun, Marczyk, & DeMatteo, 2002; Melton, Petrila, Poythress, & Slobogin, 1997; Melton et al., 2007; Zapf & Roesch, 2009).

Types of Competency Research

The competency literature is comprised of three major areas of investigation: (a) the correlates of competency, (b) performance of incompetent and competent defendants on traditional psychological tests, and (c) performance of incompetent and competent defendants on specialized competency assessment measures. Studies on the correlates of competency have primarily investigated the relationship between competency status (i.e., incompetent or competent) and various demographic, psycholegal/criminological, and clinical variables. The most commonly researched variables in this regard are: ethnicity, sex, employment status, and

marital status (demographic); type of current criminal charge (e.g., violent or non-violent), and competency evaluation history (psycholegal/criminological); and, psychiatric diagnosis and psychiatric hospitalization history (clinical). Researchers examining the performance of defendants on traditional assessment measures (e.g., the WAIS) have investigated associations between scores on these measures and competency status and/or the aforementioned demographic, psycholegal/criminological, and clinical variables. Investigators incorporating one or more of the 12 existing competency assessment instruments have also examined such correlations, but many of these studies have been psychometric investigations instead, whereby reliability and validity evidence for the competency measures was explored.

Adjudicative competency research has been published steadily since the 1960s; however, there is a dearth of review literature in the competency arena. There have been three 5-year research updates (i.e., qualitative reviews) compiled by Grisso and colleagues since 1992 (Cooper & Grisso, 1997; Grisso, 1992; Mumley, Tillbrook, & Grisso, 2003) and only one meta-analysis that was conducted nearly two decades ago (Nicholson & Kugler, 1991). Qualitative and quantitative reviews are essential for psycholegal researchers and practitioners as they provide summaries of acquired knowledge in a given area of study, which facilitates conceptual and practical advancements in the area (e.g., models, theories, standards of practice). Such reviews are particularly important in this area because of the large number of empirical investigations conducted in the past 50 years.

Qualitative Reviews of Competency Research

Grisso and colleagues (Cooper & Grisso, 1997; Grisso, 1992; Mumley et al., 2003) conducted three qualitative five-year reviews over the past two decades. The reviews were divided into seven areas, which, according to Grisso, parallel the competency assessment

process: (i) the systemic context of competency to stand trial evaluations; (ii) conceptual definitions of competence and models for competency to stand trial assessment; (iii) research on competency assessment methods; (iv) characteristics of incompetent defendants; (v) interpretation of competency evaluation data; (vi) issues in competency assessment of special populations; and (vii) treatment to restore competence.

Grisso (1992). In the first review, Grisso (1992) summarized the competency research published between 1986 and 1990 (and exceptions were made to include a few studies published in 1991). With respect to systemic issues, Grisso (1992) did not find support for the notion that referrals for competency evaluations were becoming a mechanism for committing non-dangerous individuals; however, he found support for conducting competency evaluations at the outpatient level. In terms of conceptual definitions of competence and relevant models of assessment, Grisso cited various theoretical developments and noted the development of standardized instruments would be a necessary next step in this area.

To examine the research relevant to competency assessment methods, Grisso reviewed studies investigating the psychometric properties and behavioral correlates of various competency assessment instruments. Only one study examined the relationship between behavioral observation ratings (of two independent examiners) and scores on a competency assessment instrument, wherein mock competency to stand trial hearings were conducted – a method Grisso supported and believed would be promising for future research. The following section focused on studies investigating the empirical correlates of clinicians' competency judgments (i.e., the characteristics of incompetent and competent defendants), which included the results of Nicholson and Kugler's (1991) meta-analysis.

In the following section on interpretation of competency evaluation data, Grisso mentioned studies regarding such issues as mutism, malingering, delusions, and retrospective competence, and no noteworthy findings were discussed. With respect to issues related to the competency assessment of special populations, Grisso noted that research on defendants diagnosed with Mental Retardation as well on juveniles and those diagnosed with AIDS-related dementia was basically nonexistent prior to 1986. Last, Grisso found support for treatment to restore competence. He concluded the review by providing three main directions for future research: (i) to evaluate the actual demands associated with the role of the defendant in the criminal adjudication process; (ii) to develop a model of competency, with respect to defendants' abilities, to facilitate examiners' decision-making; and, (iii) to develop a standardized competency assessment instrument.

Cooper & Grisso (1997). The second review in this series was conducted by Cooper and Grisso (1997) and included studies published between 1991 and 1995. The researchers found a trend for the decreased civil commitment of competency examinees and suggested that additional research was needed before drawing conclusions regarding outpatient versus inpatient evaluations and treatment. With respect to conceptual definitions of competency, the authors highlighted Bonnie's (1992) reconceptualization of competence as well as the decision in *Godinez v. Moran* (1993). A handful of studies investigating the psychometric properties of various assessment instruments were reviewed as well as were those examining the role of traditional psychological instruments versus forensic instruments in competency evaluations.

In addition, Cooper and Grisso interpreted their review of Nicholson and Kugler's (1991) findings as evidencing little to no support for the relationship between demographic variables and clinical judgments of competency. Cooper and Grisso cited a number of studies wherein

reports were found to be inadequate and incongruent with evolving standards of practice, and stated that no studies examining the communication of competency findings (i.e., competency reports) had been conducted before the second five-year review. The increased size of the special populations sections in this review illustrated the growing body of research in that time period in two main areas: the assessment of those diagnosed with Mental Retardation and juveniles. Last, there was a substantial increase in research relevant to competency restoration during this time; specifically, Cooper and Grisso noted that new publications during this time period increased approximately 33% (from 51 to 69) as compared to the time period from the previous review. They also predicted a significant increase in research in three main areas by the time of the following (i.e., third) five-year review: (i) specialized instruments for evaluating competence-related abilities; (ii) juveniles' competency to stand trial; and, (iii) description and evaluation of the work of forensic mental health examiners in competency evaluations.

Mumley et al. (2003). In the third review in the series, Mumley and colleagues (2003) noted that findings from studies in this time period suggested potential support for a "subversion" hypothesis (i.e., using competency evaluations for a "back door" manner to provide defendants with treatment) within some state legal systems; however, they acknowledged such inferences remain tenuous. In addition, results from studies of service delivery systems continued to be variable. Mumley and colleagues noted that two previously conceptualized advancements in the area (i.e., Bonnie's 1992 reformulation and the *Godinez* decision) continued to be addressed in the literature. The third review found that fewer studies on competency assessment methods were conducted during the respective time period than the previous two; however, four new competency assessment measures were developed and increased attention to malingering considerations in competency evaluations was also found. As hypothesized in the previous

review, numerous studies were conducted investigating the quality of competency evaluations and reports during this time period. The findings in this area suggested that, while competency assessment practices were improving, they continued to fall short of aspirational professional standards. Research on special populations addressed in the previous two reviews (i.e., defendants diagnosed with Mental Retardation and juveniles) in addition to research on female defendants increased during this time period as compared to the previous two time periods. Mumley and colleagues concluded that, while research in the competency arena has continually grown, research during this time period was notable in three main ways: (i) the development of new competency assessment instruments; (ii) the introduction of research on the quality of competency assessments and reports; and, (iii) conceptual and empirical information pertaining specifically to the competency assessment of juveniles. They hypothesized that research in these three areas would continue to grow during the time covered by the next five-year research update.

Summary. In their series of qualitative reviews, Grisso and colleagues' (Grisso, 1992; Cooper & Grisso, 1997; Mumley et al., 2003) addressed seven areas: systemic issues, conceptual guidelines for evaluators, assessment methods, empirical correlates of competency, quality of competency evaluations and reports, special populations, and treatment of incompetent defendants. The authors provided the field with a template for competency research and commentary by delineating the aforementioned topics, thereby setting the stage for theory formulation and spur future research; however, these reviews consist of limitations inherent to most, if not all, qualitative reviews. First, no formal inclusion criteria was reported; therefore, studies may not have been as heavily vetted and they may have been subsequently aggregated inappropriately (i.e., the apples and oranges concept). Second, it is difficult to provide an overall

summary of results and implications of research literature when the findings across studies are not completely consistent. Thus, the authors engaged in vote counting (i.e., dependent on significance testing), such that evidentiary support was based on the number of studies with such findings rather than calculating effect sizes. For example, in qualitative reviews, three studies finding a non-significant relationship between two variables would likely hold more weight than one study with significant findings with no attention paid the effect size statistics. Third, the reviews were independently written; therefore, the results across them were not combined, but rather, compared. Thus, the authors were only able to make relative judgments (i.e., there is an increase in research in this area compared to the previous review) as opposed to aggregating the results from the three reviews to support or fail to support specific hypotheses. As a result of these limitations, the reviews may be of limited utility to the intended audience (i.e., forensic mental health professionals, judges, and lawyers).

A quantitative research synthesis, or meta-analysis, can address the limitations of qualitative research syntheses. A meta-analysis can be valuable by allowing for the exploration of hypotheses not previously evaluated in primary studies and those that cannot be tested by primary studies alone, including the investigation of potential moderator variables. In addition, a meta-analysis can be used to track changes and differences in study outcomes across a large body of literature, which is accomplished via calculation and analyses of effect size statistics. Nevertheless, a reasonable concern is that quantitative and qualitative reviews incorporate different studies to draw conclusions regarding the same concepts. The following section provides a detailed description of the only meta-analysis conducted in the competency arena.

Quantitative Review of Competency Research

In the only meta-analysis conducted in this area of research, Nicholson and Kugler (1991) synthesized the findings of 27 studies from 1967-1989 that compared competent and incompetent defendants.[2] Nearly all of the studies (26; 86.7%) included in their synthesis used mental health professionals' decisions to define competency status. In the remaining studies, one employed a blue ribbon panel of experts to decide upon competency status, one relied solely on court decision, and two utilized some combination of decision from mental health professionals, lawyers, and the courts. Most of the reviewed studies (22; 73.3%) were conducted in inpatient facilities, while the remaining studies were conducted in outpatient settings. Data collection in ten, or one-third, of the studies took place prior to the landmark competency ruling in *Jackson v. Indiana* (1972)[3] and the remaining 20 occurred after the ruling.

Calculation and analysis of effect sizes. Nicholson and Kugler (1991) calculated Pearson correlation coefficients as measures of effect size in their meta-analysis, which is the same method Nicholson (1986) used in his meta-analysis investigating the correlates of civil commitment published five years earlier. Phi coefficients (Φ) were generated for 2 x 2 categorical analyses (e.g., the relationship between competency status and sex) and point-biserial correlations were calculated to investigate the relationship between categorical and continuous variables (e.g., competency status and years of education). Unfortunately, utilizing correlation coefficients, especially the phi coefficient, presents major limitations for use with these data. Despite acknowledging these limitations, Nicholson employed these statistics in both of his meta-analyses using formulas provided by Nunnally (1978). The following is a footnote in Nicholson's first meta-analysis (1986):

> The magnitude of a correlation depends in part on the degree of correspondence between the distributions of the variables being correlated. Where one or both of the variables are dichotomous, the effect of differences in the shapes of the distributions can be considerable. Hence, the magnitudes of the correlations in the present review are probably constrained by discrepancies between the distributions of the correlated variables (see Nunnally, 1978, chap. 4, pp. 141-146). (p. 242)

Not only do phi coefficients underestimate the population correlation coefficient (Sanchez-Meca et al., 2003), they can wildly vacillate as a function of marginal proportions and/or sample sizes alone, whereas odds ratios cannot (Kline, 2004). Haddock and colleagues (1998) illustrated this exact point with simulated data. Table 1 presents an adaptation of Haddock et al.'s illustration whereby their data on the relationship between exposure to a toxin and disease occurrence has been substituted with the association between sex and competency status. These data demonstrate the substantial effects of sample size on the phi coefficient as well as the complete stability of the odds ratio regardless of sample size.

Table 1. *Simulation of the Comparison between Sex and Competency Status*

Proportion	Female		Male			
Female	Incompetent	Competent	Incompetent	Competent	Φ	Odds Ratio
5%	40	10	190	760	.31	16.0
10%	80	20	180	720	.41	16.0
15%	120	30	170	680	.47	16.0
25%	200	50	150	600	.54	16.0
50%	400	100	100	400	.60	16.0

The use of correlation coefficients is particularly problematic for competency data because of the base-rate of incompetency, which approaches 30%. As Lipsey and Wilson (2001) articulated:

> …the maximum possible phi value for a 2 x 2 table with a 90-10 split on one variable and 50-50 split on the other is .33; considerably less than 1. For an 80-20 split the maximum only climbs to .5. The odds-ratio, on the other hand, is insensitive to changes in the marginal proportions, that is, the proportion split for each dichotomy, and is therefore well suited to represent low frequency events. (p. 60-61)

In addition to being statistically problematic, Nicholson and Kugler's (1991) use of correlation coefficients to handle 2 x 2 analyses is conceptually confusing. That is, they utilized the word "likelihood" throughout the explanation of their results. For example, "Older defendants were more likely to be found incompetent than younger defendants. In addition, women and members of minority groups were more likely to be judged incompetent" (p. 359). A correlation coefficient is a measure of association; however, it cannot directly speak to actual levels of likelihood. Nicholson and Kugler attempted to circumvent the problems associated phi coefficients by subsequently calculating Cohen's *h* statistics. This statistic remains problematic, however, because it is a member of the *d* family and, therefore, represents the difference between two proportions rather than a measure of likelihood (Rosenthal, 1994).

Nicholson and Kugler (1991) also utilized correlation coefficients as measures of effect size to investigate the relationship between competency status and continuous outcome data. Point-biserial correlation coefficients are subject to the same range restriction problems previously outlined for categorical data (Lipsey & Wilson, 2001) and they are also conceptually inappropriate for these data. The primary objective of both the present and previous meta-analysis was to synthesize data from studies comparing competent and incompetent defendants

on a number of variables. As such, effect size data in the form of difference-statistics is more consistent with a comparative approach. Furthermore, while correlation coefficients are easily interpretable and likely more familiar to most researchers and practitioners, they are actually more convoluted and abstract than difference-statistics for these data.

Results. Nicholson and Kugler (1991) found the base rate of incompetency to be just over 30% across studies. The vast majority of defendants in the reviewed studies were male; in fact, when studies using male-only and female-only samples were excluded, males comprised nearly 90% of the samples. With respect to minority status, the mean across samples was 36.6% minority defendants; however, there was substantial variability in this regard (i.e., range: 3-70%). The samples included defendants with relatively few social and economic resources. The average defendant received 9.6 years of formal education, 57% had never been married, and 67.7% were unemployed. In terms of legal and psychiatric factors, approximately 55% of defendants across samples had been charged with a violent offense, 50.4% had previous arrest histories, and 38.6% had a history of hospitalization. The mean IQ score for defendants was 82.7 (only 6.2% were formally diagnosed with Mental Retardation) and 38.5% were diagnosed with a psychotic disorder.

Four of the six demographic variables measured evidenced significant correlations with competency status (i.e., age, gender, race, and marital resources). Specifically, older defendants were more likely to be deemed incompetent than their younger counterparts; females, minority defendants, and those who were not married were more likely to be found incompetent than their respective counterparts. It is important to highlight the weak relation of these variables to competency decision (i.e., *r*s ranged from .06 to .09). No significant relations were found for level of education or employment status and competency decision. With regard to legal and

psychiatric history, previous legal involvement (r = .17) and previous hospitalization (r = .26) were significantly related to competency, such that defendants without prior legal involvement and those with a history of previous psychiatric hospitalization were more likely to be found incompetent than their counterparts. Last, no correlation was found between type of offense (i.e., violent or nonviolent) and competency status.

Numerous significant correlations were found between psychiatric variables and competency status. Diagnosis of a psychotic disorder was strongly related to competency status (r = .45); however, no such relationship was found between a diagnosis of Mental Retardation and competency. The researchers investigated correlations between eight types of symptoms of psychopathology, adapted from descriptions outlined in the Interdisciplinary Fitness Interview (IFI) manual, and competency status. Most of the symptom categories were significantly related to competency. The associated *r*s were: disorientation (.43); delusions (.36); hallucinations (.29); impaired memory (.28); impaired thought or communication (.25); and disturbed behavior (.25). Affective disturbance did not correlate with competency and impaired judgment was only non-significant (.38; p = .06).

Nicholson and Kugler (1991) also investigated performance on various traditional psychological measures and competency assessment instruments. They found a significant negative correlation between Intelligence Quotient (IQ) and competency status (-.16), such that incompetent defendants scored lower on intelligence tests than their competent counterparts. In addition, four MMPI scales were significantly related to competency. Incompetent defendants scored higher than competent defendants on scales F (Infrequency, or "fake bad"), 6 (Paranoia), and 8 (Schizophrenia), which are designed to tap into severe psychopathology, and they also scored higher on scale 5 (Masculinity-Femininity), but the researchers were unable to explain

this finding. All correlations for the MMPI scales and competency status were small, ranging from .05 to .08. Scores on four competency assessment instruments were also investigated. Performance on the *Competency Screening Test* (CST; Lipsitt, Lelos, & McGarry, 1971), the *Georgia Court Competency Test* (GCCT; Nicholson, Briggs, & Robertson, 1988), the *Competency Assessment Instrument* (CAI; Laboratory of Community Psychiatry, 1973), and the *Interdisciplinary Fitness Interview* (IFI; Golding, 1993) was associated with competency decisions, such that defendants who scored poorly on these measures were more likely to be found incompetent. Although the correlations were generally large (CST = -.37; GCCT = -.42; CAI = -.52; IFI = -.42), relatively few studies employed each measure (CST: 11; GCCT: 4; CAI: 2; IFI: 1). Contrasting these findings with those associated with the aforementioned traditional measures led the researchers to "question the importance of traditional assessment instruments in the evaluation of competency" (p. 363).

Although Nicholson and Kugler (1991) did not formally test for moderation in their meta-analysis, they speculated about potential moderators via correlation analyses; namely, the source of the competency decision (e.g., mental health professional vs. court decision); the setting of a study (e.g., inpatient vs. outpatient); and the date of data collection (e.g., before or after particular landmark rulings). They acknowledged a confluence of the competency decision variable would likely exist as a function of the high agreement often found between judges and mental health professionals on such decisions.

Limitations. Nicholson and Kugler's (1991) meta-analysis consists of a number of limitations. First, the authors only reviewed published studies, and therefore, there is the potential for the "file-drawer" problem. That is, some studies are not published for various reasons (e.g., dissertations, conference proceedings), but still contain information useful for

meta-analysis. Second, there are major limitations to the way in which Nicholson and Kugler calculated effect sizes; namely, their use of correlation coefficients (see the Method subsection below entitled: *Calculation of Effect Sizes and Statistical Modeling*). Third, specific findings were constrained because they were based on the aggregation of data from few studies within each variable category. For example, in their comparison of the relations between competency assessment instruments, traditional psychological instruments, and competency decisions, Nicholson and Kugler were only able to include data from a handful of studies. Specifically, 8 studies contained IQ scores, 5 included MMPI scores, 11 had Competency Screening Test scores, 4 had Georgia Court Competency Test scores, 2 had Competency Assessment Instrument scores, and 1 had Interdisciplinary Fitness Interview scores. The authors found stronger relations between the competency assessment instruments and competency decisions, which led them to question the use of traditional measures in competency evaluations as compared to specific competency measures; however, these findings can be misused if not cautiously interpreted. The authors acknowledged that traditional and forensic assessment measures were not compared within the same studies and that relatively few studies were analyzed for each instrument. Fourth, the authors did not formally test for moderation, which is an inherent benefit of conducting a meta-analysis. Fifth, the meta-analysis was conducted nearly 20 years ago and prior to more recent landmark court decisions (e.g., *Godinez v. Moran*, 1993). Additionally, eight competency assessment instruments have been published and numerous empirical studies have been conducted since the previous meta-analysis. The present meta-analysis (i.e., the present study) circumvents the aforementioned limitations and accounts for other potential limitations (see the Present Study section).

Hypotheses

The following hypotheses were tendered based on the findings of the aforementioned qualitative reviews, previous meta-analysis, and primary research published in the competency arena:

(i) The mean base rate of incompetency will be between 20-30%.

(ii) Demographic variables will relate to competency status; specifically, incompetency would be associated with ethnicity (i.e., Non-Caucasian); sex (i.e., female); employment (i.e., unemployed); and marital status (i.e., not married).

(iii) Psychiatric and psycholegal variables (i.e., Psychotic Disorder diagnosis, previous psychiatric hospitalizations, previous competency evaluation history, and non-violent current criminal charge) will be correlated with findings of incompetency.

(iv) Six study-level variables will serve as potential moderators for the aforementioned relations between demographic, psycholegal, criminological, and clinical variables, and competency status: (i) type of publication; (ii) source of competency decision; (iii) type of competent group; (iv) setting; (v) country; and (vi) recruitment method.

(v) Scores on competency assessment measures and traditional measures (i.e., intellectual and personality assessment instruments) will both be correlated with competency decisions; however, larger effect sizes are anticipated for the relationship between scores on competency assessment instruments and such decisions.

Two decades have passed, hundreds of empirical investigations have been conducted, eight competency assessment instruments have been developed and/or revised, and major court decisions (e.g., *Godinez v. Moran*, 1993) have been held since Nicholson and Kugler's (1991) meta-analysis; therefore, the field is in need of a second quantitative synthesis of the adjudicative competency research. The present meta-analysis is intended to provide psycholegal researchers and practitioners with a summary of the cumulative knowledge gained over 50 years of research in this area and aims to advance the state of knowledge in the field by testing hypotheses not previously tested in primary studies and those that cannot be tested by primary studies alone.

Method

The present study is a meta-analysis of 68 studies, conducted between 1967 and 2008, comparing competent and incompetent defendants on a number of demographic, psychiatric, and criminological variables.[4] Only one meta-analysis has been conducted in this area (Nicholson & Kugler, 1991) and over 100 empirical research studies have been conducted and eight competency instruments have been either published or revised since.

Literature Search

A comprehensive search consisting of five methods was conducted to identify empirical research studies in this area: (i) obtaining references of those found in acquired reports; (ii) consulting with experts in the area; (iii) searching electronic and print abstract databases; (iv) incidental browsing of libraries and bookstores; and, (v) searching citation indexes. Documents not available at local libraries were retrieved primarily through inter-library loan. A wide net was cast across five main electronic databases: (a) *PsycInfo;* (b) *PsycArticles;* (c) *Medline;* (d) *Criminal Justice Periodicals Index 1981-2007*; and (e) *National Criminal Justice Reference Service*, using three keywords: (i) competenc* to stand trial; (ii) adjudicative competenc*; and

(iii) trial competenc* (Note: using an asterisk enables searching of various endings of the root word, such as *competency* and *competence*). Electronic searches yielded 154 potentially relevant reports and the other search methods yielded an additional 32 reports, equaling 186 potential reports. Retrieval of all studies ever conducted is impossible; however, publication and sampling bias was addressed by conducting a thorough literature search, which included dissertations, and through statistical methods. The *fail-safe N* statistic, developed by Rosenthal (1979), was computed contemporaneously with effect sizes to estimate the potential effects of studies not retrieved with null results or results in the opposite direction of the mean effect size. Put differently, the *fail-safe N* is an estimate of the number of unpublished studies finding null results to render a cumulative effect size non-significant.

Inclusion/Exclusion Criteria

Although inclusion/exclusion criteria must be developed in an iterative manner, preliminary criteria were predetermined. Studies were considered for inclusion if they compared competent and incompetent defendants on at least one variable for which an effect size was calculable. Of the 186 potential reports reviewed, 88 met inclusion criteria. Of the 88 included reports, only 68 independent studies were identified and represented the total sample size (n=68) for the current meta-analysis (i.e., 20 reports were of redundant samples and added no new coding information). Of the 98 excluded reports, 88 were considered independent studies. The reference list for all included studies is presented in Appendix A and the excluded study list is presented in Appendix B. Redundant reports are included in each list. A study was typically excluded for one of four reasons. Of the excluded studies, (i) 35% did not utilize a competent comparison group; (ii) 25% were competency restoration studies; (iii) approximately 16% met the main inclusion criteria but did not present sufficient data to code; and (iv) approximately

10% included only participants diagnosed with Mental Retardation. An additional 14% of the excluded studies were not included for other reasons (e.g., samples completely consisting of malingerers or coached simulators, a juvenile comparison group only, a case study, and an attorney survey).

Coding and Interrater Reliability

Coding manuals and forms were developed iteratively and revised as needed. Two forms/manuals were used in the present study: one for study-level variables (see Appendices C and D) and one for continuous outcomes (e.g., scores on a competency instrument; see Appendices E and F). These forms were created in FileMaker Pro to facilitate citation retrieval and coding, and for the maintenance of records of retrieved reports. FileMaker Pro is particularly useful for meta-analysis research because coding is completed directly on the computer and data can be exported into Excel and other statistical software programs.

Study-level variables were defined as those related to sample characteristics (e.g., number of females in the Incompetent group) as well as those related to study design (e.g., setting of study, such as inpatient). As such, the majority of the study-level coding consisted of rates and proportions based on categorical data. Continuous outcome variables were coded for data derived from scores on both traditional and competency assessment measures.

All reports were coded by the first author and 20% (i.e., 14) of the reports were coded by a fifth-year Ph.D. student experienced in conducting meta-analyses. The second-coding procedure consisted of a number of steps. An initial training session was conducted to review the coding manual and to provide an overview of the competency literature. This session was followed by the practice coding of 10 studies chosen via an online random number generator by both the first author and second coder. The coders met to address inconsistencies once practice

coding was complete. The coding manual was subsequently revised to address all concerns elicited during the practice-coding step. A second training session was provided focusing on the revisions implemented in the coding manual. Finally, a systematic random selection procedure was employed to generate interrater reliability statistics; specifically, every third study from the possible 68 studies (listed alphabetically) was chosen for inclusion in the second coding procedure, only one of which was previously coded during the practice session.

A total of 1,284 coding decisions (study- and effect-size level) were made across 14 studies of which the first author and second coder demonstrated strong agreement. A Pearson product-moment correlation analysis was conducted on the coding of 1,102 continuous variables. High interrater reliability and significant statistical agreement and was found; intraclass correlation coefficient $r = .95$, $p < .001$. A kappa statistic (κ) was computed to determine the level of agreement between the coders on 182 categorical variables; it was .77, and the agreement rate was approximately 81%. Although the interpretation of the kappa statistic has been debated over the years, existing benchmarks would classify a kappa of .77 as an overall high level of agreement. This kappa statistic is considered "substantial" based on Landis and Koch's (1977) classification, "good" as per Altman (1991); and "excellent" as per Fleiss (2003).

Calculation of Effect Sizes and Statistical Modeling

The majority of data presented in the competency research literature can be conceptualized in the context of 2 x 2 tables because it frequently involves an investigation of the relationship between competency status (i.e., competent/incompetent) and another dichotomous variables (e.g., psychotic/not psychotic); therefore, odds ratios (ORs) were calculated as effect sizes for these categorical data. Odds ratios and their statistical variants (e.g., log-odds ratios, logit models, logistic regression models) are the recommended statistics for meta-analyses

utilizing 2 x 2 tables (Fleiss, 1981; Haddock, Rindskopf, & Shadish, 1998; Sanchez-Meca, Chacon-Moscoso, & Marin-Martinez, 2003; Sandercock, 1989; Schumacker, 2005).

In the present meta-analysis, odds ratios were calculated to investigate the relationship between competency decision (i.e., incompetent/competent) and eight categorical variables: (i) ethnicity; (ii) sex; (iii) employment status; (iv) marital status; (v) psychiatric diagnosis; (vi) psychiatric hospitalization history; (vii) competency evaluation history; (viii) current criminal charge. Each variable was dichotomized in a manner intended to make interpretation consistent with the competency literature, such that variables associated with *incompetency* served as the reference point. Specifically, ethnicity was analyzed as Not Caucasian (yes/no); sex as Female (yes/no); employment as Unemployed (yes/no); marital status as Not Married (yes/no); psychiatric diagnosis as Psychotic Disorder (yes/no); psychiatric hospitalization history as Previous Psychiatric Hospitalization (yes/no); competency evaluation history as Previous Competency Evaluation (yes/no); and, current criminal charge as Current Violent Charge (yes/no). Using odds ratios as effect sizes and dichotomizing the aforementioned variables enabled a straightforward interpretation from which the actual level of *likelihood* was elicited (e.g., "Female defendants are X times more likely to be found Incompetent"). While odds ratios are used in the initial analyses because they are easier to interpret from a descriptive standpoint (i.e., levels of likelihood), log-odds ratios are easier to interpret than odds ratios in the context of meta-regression analyses using categorical antecedent variables because they are centered at 0, whereas odds ratios are centered at 1.

All effect size calculations were performed with Comprehensive Meta-Analysis (CMA; Borenstein, Hedges, Higgins, & Rothstein, 2005) a standard meta-analytic and effect size software package. Once all effect sizes were calculated, they were weighted by the inverse of

their variance and summed to generate an overall mean effect size statistic (i.e., the cumulative odds ratio) – a process that also controls for sampling error (Hedges & Olkin, 1985). Cumulative odds ratios were generated via a random effects model, which assumes that "each observed effect size differs from the population mean by subject-level sampling error *plus* a value that represents other sources of variability assumed to be randomly distributed" (Lipsey & Wilson, 2001, p. 119). The decision to employ a random effects model, rather than a fixed effects model, is subjective and is based on the analyst's perspective on the included studies. Cooper and Hedges (1994) recommended using a random effects model if the analyst conceptualizes the studies as different from each other in ways too complex to account for by only a few study characteristics, and if the intent of the meta-analysis is "to make inferences about a universe of such diverse studies" (p. 526).

Assessment of study quality is an important process in meta-analysis and can be investigated empirically by an investigation of moderators hypothesized to relate to study quality. Six study-level moderator variables hypothesized to be relevant to competency research were coded and subsequently analyzed in the present study: (i) type of publication; (ii) source of competency decision used for comparison; (iii) type of competent comparison group; (iv) study setting; (v) sample's country of origin; and (vi) recruitment of participants.[5] The frequencies and percentages of each potential moderator across all 68 studies are presented in Table 2. Homogeneity analyses were conducted on each of the eight categorical variables after descriptive and effect size statistics were calculated, producing a *Q* statistic. A significant *Q* indicates that the variability among effect sizes is greater than expected from sampling error alone. Formal moderation analyses were conducted if homogeneity statistics indicated significant variance across studies not due to sampling error (Hedges & Olkin, 1985). The impact of the

aforementioned moderators on each effect size was investigated via meta *F*-tests and meta-regression analyses to determine the extent to which various aspects of study design effect or predict the calculated effect sizes.[6]

Table 2. *Study Descriptors of All Included Studies (n=68)*

	# of Studies (%)
Type of Publication	
Article	56 (82.4)
Dissertation	10 (14.7)
Book	2 (2.9)
Source of Competency Decision	
Psychiatrist(s)	20 (29.4)
Mixed	17 (25.0)
Mental Health Professional Team†	14 (20.6)
Court	13 (19.1)
Psychologist(s)	2 (2.9)
Not reported	2 (2.9)
Type of Competent Group	
Referred	59 (86.8)
Pure	4 (5.9)
Restored	3 (4.4)
Study-competent	1 (1.5)
Mixed	1 (1.5)
Setting	
Inpatient	46 (67.6)
Mixed	11 (16.2)
Outpatient	9 (13.2)
Other	1 (1.5)
Not reported	1 (1.5)
Country	
USA	52 (76.5)
Canada	16 (23.5)
Recruitment	
Archival/Retrospective	40 (58.8)
Prospective	28 (41.2)
Type of Sampling	
Random/Convenience	59 (86.8)
Matched	8 (11.8)
Other	1 (1.5)

† A Mental Health Professional Team was characterized by two or more mental health professionals working together to arrive at one decision, whereas a Mixed decision refers to one that was based on various independent decisions.

The main effect size statistics used in the present meta-analysis to investigate the differences between competent and incompetent defendants on continuous outcome measures (i.e., scores on competency and traditional assessment instruments) were unstandardized and standardized mean differences – also calculated by the CMA software. The unstandardized mean difference was calculated for data generated from the same exact measure or scale across studies (e.g., the MMPI-2). The standardized mean difference (a Cohen's *d* statistic) was calculated when the same construct was measured across studies by a different measure or scale (e.g., Verbal IQ scores measured by the WAIS, WAIS-III, and WASI). Cohen (1977, 1988) set forth the following widely accepted interpretive ranges for standardized mean difference effect sizes: ≤ .20 = Small; .50 = Medium; ≥ .80 = Large. These ranges can serve as useful guidelines, but they were not empirically derived and interpretations of effect sizes are dependent on the area of study (e.g., a Medium effect size according to Cohen's ranges may be considered Large in some domains); therefore, these statistics were converted into odds ratios for interpretive purposes. Meta *F*-tests and meta-regression analyses were initially planned to investigate possible moderation; however, such analyses could not be conducted due to insufficient variability across various levels of the moderators (e.g., type of setting).

Results

Studies on adjudicative competency have been conducted consistently since the 1960s, but time-related patterns are apparent. Figure 1 illustrates the distribution of 186 empirical studies conducted between 1965 and 2008 and Figure 2 represents the distribution of the 68 independent studies included in the present meta-analysis during the same time period.

Figure 1. *Empirical Studies Published between 1965-2008 (n=186)*

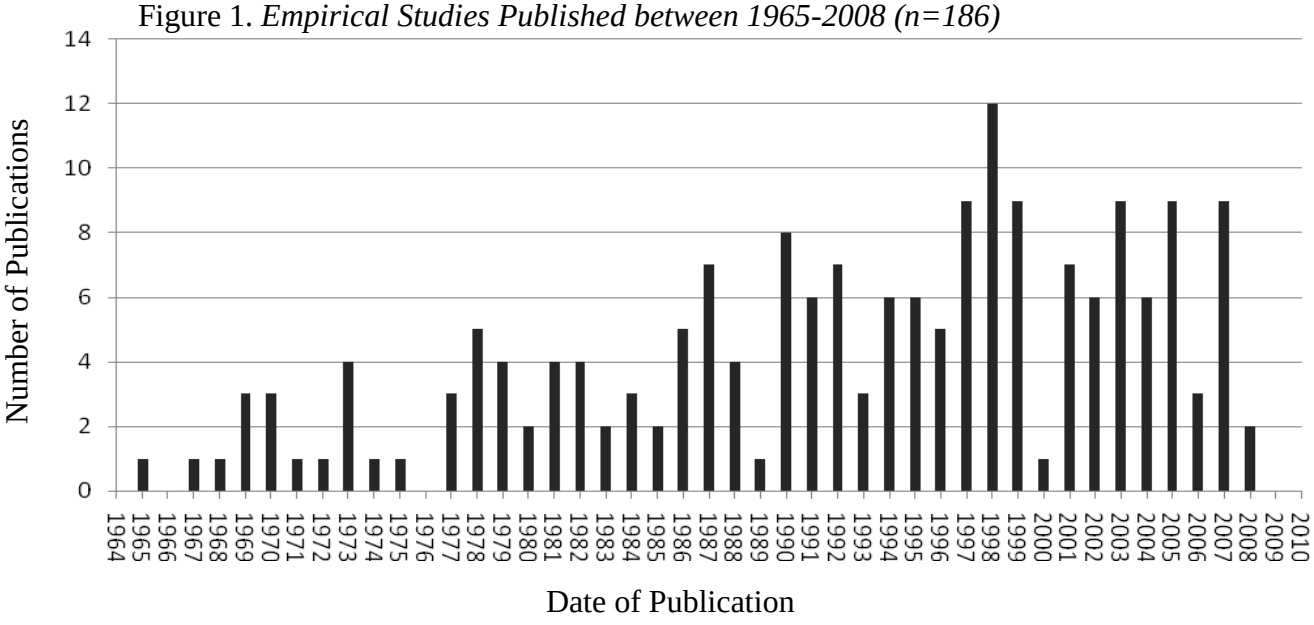
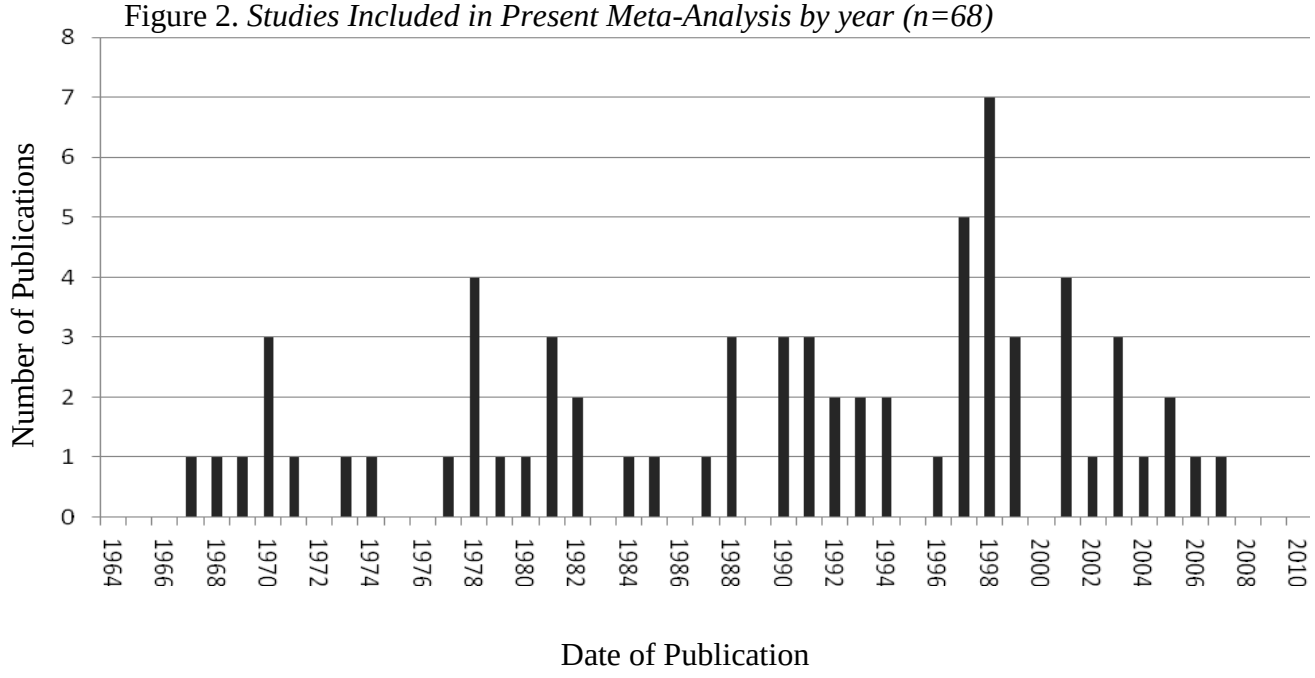


Figure 2. *Studies Included in Present Meta-Analysis by year (n=68)*



There are a number of relative peaks observable in both distributions, though they are clearer in Figure 2. Clusters of studies seem to be published around the turn of each decade contemporaneously with noteworthy events in the area. In the late 1960s and early 1970s the first competency checklists and screening measures were developed (Bukatman, Foy, & Degrazia, 1971; Robey, 1965) after the landmark ruling in *Dusky* and just prior to *Jackson v. Indiana* (1972); the late 1970s and early 1980s saw a new wave of research, much of which was conducted by Dr. Roesch, which led to the publication of a seminal text authored by Roesch and Golding (1980). Research in the early 1990s was spurred by Bonnie's (1992, 1993) reformulation of the competency construct as well as another landmark Supreme Court ruling in *Godinez v. Moran* (1993). Research in the competency arena was revitalized in the late 1990s, mainly as a function of the MacArthur Adjudicative Competence studies (Hoge et al., 1996, 1997a, 1997b; Otto et al., 1998; Poythress et al., 1998), and has remained stable throughout the 2000s.

Descriptive statistics for the entire sample of studies (n=68) are presented prior to delving into specific subgroup analyses.[7] Sample characteristics for all study participants (incompetent and competent) across all included studies are presented in Table 3 (n=26,139). Note the base rate of incompetency was 27.5% across 59 non-matched samples (median=25.3, mode=10), which supports the first hypothesis; a one-sample *t*-test was conducted to produce a 95% confidence interval around the mean estimate (25.7 – 33.4). It is also worth noting that only approximately half of the studies included female participants in their samples.

Table 3. *Sample Characteristics for all Included Studies (n=68)*

Characteristic	No. of Studies	Mean	Range
Study/Sample			
Date of Publication	68	1989.9	1967-2007
Sample Size (n)	68	384.5 (median = 176)	21-8,416
% Incompetent	59	27.5 (median = 25.33) (mode = 10)	7.0-70.0
Demographics			
Age	22	33.4	29.8-37.6
% Male	41	83.0	0-100 (50-100)*
Included Females	37	-	-
% Caucasian	22	53.4	17.0-84.0
% Not Married	10	80.7	54.0-92.0
% Unemployed	8	64.5	24.0-88.0
Education Level (in years)	14	10.4	7.8-12.0
Diagnosis			
% Psychotic Disorder	25	44.4	20.0-82.0
% Mental Retardation	16	6.3	0-23.0
% Mood Disorder	15	13.4	0-32.0
% Personality Disorder	16	18.3	0-47.0
% Substance Use Disorder	16	17.8	0-72.0
Psychiatric History			
% Previous Psych. Hospitalization(s)	5	46.1	22.0-56.0
Competency History			
% w/ Previous Competency Eval(s)	3	31.5	16.0-52.0
Criminal History			
% w/ Prior Arrest(s)	5	61.5	29.0-75.0
% Current Violent Crime	18	52.9	25.0-75.0

*Reflects the range in mixed-sex samples.

Table 4 presents characteristics of the incompetent (n=6,428) and competent (n=19,711) sub-samples of participants across all included studies. Although the sample studies are relatively large when combined, most of the data was derived from few studies (as is illustrated in the second column) and, therefore, the following descriptive statistics must be placed in that context. Compared to their competent counterparts, incompetent defendants were slightly older (35 years old versus 31.8), predominantly not Caucasian (52.3% versus 43.1%), had a much higher unemployment rate (70.8% versus 58.2%) and a greater percentage were not married (84% versus 77.3%). The biggest differences between the two groups were on psychiatric variables. Most incompetent defendants had been diagnosed with a Psychotic Disorder (66.5%) and had a previous psychiatric hospitalization (53.4%), while very few were diagnosed with a Personality Disorder (8.2%). This breakdown is very different than the competent group, whereby only 22.2% were diagnosed with a Psychotic Disorder, 32.3% had a previous psychiatric hospitalization, and 27.9% were diagnosed with a Personality Disorder. While there were a number of distinct differences between the two groups, incompetent and competent defendants were characteristically similar across some variables; specifically, the vast majority of both samples were male (84.1% incompetent, 81.9% competent); had a prior arrest history (59.6% incompetent, 63.4% competent); approximately half had a current violent criminal charge (50.8% incompetent, 55.1% competent); and both groups averaged approximately 10 years of education (10.4 years for the incompetent group, 10.5 years for the competent group).

Table 4. *Incompetent and Competent Sub-Sample Characteristics across Studies (n=68)*

	Incompetent Defendants (n=6,428)			**Competent Defendants** (n=19,711)		
Characteristic	**No. of Studies**	**Mean**	**Range**	**No. of Studies**	**Mean**	**Range**
Demographics						
Age	22	35.0	29-40	22	31.8	27-38
% Male	41	84.1	0-100 (45.0-100)[1]	41	81.9	0-100 (53-100)[2]
% Caucasian	23	47.7	0-83	22	56.9	12-91
% Not Married	10	84.0	56-100	10	77.3	52-86
% Unemployed	8	70.8	29-96	8	58.2	19-80
Education Level (in years)	14	10.4	7-12	14	10.5	8-12
Diagnosis						
% Psychotic Disorder	25	66.5	30-100	25	22.2	0-64
% Mental Retardation	16	7.5	0-25	16	5.2	0-27
% Mood Disorder	15	13.4	0-45	15	13.4	0-38
% Personality Disorder	16	8.2	0-31	17	27.9	0-73
% Substance Use Disorder	16	13.0	0-60	17	22.0	0-84
Psychiatric History						
% Previous Psych. Hospitalization(s)	5	53.4	18-79	6	32.3	0-50
Competency History						
% with Previous Competency Eval(s)	4	23.6	9-40	4	25.9	0-80 (0-14)[3]
Criminal History						
% with Prior Arrest(s)	4	59.6	18-78	4	63.4	40-78
% Current Violent Crime	18	50.8	25-83	18	55.1	20-75

[1] When female-only samples were removed (n=3)
[2] When female-only samples were removed (n=3) and one study with only 1% of competent males
[3] When one study with 80% was removed

Analysis of Categorical Data

The first two hypotheses set forth in this research were: demographic, psychiatric, and psycholegal variables were expected to relate to competency status; therefore, the relationships between competency status and eight categorical variables were investigated: (i) ethnicity; (ii) sex; (iii) employment status; (iv) marital status; (v) psychiatric diagnosis; (vi) psychiatric hospitalization history; (vii) competency evaluation history; (viii) current criminal charge. Each variable was dichotomized and analyzed in relation to competency status, and the following statistics were calculated: odds ratios and their associated 95% confidence intervals, *z*- and *p*-values, study weights, the cumulative random effects odds ratio (i.e., combined effect size of included studies on a particular variable), the median odds ratio, *fail-safe N* (a publication bias statistic), and *Q* (a homogeneity statistic). Associated forest plots were also included.

Moderator Analysis. Three steps were taken to formally test for moderation: subgroup analyses, meta *F*-tests, and meta-regression analyses. Subgroup analyses consisted of calculating odds ratios for each level of six study-level variables hypothesized to serve as potential moderators: (i) type of publication; (ii) source of competency decision; (iii) type of competent group; (iv) setting; (v) country; and, (vi) recruitment. The meta *F*-test represents an analog to ANOVA, whereby each moderator is formally tested for statistically significant differences between its levels. For example, a meta *F*-test analysis of *type of publication* within the *marital status* variable would entail a calculation of the odds ratios elicited from journal articles, dissertations, and books for which Not Married (yes/no) was coded to determine if the effect sizes for the marital status variable statistically differ across types of publication. Bonferroni-type corrections were employed to account for the potential of inflated Type I error. Meta-regression analyses are conceptually equivalent to multiple regression analyses insofar as

predictive models are tested; however, in meta-regression analyses, the effect size serves as the outcome variable and the moderators being explored serve as the antecedent variables (i.e., predictors).

Ethnicity. Table 5 presents the random-effects odds ratios and 95% confidence intervals for studies with ethnicity data (n=22). The cumulative odds ratio was 1.39 (95% CI: 1.08-1.77, median=1.38), such that non Caucasian defendants were approximately one and a half times more likely to be found incompetent than Caucasian defendants, and the *fail-safe N* is 133 (i.e., there would need to be 133 unpublished studies reporting null results to reduce the odds ratio to non-significance). Homogeneity analysis was conducted to determine if the variability across odds ratios is larger than expected from sampling error alone. Signinficant heterogenity was found: $Q(21)= 95.1$, $p<.01$; therefore, subgroup analysis was conducted to explore such variability across potential moderators.

Table 5. *Odds Ratios and Confidence Intervals for Studies with Ethnicity Data (n=22)*

Study name	Odds ratio	Lower limit	Upper limit	Z-Value	p-Value	Relative weight
Brown et al. (1994)	0.648	0.293	1.432	-1.072	0.284	4.06
Cooke et al. (1974)	2.277	1.444	3.590	3.543	0.000	5.71
Cooper & Zapf (2003)	1.714	1.048	2.805	2.146	0.032	5.52
Gothard (1993)	0.163	0.049	0.543	-2.956	0.003	2.61
Grandjean (2004)	2.750	0.805	9.391	1.614	0.106	2.54
Haney & Michielutte (1968)	1.187	0.732	1.923	0.695	0.487	5.57
Heller et al. (1981)	1.042	0.620	1.752	0.157	0.875	5.39
Hoge et al. (1997a)	1.380	0.890	2.139	1.440	0.150	5.79
Hoge et al. (1996)	1.618	0.682	3.838	1.091	0.275	3.76
Jackson et al. (2005)	3.517	1.630	7.589	3.204	0.001	4.17
Lesser (1990)	1.151	0.574	2.307	0.397	0.691	4.51
Otto et al. (1998)	3.587	2.507	5.133	6.988	0.000	6.18
Riley (1997)	1.372	0.883	2.132	1.407	0.160	5.78
Robertson et al. (1997)	0.702	0.279	1.767	-0.751	0.453	3.53
Roesch (1978b)	1.366	0.843	2.216	1.265	0.206	5.57
Rogers et al. (1988)	3.179	1.562	6.470	3.190	0.001	4.44
Rosenfeld & Ritchie (1998)	0.424	0.195	0.922	-2.165	0.030	4.13
Shields (2005)	1.406	0.639	3.094	0.846	0.398	4.08
Sikorski & Benedek (1977)	1.077	0.412	2.814	0.152	0.879	3.39
Warren et al. (2006)	1.027	0.919	1.146	0.467	0.640	7.02
Wildman et al. (1990)	2.867	0.936	8.775	1.845	0.065	2.85
Zapf & Roesch (1998)	1.528	0.587	3.977	0.868	0.385	3.40
Overall	1.385	1.085	1.767	2.615	0.009	

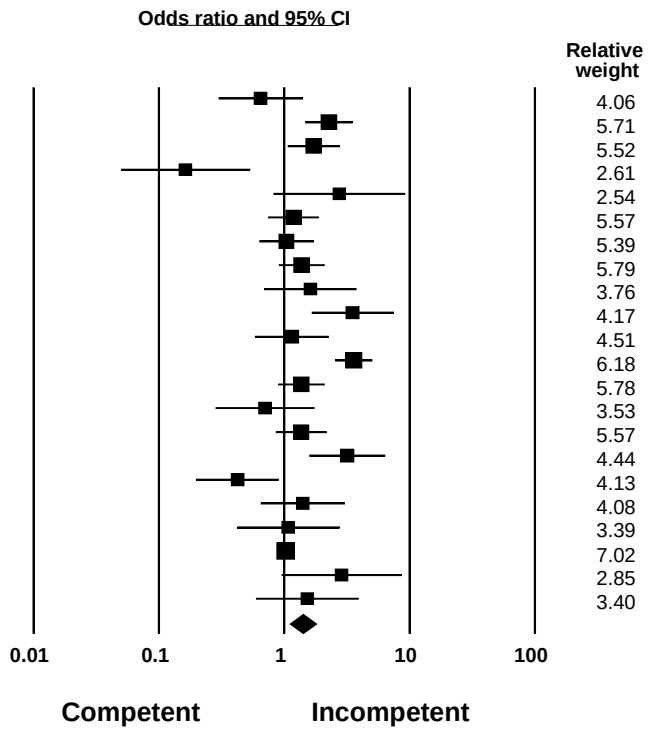


Table 6 presents the odds ratios and 95% confidence intervals for the likelihood of being found incompetent if not Caucasian across six potential moderator variables. *F*-tests were conducted to formally determine if the odds ratios statistically differ across the levels of each moderator. Odds ratios for two of the six moderator groups (i.e., type of competent group and recruitment) significantly differed at a .01 alpha level. Specifically, the effect size for the *pure* competent comparison group (OR=2.33) was significantly larger than both the *referred* (OR=1.20) and *restored* groups (OR=1.53); and, the odds ratio for studies employing a *prospective* recruitment method (OR=1.77) was significantly larger than those utilizing an *archival/retrospective* method (OR=1.17). The meta-regression model including the six moderators as predictors and the cumulative odds ratio as the outcome was non-significant.

Table 6. *Subgroup Analysis of the Likelihood of being found Incompetent if Not Caucasian*

Subgroup	No. of Studies	Odds Ratio (95% CI)
Type of Publication		
Article	17	1.47 (1.11-1.94)
Dissertation	5	1.07 (0.56-2.02)
Source of Competency Decision		
Court	5	1.91 (1.16-3.15)
MHP team	5	1.30 (0.94-1.80)
Psychiatrist(s)	6	0.87 (0.58-1.32)
Psychologist(s)	4	1.71 (0.78-3.73)
Mixed	1	1.41 (0.64-3.10)
Not reported	1	2.28 (1.44-3.60)
Type of Competent Group		
Referred	16	1.20 (0.94-1.53)
Pure	4	2.33 (1.31-4.12)*
Restored	2	1.53 (0.69-3.40)
Setting		
Inpatient	15	1.44 (1.10-1.97)
Outpatient	3	0.96 (0.47-1.96)
Mixed	3	1.70 (0.78-3.77)
Other	1	1.18 (0.73-1.92)
Country		
USA	19	1.36 (1.05-1.76)
Canada	3	1.56 (0.64-3.84)
Recruitment		
Archival/Retrospective	12	1.17 (0.94-1.45)
Prospective	10	1.77 (1.15-2.72)

Sex. Table 7 presents the odds ratios and confidence intervals for studies with available sex data (n=18). The cumulative odds ratio was 1.12 (95% CI: 0.86-1.50, median=1.15), such that female defendants were essentially equally as likely as male defendants to be found incompetent (*fail-safe N* = 0). Significant heterogenity was found: $Q(17)= 48.5$, $p<.01$ (see Table 8 for the subgroup analysis).

Odds ratios for three of the six moderator groups significantly differed: type of publication, country, and recruitment. The effect size for the relationship between sex and competency status presented in the book by Roesch and Golding (1980) was significantly lower (OR=0.001) than those found in the journal articles (OR=1.19) and dissertations (0.98). This finding is not particularly compelling, however, because Roesch and Golding's research only included two female participants – both of whom were deemed competent. A more salient finding was the difference in effect sizes between studies conducted in the United States and Canada. Female defendants were twice as likely (OR=2.03) to be found incompetent than males in the four studies conducted in Canada for which sex data was available (i.e., Crocker et al., 2002; Robertson et al., 1997; Roesch et al., 1981; Rogers et al., 1998) as compared to the 14 studies conducted in the United States, whereby the finding was neutral (OR=1.10). Last, the cumulative odds ratio for the 14 studies employing an archival/retrospective sample recruitment method was significantly larger than the four studies utilizing prospective sampling; however, both effect sizes were relatively neutral (ORs=1.20 and 0.77). The predictive meta-regression model was analyzed and found to be non-significant.

Table 7. *Odds Ratios and Confidence Intervals for Studies with Sex Data (n=18)*

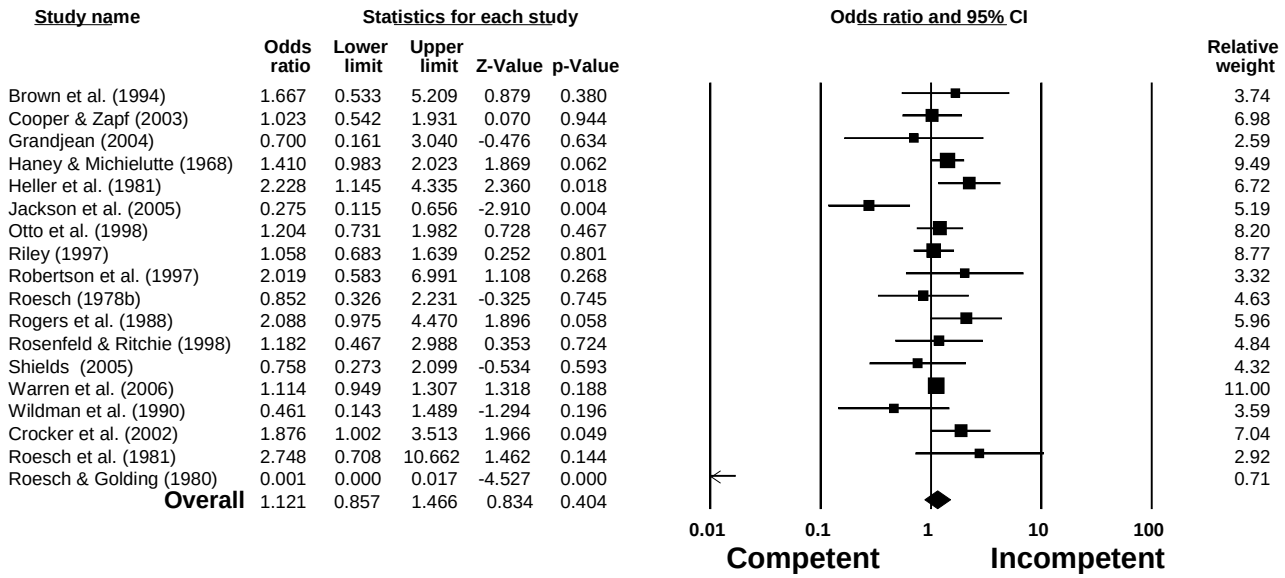


Table 8. *Subgroup Analysis of the Likelihood of being found Incompetent if Female*

Subgroup	No. of Studies	Odds Ratio (95% CI)
Type of Publication		
Article	14	1.19 (1.05-1.35)
Dissertation	3	0.98 (0.66-1.44)
Book	1	0.001 (0.00-0.02)
Source of Competency Decision		
Court	4	1.11 (0.88-1.41)
MHP team	4	0.76 (0.40-1.50)
Psychiatrist(s)	5	1.67 (1.12-2.50)
Psychologist(s)	1	0.76 (0.27-2.10)
Mixed	4	1.15 (0.99-1.33)
Type of Competent Group		
Referred	15	1.20 (1.05-1.35)
Pure	2	0.84 (0.54-1.30)
Restored	1	0.70 (0.16-3.04)
Setting		
Inpatient	11	1.06 (0.78-1.44)
Outpatient	3	1.42 (0.94-2.14)
Mixed	3	1.12 (0.97-1.29)
Other	1	1.41 (0.98-2.02)
Country		
USA	14	1.10 (0.98-1.25)
Canada	4	2.03 (1.33-3.12)*
Recruitment		
Archival/Retrospective	14	1.20 (1.07-1.40)
Prospective	4	0.77 (0.52-1.14)

Employment. Eight studies (n=8) included information relevant to employment (i.e., frequencies of unemployed competent and incompetent defendants) and their associated odds ratios and confidence intervals are presented in Table 9. The cumulative odds ratio is 2.07 (95% CI: 1.38-3.10, median=1.77); therefore, unemployed defendants were twice as likely to be found incompetent as are employed defendants (*fail-safe N* = 54). Signinficant heterogenity was found: $Q(7)= 15.8$, p=.027, and the subgroup analysis is presented in Table 10. Meta *F*-tests for type of publication and setting were not computable because of of the limited variability within the moderator groups and no significant differences were found for comparisons across the other four moderators. A meta-regression model with only three predictors (i.e., type of competent group, country, recruitment) was investiaged due to the lack of variability within the other moderator groups, and the model was not significant.

Table 9. *Odds Ratios and Confidence Intervals for Studies with Employment Data (n=8)*

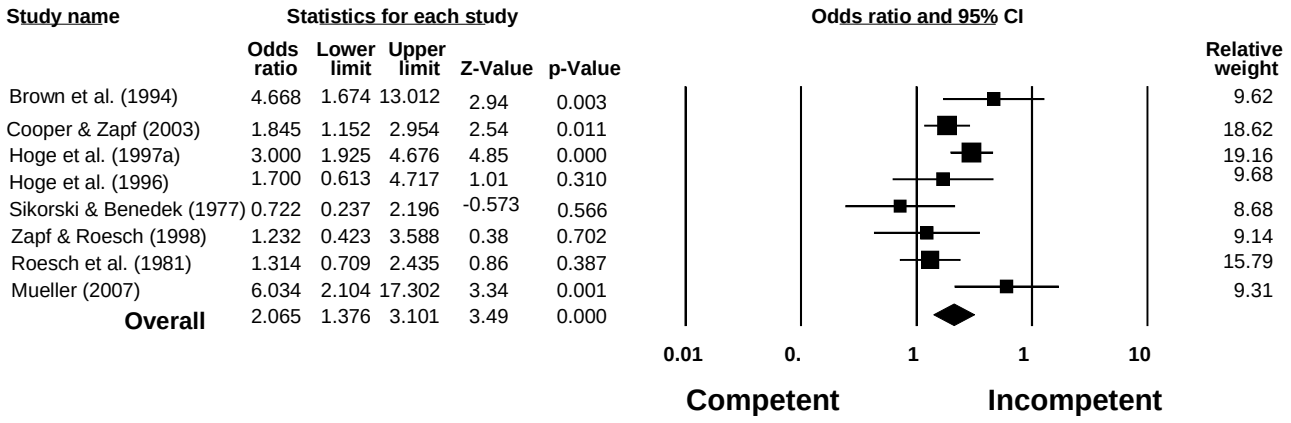


Table 10. *Subgroup Analysis of the Likelihood of being found Incompetent if Unemployed*

Subgroup	No. of Studies	Odds Ratio (95% CI)
Type of Publication		
Article	7	1.90 (1.30-2.76)
Dissertation	1	6.03 (2.10-17.30)
Source of Competency Decision		
Court	1	3.00 (1.93-4.70)
MHP team	1	1.70 (0.61-4.72)
Psychiatrist(s)	3	1.87 (0.85-4.13)
Mixed	3	2.00 (0.76-5.30)
Type of Competent Group		
Referred	6	1.93 (1.13-3.30)
Pure	2	2.74 (1.82-4.12)
Setting		
Inpatient	6	1.84 (1.11-3.10)
Outpatient	1	1.85 (1.20-2.95)
Mixed	1	6.03 (2.10-17.30)
Country		
USA	6	2.42 (1.50-3.90)
Canada	2	1.30 (0.80-2.21)
Recruitment		
Archival/Retrospective	5	2.10 (1.13-3.84)
Prospective	3	2.22 (1.30-3.80)

Marital status. Effect sizes and confidence intervals calculated from studies in which marital status data was available (n=10) are presented in Table 11. The cumulative odds ratio was 1.43 (95% CI: 1.09-1.89, median=1.65), such that defendants who are not married were approximately one and a half times more likely than married defendants to be found incompetent (*fail-safe N* = 15). Signinficant heterogenity was not found: $Q(9)= 6.43$, $p > .05$; however, findings from the subgroup analysis are presented in Table 12 solely for descriptive purposes. Neither meta *F*-tests nor meta-regression analyses were conducted because of the non-significant heterogeneity statistic.

Table 11. *Odds Ratios and Confidence Intervals for Studies with Marital Status Data (n=10)*

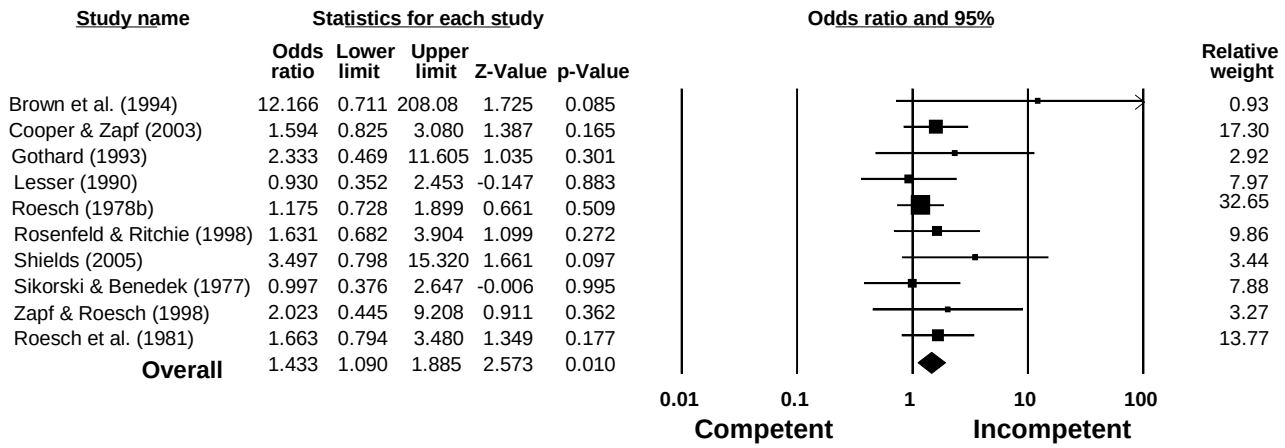


Table 12. *Subgroup Analysis of the Likelihood of being found Incompetent if Not Married*

Subgroup	No. of Studies	Odds Ratio (95% CI)
Type of Publication		
Article	7	1.42 (1.05-1.90)
Dissertation	3	1.63 (0.71-3.75)
Source of Competency Decision		
Court	0	
MHP team	2	1.12 (0.73-1.73)
Psychiatrist(s)	3	1.91 (1.00-3.65)
Mixed	4	1.49 (0.96-2.34)
Psychologist(s)	1	3.49 (0.79-15.32)
Type of Competent Group		
Referred	9	1.49 (1.12-1.98)
Restored	1	0.93 (0.35-2.45)
Pure	0	
Setting		
Inpatient	8	1.37 (0.99-1.89)
Outpatient	2	1.61 (0.95-2.72)
Mixed	0	
Country		
USA	8	1.38 (1.02-1.86)
Canada	2	1.73 (0.89-3.35)
Recruitment		
Archival/Retrospective	7	1.45 (1.08-1.95)
Prospective	3	1.35 (0.65-2.79)

Psychiatric Diagnosis. Table 13 presents the odds ratios and confidence intervals for studies with diagnostic data (n=25). The cumulative odds ratio was 7.96 (95% CI: 5.99-10.60, median=9.28), such that defendants diagnosed with a Psychotic Disorder were nearly eight times more likely to be found incompetent than those without such a diagnosis. These odds are considerably larger than any of those found throughout the present study. There would need to be 5,901 unpublished studies reporting null results to reduce the odds ratio to non-significance (i.e., *fail-safe N*). Significant heterogenity was found: $Q(24)= 119.1$, $p<.01$; therefore, a subgroup analysis was conducted (see Table 14).

.Table 13. *Odds Ratios and Confidence Intervals for Studies with Diagnostic Data (n=25)*

Study name	Odds ratio	Lower limit	Upper limit	Z-Value	p-Value	Relative weight
Brown et al. (1994)	28.679	10.249	80.246	6.393	0.000	3.64
Cooper & Zapf (2003)	4.914	2.810	8.592	5.584	0.000	5.45
Gothard (1993)	34.450	6.717	176.690	4.243	0.000	2.11
Heller et al. (1981)	6.586	4.045	10.724	7.578	0.000	5.74
Hoge et al. (1997a)	11.289	6.780	18.797	9.318	0.000	5.65
Hoge et al. (1996)	207.400	11.822	3638.623	3.650	0.000	0.87
Lesser (1990)	1.712	0.815	3.596	1.419	0.156	4.70
Robertson et al. (1997)	13.800	4.573	41.643	4.658	0.000	3.39
Roesch (1978b)	21.206	10.742	41.865	8.802	0.000	4.96
Rogers et al. (1988)	3.747	1.950	7.199	3.964	0.000	5.07
Shields (2005)	8.938	3.965	20.148	5.281	0.000	4.42
Sikorski & Benedek (1977)	12.857	4.053	40.789	4.336	0.000	3.24
Warren et al. (2006)	3.931	3.507	4.407	23.499	0.000	6.82
Zapf & Roesch (1998)	3.900	1.248	12.184	2.342	0.019	3.29
Roesch et al. (1981)	3.850	2.171	6.827	4.613	0.000	5.40
Mueller (2007)	5.393	3.030	9.597	5.729	0.000	5.38
Cochrane et al. (2001)	9.284	6.871	12.545	14.509	0.000	6.40
Johnson et al. (1990)	5.976	1.185	30.136	2.165	0.030	2.14
Lipsitt et al. (1971)	21.273	2.434	185.913	2.764	0.006	1.39
Nottingham & Mattson (1981)	66.000	4.570	953.231	3.075	0.002	0.98
Pfeiffer et al. (1967)	22.661	7.398	69.407	5.464	0.000	3.35
Poythress et al. (1998)*	12.952	3.919	42.811	4.199	0.000	3.12
Reich & Wells (1985)	6.365	3.834	10.567	7.157	0.000	5.66
Viljoen et al. (2003)	15.393	0.884	268.102	1.875	0.061	0.87
Warren et al. (1991)	6.171	4.006	9.508	8.253	0.000	5.95
Overall	7.963	5.986	10.594	14.245	0.000	

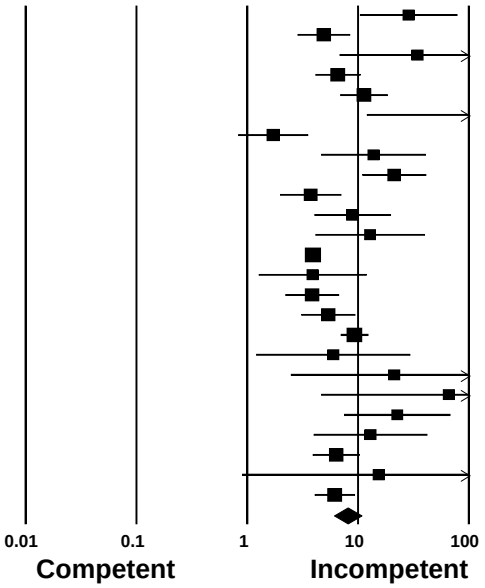


Table 14. *Subgroup Analysis of the Likelihood of being found Incompetent if Psychotic Diagnosis*

Subgroup	No. of Studies	Odds Ratio (95% CI)
Type of Publication		
Article	21	8.52 (6.24-11.64)
Dissertation	4	6.13 (2.37-15.87)
Source of Competency Decision		
Court	5	9.03 (6.44-12.65)
MHP team	6	13.53 (3.76-48.71)
Psychiatrist(s)	5	7.70 (3.34-17.55)
Psychologist(s)	1	8.94 (3.96, 20.15)
Mixed	8	6.60 (4.60-9.52)
Type of Competent Group		
Referred	21	7.94 (5.93-10.64)
Pure	2	33.92 (2.13-539.21)*
Restored	1	1.71 (0.82-3.60)
Mixed	1	12.95 (3.92-42.81)
Setting		
Inpatient	18	9.99 (6.63-15.10)
Outpatient	3	5.99 (4.45-8.10)
Mixed	4	5.30 (3.61-7.66)
Country		
USA	20	8.85 (6.40-12.31)
Canada	5	4.80 (2.99-7.66)
Recruitment		
Archival/Retrospective	16	7.60 (5.60-10.30)
Prospective	9	10.88 (4.71-25.10)

Only levels within the type of competent comparison group significantly differed, studies using *pure* competent comparison groups found defendants much more likely to be found incompetent when diagnosed with a Psychotic Disorder as compared to studies consisting of *referred* (OR=7.94), *mixed* (OR=12.95), and *restored* (OR=1.71) competent defendants. This finding must be interpreted with caution because only two studies used *pure* competent comparison groups (i.e., Hoge et al., 1996; 1997a); in addition, the odds ratio (33.92) is somewhat misleading when aggregated. The odds ratio calculated from the Hoge et al. (1997a) study was 11.3, as approximately 65% of incompetent defendants were diagnosed with a Psychotic Disorder (i.e., 103 of 159) and 14% of competent defendants had such a diagnosis

(i.e., 29 of 207). The odds ratio generated from Hoge et al. (1996) is 207.4 because 30 of the 42 incompetent defendants were diagnosed with a Psychotic Disorder, while none of the competent defendants (n=42) had been given that diagnosis. A meta-regression analysis resulted in a non-significant model ($p = .07$).

Psychiatric hospitalization history. Effect sizes and confidence intervals calculated from studies in which psychiatric hospitalization history was available (n=5) are presented in Table 15 (note: Poythress et al., 1998 is marked with an asterisk because this study utilized an all-female sample). The cumulative odds ratio was 1.86 (95% CI: 1.09-3.20, median=1.58), such that defendants who had a previous psychiatric hospitalization were nearly twice as likely as defendants without such history to be found incompetent (*fail-safe N* = 48). Significant heterogeneity was not found: $Q(4)= 15.5$, $p < .01$; however, neither meta *F*-tests nor meta-regression analyses could be conducted. As noted, only five studies made psychiatric hospitalization history data available, and as such, virtually no variability across moderator groups existed (see Table 16 for the subgroup analysis, which is presented for descriptive purposes).

Table 15. *Odds Ratios and Confidence Intervals for Studies with Psychiatric History Data (n=5)*

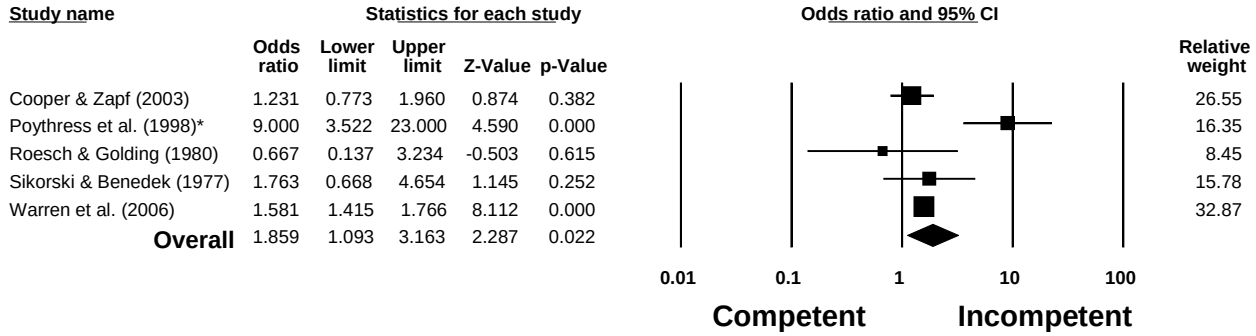


Table 16. *Subgroup Analysis of the Likelihood of being found Incompetent if Previous Psychiatric Hospitalization*

Subgroup	No. of Studies	Odds Ratio (95% CI)
Type of Publication		
Article	4	2.05 (1.20-3.61)
Book	1	0.67 (0.14-3.23)
Source of Competency Decision		
Court	1	9.00 (3.52-23.00)
MHP team	1	0.67 (0.14-3.23)
Mixed	3	1.60 (1.40-1.74)
Type of Competent Group		
Referred	4	1.60 (1.40-1.73)
Mixed	1	9.00 (3.52-23.00)
Setting		
Inpatient	2	1.33 (0.60-3.20)
Outpatient	1	1.23 (0.77-1.96)
Mixed	2	3.54 (0.65-19.34)
Country		
**n/a (All USA)		
Recruitment		
Archival/Retrospective	4	1.56 (1.40-1.73)
Prospective	1	9.00 (3.52-23.00)

Competency evaluation history. Only three studies (n=3) presented data on defendants' competency evaluation history (see Table 17 for effect size data). For interpretive purposes, competency rather than incompetency was used as the criterion in this analysis due to the nature of the data. Regardless, the cumulative odds ratio was essentially neutral (OR=1.07; 95% CI: 0.10-11.20; median=2.47); that is, defendants who had a prior competency evaluation evaluation were no more likely to be found competent (or incompetent) than those who did not have such an evaluation (*fail-safe N* = 0). Although heterogeneity was found: $Q(2)= 90.04$, $p < .01$, neither *F*-tests nor meta-regression analyses were conducted due to the invariability within moderator groups (see Table 18 for the subgroup analysis presented for descriptive purposes).

Table 17. *Odds Ratios and Confidence Intervals for Studies with Competency Evaluation History Data (n=3)*

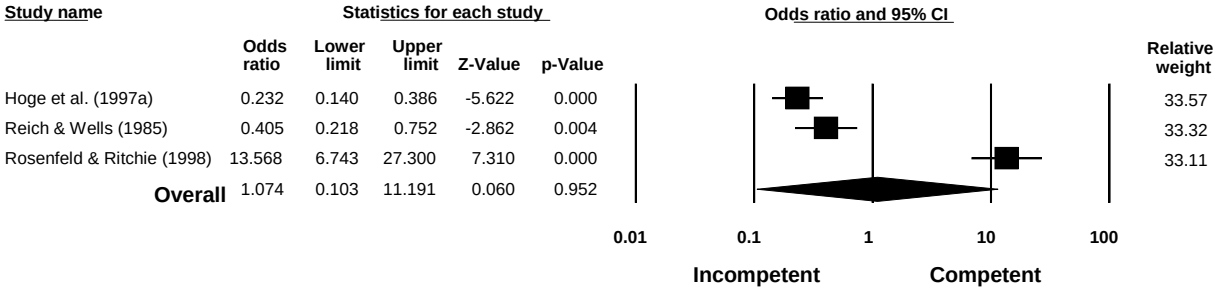


Table 18. *Subgroup Analysis of the Likelihood of being found Competent if Previous Competency Evaluation*

Subgroup	No. of Studies	Odds Ratio (95% CI)
Type of Publication	*n/a (ALL Articles)	
Source of Competency Decision		
Court	2	0.30 (0.20-0.51)
Mixed	1	13.60 (6.74-27.30)
Type of Competent Group		
Referred	2	2.33 (0.10-73.00)
Mixed	1	0.23 (0.14-0.40)
Setting		
Inpatient	1	0.23 (0.14-73.00)
Outpatient	2	2.33 (0.10-1.96)
Country	**n/a (All USA)	
Recruitment		
Archival/Retrospective	2	2.33 (0.10-73.00)
Prospective	1	0.23 (0.14-73.00)

Current criminal charge. Odds ratios and confidence intervals for studies (n=18) for which the type of current criminal charge (i.e., violent or non-violent) was available are presented in Table 19. Competency rather than incompetency was used as the criterion once again due to the nature of the data. Defendants with a current violent criminal charge were 1.25 times more likely to be found competent than those with a current non-violent charge (OR=1.25; 95% CI: 1.00-1.60; median=0.87; *fail-safe N* = 60). Significant heterogeneity was found: $Q(17)= 74.94$, $p < .01$, and the subgroup statistics are presented in Table 20. Meta *F*-tests for all potential moderator variables were conducted (with the exception of type of competent group), but no significant differences were found within groups. A meta-regression model was tested, but was also found to be non-significant.

Table 19. *Odds Ratios and Confidence Intervals for Studies with Criminological Data (n=18)*

Study name	Odds ratio	Lower limit	Upper limit	Z-Value	p-Value
Bittman & Convit (1993)	4.550	2.904	7.129	6.614	0.000
Bluestone & Mellela (1978)	1.046	0.707	1.546	0.223	0.824
Cochrane et al. (2001)	1.110	0.836	1.473	0.722	0.470
Cooper & Zapf (2003)	2.215	1.383	3.548	3.307	0.001
Gothard (1993)	0.571	0.189	1.729	-0.991	0.322
Heller et al. (1981)	1.368	0.845	2.216	1.274	0.203
Laczko et al. (1970)	0.698	0.444	1.095	-1.566	0.117
Lesser (1990)	1.438	0.711	2.906	1.011	0.312
Riley (1997)	1.038	0.671	1.606	0.167	0.867
Robertson et al. (1997)	2.232	0.891	5.589	1.714	0.087
Roesch (1978b)	0.567	0.342	0.940	-2.200	0.028
Roesch et al. (1981)	2.347	1.303	4.227	2.841	0.004
Rosenfeld & Ritchie (1998)	0.653	0.367	1.162	-1.449	0.147
Sachsenmaier (1991)	1.455	0.893	2.368	1.507	0.132
Sikorski & Benedek (1977)	0.431	0.130	1.428	-1.377	0.169
Warren et al. (1991)	1.682	1.122	2.521	2.518	0.012
Warren et al. (2006)	1.183	1.061	1.319	3.017	0.003
Zapf & Roesch (1998)	0.951	0.325	2.781	-0.091	0.927
Overall	1.251	1.004	1.558	1.992	0.046

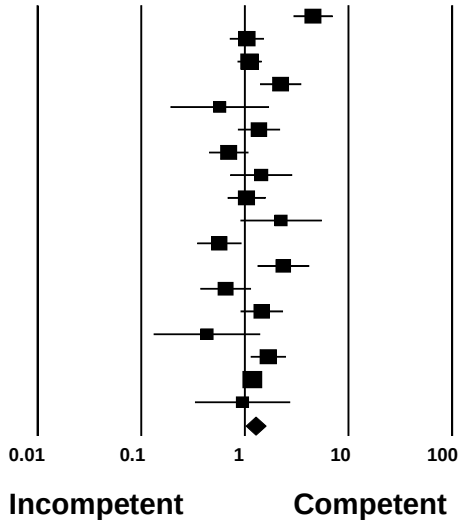


Table 20. *Subgroup Analysis of the Likelihood of being found Competent if Current Violent Charge*

Subgroup	No. of Studies	Odds Ratio (95% CI)
Type of Publication		
Article	14	1.23 (1.13-1.34)
Dissertation	4	1.20 (0.90-1.60)
Source of Competency Decision		
Court	1	1.04 (0.70-1.61)
MHP team	3	0.93 (0.64-1.35)
Psychiatrist(s)	5	1.52 (1.22-1.90)
Mixed	9	1.21 (1.10-1.32)
Type of Competent Group		
Referred	17	1.22 (1.13-1.33)
Restored	1	1.44 (0.71-2.91)
Setting		
Inpatient	11	1.10 (0.90-1.24)
Outpatient	4	1.95 (1.52-2.50)
Mixed	3	1.20 (1.10-1.33)
Country		
USA	15	1.21 (1.11-1.31)
Canada	3	1.98 (1.30-3.10)
Recruitment		
Archival/Retrospective	15	1.23 (1.13-1.33)
Prospective	3	1.10 (0.63-1.80

Analysis of Continuous Data

The final hypothesis in this study was that scores on traditional measures (i.e., intellectual and personality assessment instruments) would be correlated with competency decisions, but larger effect sizes were anticipated for the relationship between scores on competency assessment instruments and such decisions. Scores on competency assessment instruments (e.g., GCCT, FIT) and traditional measures (e.g., MMPI, WAIS) were coded as continuous outcome data in the present meta-analysis. The previous meta-analysis conducted by Nicholson and Kugler (1991) presented effect size data from nine sources of traditional and competency assessment measures: intelligence tests (8 studies); MMPI scales F, 5, 6, and 8 (5 studies); the Competency Screening Test (11 studies); the Georgia Court Competency Test (4 studies); the Competency Assessment Instrument (2 studies); and the Interdisciplinary Fitness Interview (1

study). The effect size data generated in the present study differs from those calculated by the previous meta-analysts in two main respects. First, Nicholson and Kugler (1991) utilized Pearson correlation coefficients to represent effect sizes rather than standardized and unstandardized mean difference statistics as used in the present meta-analysis. Although the calculation of correlation coefficients is not incorrect per se, they may have been used inappropriately if Nicholson and Kugler did not use comparative data to generate such statistics. The second distinction will illustrate this point, as it relates to the effect size data itself. The main inclusion criteria for both the present and previous meta-analysis are identical; namely, included studies must have compared competent and incompetent groups; however, initial inclusion of a study does not equal inclusion of all of its presented data. The validation study of the Competeny Screening Test (CST) conducted by Lipsitt and colleagues (1971) illustrates this point.

Lipsitt and colleagues incorporated six samples in their study: (i) defendants referred to Bridgewater State Hospital for a competency evaluation (n=43); (ii) defendants for whom no question of competency was posed (n=11); (iii) patients civilly committed to Bridgewater State Hospital (n=47); (iv) patients civilly committed to Boston State Hospital (n=19); (v) college undergraduates (n=13); and, (vi) a men's breakfast club associated with a church (n=28). The authors presented the CST means and standard deviations for each of the six aforementioned groups; however, note that none of these groups represent an incompetent group, and therefore, appropriate effect sizes cannot be calculated. Lipsitt and colleagues subsequently divided the group of 43 referred defendants (i.e., Bridgewater experimental group) into two groups based on their CST scores: Low (n=23) and High (n=20). The CST ranges from 0 to 44, such that higher scores relate to competency and lower scores are associated with incompetency. Lipsitt and colleagues classified a score of 21 or higher as "High"; however, this cutoff score was not

derived empirically, but rather because the "research staff found that a qualitative difference in responses appeared at about a score of 20" (p. 106). Furthermore, only sample sizes (not scores) were presented for the bifurcated referred group. To summarize, the validation study of the CST conducted by Lipsitt and colleagues provides no comparative CST data. The CST means and standard deviations initially presented are those of five purely competent groups and one group referred for competency evaluations, which are inappropriate comparisons for this study; and, although the referred defendants were subsequently divided into competent and incompetent groups, CST data was not reported.

Unfortunately, Nicholson and Kugler (1991) only indicated the number of studies included in each effect size analysis and not *which* studies were used. It is impossible to know exactly which studies they utilized to calculate effect sizes, and, therefore, a direct comparison of the present and previous meta-analysis is also impossible. That said, the present meta-analysis includes the same studies included by Nicholson and Kugler (plus those conducted since); thus, the discrepancy of the calculated effect sizes between the two meta-analyses relates to coding decisions and not the studies themselves.[8] The following two sections present descriptive information and effect size data generated from studies using competency assessment instruments and those using traditional measures.

Competency assessment instruments. Twelve competency assessment instruments have been developed over the past 40 years and range from informal checklists to structured, criterion–based scoring instruments[9], including: the Competency Screening Test (CST; Lipsitt et al., 1971), the Competency to Stand Trial Assessment Instrument (CAI; Laboratory of Community Psychiatry, 1973), the Georgia Court Competency Test (GCCT/GCCT-MSH; Nicholson et al., 1988), the Interdisciplinary Fitness Interview (IFI/IFI-R; Golding, 1993), the

Fitness Interview Test (FIT/FIT-R; Roesch, Zapf, Eaves, & Webster, 1998), the Computer-Assisted Determination of Competency to Proceed (CADCOMP; Barnard et al., 1991), the Competence Assessment for Standing Trial for Defendants with Mental Retardation (CAST-MR; Everington & Luckasson, 1992), the Metropolitan Toronto Forensic Service (METFORS) Fitness Questionnaire (MFQ; Nussbaum, Mamak, Tremblay, Wright, & Callaghan, 1998), the MacArthur Competence Assessment Tool–Criminal Adjudication (MacCAT-CA; Poythress et al., 1999), the Mosley Forensic Competency Scale (MFCS; Mosley, Thyer, & Larrison, 2001), the Evaluation for Competency to Stand Trial–Revised (ECST-R; Rogers, Tillbrook, & Sewell, 2004), and the Test of Malingered Incompetence (TOMI; Colwell, Colwell, Perry, Wasieleski, Billings, 2008). For a full description of each of the aforementioned instruments, including instrument development, administration, scoring, and psychometric properties, readers are referred to other sources (Cooper & Grisso, 1997; Goldstein, 2003; Grisso, 1986, 1992, 2003; Melton et al., 2007; Mumley, Tillbrook, & Grisso, 2003; Pirelli, 2008; Roesch, Zapf, Golding, & Skeem, 1999; Zapf & Viljoen, 2003).

Table 21 presents the research studies conducted on each competency instrument, including their respective total sample sizes and the inclusion/exclusion and coding status in the present meta-analysis. (Note: dashes (-) are used to symbolize missing information related to specific sample sizes, and the plus/minus symbol (±) is used to acknowledge approximate total sample sizes). As mentioned earlier, a number of studies incorporated competency assessment measures into their designs; however, most of the published reports did not present data from which effect sizes were calculable.

Table 21. Inclusion Status and Sample Sizes for Studies Using Competency Instruments (n=59)

Instrument	Study	Included (Yes/No)	Total n	Incompetent n	Competent n
Competency Screening Test (CST)	Lipsitt et al. (1971)*	**Y**	43	19	24
	Shatin (1979)*; Shatin & Brodsky (1979)	N/**Y**	21	9	12
	Roesch & Golding (1980)*	**Y**	128	5	123
	Nottingham & Mattson (1981)*	**Y**	50	4	46
	Randolph et al. (1981)*	**Y**	25	15	10
	Randolph et al. (1982)*	**Y**	39	10	29
	Nicholson (1988); Nicholson, Briggs, & Robertson (1988)*; Nicholson, Robertson, et al. (1988)*	**Y**	132	11	121
	Bagby et al. (1992)	**Y**	311	121	190
	Chellsen (1986)*	N	25	25	0
	Paramesh (1987)	N	260	-	-
	Schreiber et al. (1987)*	N	120	-	-
	Roach (1994)	N	72	30	42
	Smith & Hudson (1995); Smith (1996)	N	55	-	-
	Ustad et al. (1996)	N	111	111	0
	TOTAL	15 = 8**Y**/7N	**1,392**	**358±**	**597±**
Competency to Stand Trial Assessment Instrument (CAI)	Roesch (1978); Roesch & Golding (1980)	N	30	4	26
	Schreiber et al. (1987)	N	120	-	-
	Siegel & Elwork (1990)	N	41	41	0
	Robbins, Waters, & Herbert (1997)	N	60	17	43
	Bertsch, Younglove, & Kerr (2002)	N	20	10	10
	Bennett (2006)	N	60	60	0
	TOTAL	6 = 0**Y**/6N	**331**	**132±**	**79±**
Georgia Court CompetencyTest (GCCT/ GCCT-MSH)	Nicholson, Briggs, & Robertson (1988); Nicholson, Robertson, et al. (1988); Nicholson & Johnson (1991)	**Y**	132	11	121
	Johnson et al. (1990)*; Nicholson & Johnson (1991)	**Y**	120	9	111
	Wildman et al. (1990)*	**Y**	100	52	48
	Bagby et al. (1992)	**Y**	311	121	190
	Gothard (1993); Gothard, Rogers, and Sewell (1995); Gothard, Viglione, et al. (1995)	**Y**	108	23	85
	Rogers et al. (1996)	N	125	20	105
	Roach (1994)	N	72	30	42
	Ustad et al. (1996)	N	111	111	0
	Bertman (2000); Bertman et al. (2003)	N	26	26	0
	Manguno-Mire et al. (2007)	N	21	12	0
	TOTAL	10 = 5**Y**/5N	**1,126**	**415**	**702**
Interdisciplinary Fitness Interview-Revised (IFI-R)	Golding et al. (1984)*	**Y**	75	17	58
	Barnard et al. (1991); Barnard et al. (1992)	N	50/99	50/99	0
	Holmes (1991)	N	-	-	-
	TOTAL	3 = 1**Y**/2N	**174±**	**116±**	**58±**

Fitness Interview Test (FIT/ FIT-R)	Bagby et al. (1992)	**Y**	311	121	190
	McDonald et al. (1991)	**Y**	243	99	144
	Viljoen et al. (2003)	**Y**	96	13	83
	Whittemore et al. (1997)	**Y**	236	26	210
	Zapf & Roesch (1998)	**Y**	178	20	158
	Zapf et al. (2001)	**Y**	100	10	90
	Menzies et al. (1983)*	N	270	-	-
	Zapf & Roesch (1997, 1998)	N	57/178	20	158
	Zapf (1999); Zapf & Roesch (2001, 2005)	N	100	-	-
	Viljoen et al. (2002); Viljoen & Zapf (2002)	N	212/160	-	-
	TOTAL	10 = 6**Y**/4N	**1,924**	**309±**	**1,033±**
Computer-Assisted Determination of Competency to Proceed (CADCOMP)	Barnard et al. (1991); Barnard et al. (1992)	N	50/99	50/99	0
	Nicholson et al. (1994)	N	133	133	0
	Roach (1994)	N	72	30	42
	Buigas (1996)	N	74	-	-
	TOTAL	4 = 0**Y**/4N	**378**	**262±**	**42±**
Competence Assessment for Standing Trial for Defendants with Mental Retardation (CAST-MR)	Everington (1989); Everington (1990)	N	93	11	82
	Everington & Dunn (1995)	N	35	20	15
	Peacock (2005)	N	68	9	59
	Bennett (2006)	N	60	60	0
	Everington, Notario-Smull, & Horton (2007)	N	95	0	95
	Stoops et al. (2007)	N	1	1	0
	TOTAL	6 = 0**Y**/6N	**352**	**101**	**251**
Metropolitan Toronto Forensic Service (METFORS) Fitness Questionnaire (MFQ)	Nussbaum et al. (1998)	**Y**	44	15	29
	Nussbaum & Amaral (2001)	N	144	-	-
	TOTAL	2 = 1**Y**/1N	**188**	**15±**	**29±**
MacArthur Competence Assessment Tool - Criminal Adjudication (MacCAT-CA)	Otto et al. (1998) ; Poythress et al. (1999); Zapf et al. (2005)	**Y**	729	283	446
	Tillbrook (2001)	**Y**	70	33	37
	Viljoen et al. (2003)	**Y**	96	13	83
	Zapf (1999); Zapf & Roesch (2001, 2005)	N	100	-	-
	Redlich, Silverman, & Steiner (2003)	N	17	0	17
	Ryba (2005)	N	77	37	0
	Bennett (2006)	N	60	60	0
	Pinals et al. (2006)	N	-	-	-
	TOTAL	8 = 3**Y**/5N	**1,149±**	**426±**	**583±**
Mosley Forensic Competency Scale (MFCS)	Mosley et al. (2001)	**Y**	75	19	56
Evaluation for Competency to Stand Trial-Revised (ECST-R)	Grandjean (2004); Rogers et al. (2003)	**Y**	48	30	18
	Jackson et al. (2005)	**Y**	137	41	96
	Tillbrook (2001); Rogers et al. (2002, 2003)	**Y**	70	33	37
	Rogers et al. (2004)	N	129	42	87
	Gabel (2007); Vitacco et al. (2007)	N	100	-	-
	TOTAL	5 = 3**Y**/2N	**484**	**146±**	**238±**
Test of Malingered Incompetence (TOMI)	Colwell et al. (2008)	N	**392**	**30**	**362**

*used by Nicholson & Kugler (1991)

Although numerous studies have incorporated competency measures into their designs, only eight independent studies have compared scores of competent and incompetent defendants on such measures to the extent that an effect size was calculable. Furthermore, sufficient data was only available for five of these measures (i.e., CST, GCCT-MSH, FIT, MFQ, MFCS), and only the CST and GCCT-MSH have such data from more than one independent study.[10] Descriptive and effect size data are presented in the following tables; however, neither meta-*F* nor meta-regression analyses were performed due to insufficient variability across various levels of the moderators (e.g., type of setting).

Table 22 presents two studies that investigated the CST and presented data sufficient to calculate effect sizes. Competent defendants (M=17.0, SD=8.8) scored approximately 10 points higher than their incompetent counterparts (M=26.7, SD=8.3) on the CST (unstandardized mean difference = 9.8). The small and disproportionate sample sizes in the two studies are noteworthy.

Table 22. Unstandardized Mean Difference Statistics for the CST (n=2)

Study Name	Incompetent Group Mean (SD)	Incompetent Group n	Competent Group Mean (SD)	Competent Group n	Difference in Means† (CIs)	Standard Error (Variance)	Z-value
Randolph et al. (1981)	19.9 (9.26)	15	29.0 (3.8)	10	9.1 (3.0 - 15.2)	3.10 (9.64)	2.93**
Nicholson (1988)	14.1 (8.1)	11	24.4 (8.5)	121	10.3 (1.6 - 4.8)	2.70 (7.12)	3.86***
OVERALL	17.0 (8.8)	26	26.7 (8.3)	131	9.8 (5.8 - 13.8)	2.02 (4.09)	4.84***

**p < .01; ***p < .001

Table 23 presents four studies that investigated the GCCT-MSH and reported sufficient data. Competent defendants (M=81.3, SD=16.9) scored approximately 25 points higher than their incompetent counterparts (M=55.6, SD=25.9) on the GCCT-MSH (unstandardized mean difference = 25.76, median=27.1). The small sample sizes across studies are noteworthy.

Table 23. Unstandardized Mean Difference Statistics for GCCT-MSH (n=4)

Study Name	Incompetent Group Mean (SD)	Incompetent Group n	Competent Group Mean (SD)	Competent Group n	Difference in Means† (CIs)	Standard Error (Variance)	Z-value
Nicholson (1988)	51.3 (24.8)	11	81.6 (18.3)	121	30.3 (18.7 – 41.9)	5.95 (35.35)	5.10***
Johnson et al. (1990)	66.0 (25.9)	9	79.4 (16.5)	111	13.4 (1.7 – 25.2)	5.99 (35.95)	2.24*
Wildman et al. (1990)	44.4 (30.5)	52	79.8 (21.5)	48	35.4 (24.9 – 45.8)	5.32 (28.28)	6.66***
Gothard, Viglione, et al. (1995)	60.5 (15.4)	30	84.5 (7.3)	55	23.9 (19.2 – 28.8)	2.46 (6.06)	9.74***
OVERALL	55.6 (25.9)	102	81.3 (16.9)	335	25.76 (17.9 – 33.6) median = 27.1	3.98 (15.86)	6.47***

*p < .05; ***p < .001

The FIT, MFQ, and MFCS were used in one included study each; therefore, those data were combined with the data on the CST and GCCT-MSH to investigate the difference between competent and incompetent defendants on competency measures in general. Standardized mean difference statistics (i.e., Cohen's *d*) were calculated rather than unstandardized differences to account for the use of different measures employed to operationalize the same construct (e.g., competency to stand trial). Table 24 presents the nine sources from which the effect sizes were generated. Incompetent and competent defendants' scores across competency assessment instruments significantly differed and a rather large effect size was found: standardized effect size *(d)* = 1.4 (1.1-1.7), p < .001, which is equal to an odds ratio of 2.5.[11] The relatively small total sample sizes for both incompetent (n=214) and competent defendants (n=574) are noteworthy in this analysis.

Table 24. Standardized Mean Difference (Cohen's *d*) Statistics for Competency Measures (n=8)

Study Name	Test Used	Incompetent Group Mean (SD)	Incompetent Group n	Competent Group Mean (SD)	Competent Group n	Standardized Difference in Means† (CIs)	Standard Error (Variance)	Z-value	Odds Ratio (OR)
Randolph et al. (1981)	CST	19.9 (9.3)	15	29.0 (3.8)	10	1.2 (0.6 – 1.9)	0.32 (0.11)	3.76***	2.18
Nicholson (1988)	CST	14.1 (8.1)	11	24.4 (8.5)	121	1.2 (0.3 – 2.1)	0.44 (0.19)	2.71**	2.18
Nicholson (1988)	GCCT-MSH	51.3 (24.8)	11	81.6 (18.3)	121	1.6 (0.9 – 2.3)	0.33 (0.11)	4.86***	2.90
Johnson et al. (1990)	GCCT-MSH	66.0 (25.9)	9	79.4 (16.5)	111	0.8 (0.1 – 1.5)	0.35 (0.12)	2.21*	1.45
Wildman et al. (1990)	GCCT-MSH	44.4 (30.5)	52	79.8 (21.5)	48	1.3 (0.9 – 1.8)	0.22 (0.05)	6.02***	2.36
Gothard et al. (1995)	GCCT-MSH	60.5 (15.4)	30	84.5 (7.3)	55	2.2 (1.7 – 2.8)	0.28 (0.08)	7.80***	3.99
McDonald et al. (1991)	FIT	42.9 (26.2)	63	28.0 (6.6)	144	0.9$^{\Delta}$ (0.7 – 1.3)	0.16 (0.03)	6.09***	1.63
Nussbaum et al. (1998)	MFQ	9.8 (3.8)	15	15.9 (3.0)	29	1.9 (1.1 – 2.6)	0.38 (0.14)	4.97***	3.45
Mosley et al. (2001)	MFCS	10.2 (7.4)	19	17.0 (4.1)	56	1.3 (0.8 – 1.9)	0.29 (0.08)	4.64***	2.36
OVERALL		n/a	214$^{\pm}$	n/a	574$^{\pm}$	1.4 (1.1 – 1.7)	0.15 (0.02)	8.99***	2.54

†positive difference denotes higher scores for the *Competent* group

$^{\Delta}$The FIT standardized mean difference score was calculated as a positive integer to maintain overall consistency with the model because it is the only measure presented in the table whereby higher scores are associated with "incompetency".

*p < .05; **p < .01; *** p < .001

± Total sample sizes reflect independent samples only (i.e., the Nicholson (1988) sample was included once in the calculations)

Traditional assessment instruments. Traditional assessment instruments are those developed primarily to measure broad psychological constructs (e.g., intelligence or personality). Although contemporary practice standards encourage the use of competency assessment instruments in evaluations (Grisso, 2003; Melton et al., 2007; Zapf & Roesch, 2009), many psychologists continue to rely heavily on traditional measures in forensic evaluations (Archer, Buffington-Vollum, Stredny, & Handel, 2006; Borum & Grisso, 1995; Nicholson & Norwood, 2000; Ryba, Cooper, & Zapf, 2003; Skeem & Golding, 1998). Table 25 presents three traditional measures most commonly researched in the competency arena: the Minnesota Multiphasic Personality Inventory (MMPI/MMPI-2); the Wechsler Adult Intelligence Scales (WASI, WAIS, WAIS-R, WAIS-III); and, the Brief Psychiatric Rating Scale (BPRS). A number of other traditional measures have been included by researchers over the years; however, most have only been included in single studies (see Table 26). Each of these measures consists of numerous scales and subscales, but most of the studies conducted in this area have only presented data on total scores or scores on a few scales. The available data for each measure is presented in the following tables. Effect sizes were calculated in the form of standardized mean differences, but neither meta-*F* nor meta-regression analyses were conducted due to the insufficient variability across the various levels of moderators.

Table 25. Inclusion Status and Sample Sizes for Studies Using Traditional Instruments (n=46)

Instrument	Study	Included (Yes/No)	Total n	Incompetent n	Competent n
Minnesota Multiphasic Personality Interview (MMPI/ MMPI-2)	Pfeiffer et al. (1967)*	**Y**	89	34	55
	Cooke (1969)*	**Y**	215	93	122
	Maxson & Neuringer (1970)*	**Y**	594	56	538
	Cooke et al. (1974)	**Y**	325	126	199
	Rogers et al. (1988)*	**Y**	459	56	403
	Johnson et al. (1990)*	**Y**	120	9	111
	Lesser (1990)	**Y**	136	52	83
	Wildman et al. (1990)*	**Y**	100	52	48
	Sachsenmaier (1991)	**Y**	445	97	348
	Otto et al. (1998)	**Y**	729	283	446
	Carbonell et al. (1992)	N	152	-	-
	Miller (2004)	N	50	50	0
	Wygant et al. (2007)	N	87	-	-
	TOTAL	13 = 10**Y**/3N	**3,501**	**908±**	**2,353±**
Wechsler Abbreviated Scale of Intelligence (WASI)	Granjean (2004)	**Y**	48	30	18
Wechsler Adult Intelligence Scale (WAIS)	Pfeiffer et al. (1967)*	**Y**	89	34	55
	Cooke (1969)*	**Y**	215	93	122
	Heller et al. (1981, 1983)*	**Y**	410	106	304
	Laczko et al. (1970)*	**Y**	421	104	317
	Shatin & Brodsky (1979)*; Shatin (1979)	**Y**/N	21	9	12
	Smith & Broughton (1994)	N	160	-	-
	Smith & Hudson (1995); Smith (1996)	N	55	-	-
	Vernon et al. (1999)	N	28	-	-
	TOTAL	9 = 5**Y**/4N	**1,399**	**346±**	**810±**
Wechsler Adult Intelligence Scale-Revised (WAIS-R)	Johnson et al. (1990)*	**Y**	120	9	111
	Lesser (1990)	**Y**	135	52	83
	Sachsenmaier (1991)	**Y**	445	97	348
	Hoge et al. (1996)	**Y**	84	42	42
	Hoge et al. (1997a)	**Y**	366	159	207
	Otto et al. (1998)	**Y**	729	283	446
	Poythress et al. (1998)	**Y**	106	38	68
	Nestor et al. (1999)	**Y**	181	53	128
	Pierrel (1986)	N	73	-	-
	Gannon (1990)	N	50	50	0
	Carbonell et al. (1992)	N	152	-	-
	Everington & Dunn (1995)	N	35	20	15
	Jones (1995)	N	271	271	0
	Redding (1997)	N	29	-	-
	Bertman (2000); Bertman et al. (2003)	N	26	26	0
	Nicholson, Briggs, & Robertson (1988); Nicholson, Robertson, et al. (1988); Nicholson & Johnson (1991)	**Y**	132	11	121
	Anderson (1999); Anderson & Hewitt (2002)	N	75	75	0
	Viljoen et al. (2002); Viljoen & Zapf (2002)	N	212/160	-	-
	Everington et al. (2007)	N	95	-	-
	TOTAL	19 = 8**Y**/11N	**3,316**	**1,186±**	**1,569±**

Wechsler Adult Intelligence Scale-3rd Ed. (WAIS-III)	Shields (2005)	**Y**	218	35	183
	Bertsch et al. (2002)	N	20	10	10
	Ryba (2005)	N	77	37	0
	Patterson (2005)	N	617	617	0
	Peacock (2005)	N	68	9	59
	Bennett (2006)	N	60	60	0
	Everington et al. (2007)	N	95	-	-
	TOTAL	7 = 1**Y**/6N	**1,155**	**768±**	**252±**
Brief Psychiatric Rating Scale (BPRS)	Johnson et al. (1990)*	**Y**	120	9	111
	Hoge et al. (1996)	**Y**	84	42	42
	Hoge et al. (1997a)	**Y**	366	159	207
	Otto et al. (1998)	**Y**	729	283	446
	Poythress et al. (1998)	**Y**	106	38	68
	Zapf et al. (2001)	**Y**	100	10	90
	Viljoen et al. (2003)	**Y**	96	13	83
	Quinsey et al. (1975)	N	56	24	0
	Roach (1994)	N	72	30	42
	Redding (1997)	N	29	-	-
	Bertman (2000); Bertman et al. (2003)	N	26	26	0
	Ryba (2005)	N	77	37	0
	TOTAL	12 = 7**Y**/5N	**1,861**	**671±**	**1,089±**

*used by Nicholson & Kugler (1991)

Table 26. Other Traditional Measures Used Across Studies (n=14)

Study	**Measures Used**
Pfeiffer et al. (1967)	Rorschach, Thematic Apperception Test (TAT), Bender Gestalt
Heller et al. (1981, 1983)	Western Personnel Test (IQ)
Simon (1987)	Quick Test, Proverbs Test
Lesser (1990)	Bender Gestalt, Psychopathy Checklist (PCL), Schedule for Affective Disorders and Schizophrenia (SADS), Structured Clinical Interview for DSM-III (SCID), Michigan Alcoholism Screening Test (MAST)
Wildman et al. (1990)	Peabody Picture Vocabulary Test
Hoge et al. (1997a)	MacArthur Structured Assessment of the Competencies of Criminal Defendants (MacSAC-CD), Perceived Criminal Injustice Scale
Whittemore et al. (1997)	Test of Charter Comprehension (ToCC), Structured Clinical Interview for DSM-III –R – Patient Version (SCID-P)
Nussbaum et al. (1998)	Wechsler Memory Scales (WMS), Rey Complex Figure Test, Trail Making Test, Controlled Oral Word Association Test FAS, Common Item Estimation Test
Matthews (1999)	Scale to Assess Unawareness of Mental Disorder (SUMD), Positive and Negative Syndrome Scale (PANNS), Test of Nonverbal Intelligence-2 (TONI-2)
Nestor et al. (1999)	Wecshler Memory Scale-Revised (WMS-R), Trail Making Test, Wisconsin Card Sort Test (WCST), Wide Range Achievement Test-Revised (WRAT-R)
Zapf et al. (2001)	Structured Clinical Interview for DSM-III –R – Patient Version (SCID-P)
Grandjean (2004)	WMS, Trail Making Test, COWAT, Stroop Color and Word Test, Continuous Performance Test, Coglab, WCST, Social Knowledge Questionnaire, Insight scale for Psychosis, Rey Fifteen Item Memory Test
Jackson et al. (2005)	Miller Forensic Assessment of Symptoms Test (MFAST)
Ryba (2005)	Brief Test of Attention, Trail Making Test

The Wechsler instruments have been included in 36 of all competency studies (included and excluded), but very few of them have presented comparative data on competent and incompetent defendants across three main indices of cognitive functioning: Full Scale IQ, Performance IQ, and Verbal IQ. The following results are based on analysis of data from few studies with relatively small sample sizes overall.

Three studies investigated the Full Scale IQ (FSIQ) score differences between competent and incompetent defendants. Nestor, Daggett, Haycock, and Price (1999) and Otto and colleagues (1998) utilized the WAIS-R, while Grandjean (2004) incorporated the WASI. Table 27 presents the descriptive and effect size statistics for these data. Competent defendants (M=86.8, SD=14.0) scored approximately six Full Scale IQ points higher than their incompetent counterparts (M=80.6, SD=14.1). The standardized mean difference (0.32, median=0.42) can be classified as small to medium and is equal to an odds ratio of 0.58.

Four studies presented comparative Performance IQ (PIQ) data from the Wechsler scales; specifically, the WASI (Grandjean, 2004), the WAIS-R (Lesser, 1990; Nestor et al., 1999), and the WAIS-III (Shields, 2005). Competent defendants (M=84.9, SD=14.0) scored approximately five Performance IQ points greater than incompetent defendants (M=79.6, SD=13.4; see Table 28), which is also associated with a small to medium effect size statistic (standardized mean difference = 0.38, median=0.27) and is equal to an odds ratio of 0.69

Seven studies investigated differences between competent and incompetent defendants on Wechsler verbal indices, which included Verbal IQ (VIQ) scores as well as those on a Verbal Cognitive Functioning (VCF) index. The VCF index was calculated by the three included MacArthur studies (i.e., Hoge et al., 1996, 1997a; Poythress et al., 1998) using the Vocabulary, Similarities, and Digit Span subtests of the WAIS-R. Five of the seven studies utilized the

WAIS-R, while the WASI and WAIS-III were included by one study each. Table 29 presents descriptive and effect size statistics on the VIQ and VCF indices. Comparable to the aforementioned FSIQ and PIQ findings, competent defendants (M=87.2, SD=13.5) scored approximately five IQ points higher than incompetent defendants (M=82.1, SD=12.3), which translates into a small to medium effect size (standardized mean difference = 0.37, median=0.36) and is equal to an odds ratio of 0.67.

Table 27. Standardized Mean Difference (Cohen's *d*) Statistics for Wechsler Full Scale IQ (FSIQ) (n=3)

Study Name	Test Used	Incompetent Group Mean (SD)	Incompetent Group n	Competent Group Mean (SD)	Competent Group n	Standardized Difference in Means† (CIs)	Standard Error (Variance)	Z-value	Odds Ratio (OR)
Otto et al. (1998)	WAIS-R	83.2 (14.1)	283	85.4 (13.5)	446	0.16 (0.007 – 0.3)	0.08 (0.01)	2.06*	0.29
Nestor et al. (1999)	WAIS-R	82.0 (14.4)	53	88.3 (15.1)	128	0.42 (0.1 – 0.7)	0.17 (0.03)	2.54**	0.76
Grandjean (2004)	WASI	76.5 (13.6)	30	86.8 (18.4)	18	0.66 (0.1 – 1.3)	0.31 (0.09)	2.22*	1.19
OVERALL		80.6 (14.1)	366	86.8 (14.0)	592	0.32 (0.1 – 0.6) median = 0.42	0.14 (0.02)	2.35*	0.58

Table 28. Standardized Mean Difference (Cohen's *d*) Statistics for Wechsler Performance Scale IQ (PIQ) (n=4)

Study Name	Test Used	Incompetent Group Mean (SD)	Incompetent Group n	Competent Group Mean (SD)	Competent Group n	Standardized Difference in Means† (CIs)	Standard Error (Variance)	Z-value	Odds Ratio (OR)
Lesser (1990)	WAIS-R	77.9 (12.1)	52	80.0 (11.5)	83	0.18 (-.02 – 0.5)	0.18 (0.03)	1.01≠	0.33
Nestor et al. (1999)	WAIS-R	82.8 (14.1)	53	89.0 (15.5)	128	0.41 (0.1 – 0.7)	0.17 (0.03)	2.50**	0.74
Grandjean (2004)	WASI	82.7 (13.4)	30	88.3 (18.8)	18	0.36 (-0.2 – 0.9)	0.30 (0.09)	1.18≠	0.65
Shields (2005)	WAIS-III	75.1 (14.2)	35	82.6 (13.4)	183	0.55 (0.2 – 0.9)	0.19 (0.04)	2.96***	0.99
OVERALL		79.6 (13.4)	170	84.9 (14.0)	412	0.38 (0.2 – 0.6) median = 0.27	0.09 (0.01)	3.91***	0.69

*p < .05; **p < .01; ***p < .001

Table 29. Standardized Mean Difference (Cohen's *d*) Statistics for Wechsler Verbal Indices (VIQ and VCF) (n=7)

Study Name	Test Used	Incompetent Group Mean (SD)	Incompetent Group n	Competent Group Mean (SD)	Competent Group n	Standardized Difference in Means† (CIs)	Standard Error (Variance)	Z-value	Odds Ratio (OR)
Lesser (1990)	WAIS-R	80.5 (10.8)	52	82.0 (11.4)	83	0.13 (-0.2 – 0.5)	0.17 (0.03)	0.76[≠]	0.24
Hoge et al. (1996)	WAIS-R (VCF)	82.4 (10.3)	42	91.4 (15.2)	42	0.69 (0.3 – 1.1)	0.23 (0.05)	3.09**	1.25
Hoge et al. (1997a)	WAIS-R (VCF)	87.9 (11.5)	159	91.4 (12.9)	207	0.28 (0.1 – 0.5)	0.11 (0.01)	2.62**	0.51
Poythress et al. (1998)[±]	WAIS-R (VCF)	88.7 (11.1)	38	85.4 (10.9)	68	-0.31 (-0.7 – 0.09)	0.20 (0.04)	-1.50[≠]	(-) 0.56
Nestor et al. (1999)	WAIS-R	83.1 (15.2)	53	88.5 (15.3)	128	0.36 (0.04 – 0.7)	0.16 (0.03)	2.18*	0.65
Grandjean (2004)	WASI	74.5 (14.4)	30	87.2 (16.6)	18	0.83 (0.2 – 1.4)	0.31 (0.09)	2.67**	1.51
Shields (2005)	WAIS-III	77.8 (14.3)	35	84.2 (13.8)	183	0.46 (0.1 – 0.8)	0.19 (0.04)	2.48**	0.83
OVERALL		82.1 (12.3)	409	87.2 (13.5)	729	0.37 (0.2 – 0.5) median = 0.36	0.07 (0.005)	5.03***	0.67

[±]All-female study

*p < .05; **p < .01; ***p < .001; [≠] ns

The MMPI and/or the MMPI-2 have been included in 13 studies; however, the reported data is limited to validity and clinical scale scores. Furthermore, only two studies (i.e., Maxson & Neuringer, 1970; Sachsenmaier, 1991) have presented comparative data of competent and incompetent defendants on three scales for which effect sizes are calculable. Data on the MMPI scales F, 6, and 8, are presented in tables 30, 31, and 32, respectively. Although it was not explicated in her dissertation, it seems Sachsenmaier (1991) reported scale raw scores; therefore, standardized mean difference statistics were computed. Both Maxson and Neuringer (1970) and Sachsenmaier (1991) found incompetent defendants to produce higher scores across all three scales. The standardized mean difference effect sizes for the MMPI F scale, scale 6, and scale 8 were 0.33, 0.39, and 0.33, which are all considered small to medium. The associated odds ratio statistics are 0.59, 0.71, and 0.59.

Table 30. Standardized Mean Difference Statistics for MMPI F scale (n=2)

Study Name	Incompetent Group Mean (SD)	Incompetent Group n	Competent Group Mean (SD)	Competent Group n	Standardized Difference in Means† (CIs)	Standard Error (Variance)	Z-value	Odds Ratio (OR)
Maxson & Neuringer (1970)	69.1 (11.2)	56	64.7 (11.7)	538	0.38 (0.1-0.7)	0.14 (0.02)	2.68**	0.69
Sachsenmaier (1991)	15.2 (10.4)	97	12.2 (10.2)	348	0.29 (0.1-0.5)	0.12 (0.01)	2.57**	0.53
OVERALL		153		886	0.33 (0.2-0.5)	0.09 (0.01)	3.68***	0.59

Table 31. Standardized Mean Difference Statistics for MMPI scale 6 (n=2)

Study Name	Incompetent Group Mean (SD)	Incompetent Group n	Competent Group Mean (SD)	Competent Group n	Standardized Difference in Means† (CIs)	Standard Error (Variance)	Z-value	Odds Ratio (OR)
Maxson & Neuringer (1970)	73.4 (16.9)	56	67.2 (16.5)	538	0.38 (0.1-0.7)	0.14 (0.02)	2.66**	0.69
Sachsenmaier (1991)	19.7 (11.0)	97	16.1 (8.3)	348	0.40 (0.2-0.6)	0.12 (0.01)	3.49***	0.73
OVERALL		153		886	0.39 (0.2-0.6)	0.09 (0.01)	4.38***	0.71

Table 32. Standardized Mean Difference Statistics for MMPI scale 8 (n=2)

Study Name	Incompetent Group Mean (SD)	Incompetent Group n	Competent Group Mean (SD)	Competent Group n	Standardized Difference in Means† (CIs)	Standard Error (Variance)	Z-value	Odds Ratio (OR)
Maxson & Neuringer (1970)	59.8 (29.9)	56	50.1 (28.6)	538	0.34 (0.1-0.6)	0.14 (0.02)	2.39**	0.62
Sachsenmaier (1991)	40.3 (13.5)	97	36.2 (12.5)	348	0.32 (0.1-0.5)	0.12 (0.01)	2.79**	0.58
OVERALL		153		886	0.33 (0.2-0.5)	0.09 (0.01)	3.68***	0.59

*p < .05; **p < .01; ***p < .001

Twelve studies utilized the BPRS in their design – four of which investigated the total score differences between competent and incompetent defendants and presented data from which an effect size could be generated. Each study found higher BPRS total scores for incompetent defendants compared to competent defendants (see Table 33), which relates to more severe psychiatric symptomatology. The overall discrepancy was 7.4 points (median=4.9), such that incompetent defendants scored around 40 (M=39.6, SD=9.2), whereas competent defendants' mean BPRS total score was 32.0 (SD=7.9).

Table 33. Unstandardized Mean Difference Statistics for BPRS (n=4)

Study Name	Incompetent Group Mean (SD)	Incompetent Group n	Competent Group Mean (SD)	Competent Group n	Difference in Means† (CIs)	Standard Error	Z-value
Hoge et al. (1996)	45.8 (10.1)	42	28.5 (4.1)	42	17.3 (13.9 - 20.6)	1.69 (2.85)	10.26***
Hoge et al. (1997a)	37.0 (7.6)	159	33.8 (7.7)	207	3.2 (1.6 - 4.8)	0.81 (0.65)	3.95***
Otto et al. (1998)	38.8 (10.0)	283	32.6 (8.2)	446	6.3 (4.9 - 7.6)	0.68 (0.47)	9.17***
Poythress et al. (1998)[±]	36.6 (7.4)	38	33.1 (7.7)	68	3.5 (0.5 – 6.5)	1.54 (2.36)	2.29*
OVERALL	39.6 (9.2)	522	32.0 (7.9)	763	7.4 (2.9 - 11.9) median = 4.9	2.32 (5.39)	3.19***

*p < .05; **p < .01; ***p < .001; [≠] ns

[±]All-female study

Discussion

The present study was the second meta-analysis conducted on adjudicative competency research. The first, conducted by Nicholson and Kugler (1991), included 27 independent studies from 1967 through 1989. The aims of the present meta-analysis were to address the limitations of the previous meta-analysis through improved coding and statistical methods, particularly related to the type of effect size statistics used, and to serve as an update two decades later. The main objective of the present study was consistent with that of the previous meta-analysis: to quantitatively synthesize the comparative research on competent and incompetent defendants. The following presentation of findings represents a summarization of nearly 50 years of research in this area.

Main Findings

Hundreds of articles, chapters, and books have been published in the competency arena since the early 1960s. A comprehensive literature search yielded 186 empirical research reports in this area; 88 of which met inclusion criteria for the present study. Twenty of these reports were classified as redundant samples; therefore, 68 independent studies have been conducted wherein competent and incompetent participants have been compared on at least one variable and for which sufficient data were reported to elicit an effect size. Included studies were predominantly articles (56/68) conducted with samples from the United States (52/68). Approximately half (37/68) of these studies included female participants, and the setting of most studies was inpatient (46/68), followed by a mix of inpatient and outpatient (11/68) and outpatient alone (9/68). Two studies did not report the setting. Most of the studies (59/68) utilized a referred sample to serve as the competent comparison group, while four studies used a "pure" competent group (i.e., no question of competency was posed), three studies included a restored competent group, one study used a study-competent group (Randolph, Hicks, & Mason,

1981), such that participants were initially deemed incompetent, but were classified as competent by the researchers, and one study used a mixed sample (Poythress et al., 1988). The majority of studies (40/68) employed an archival/retrospective data collection design, most of which (59) were convenience samples. The source of the competency decision used for the ultimate comparison or criterion was such that 20 studies used the decision of psychiatrist(s), 17 used a mix of psychiatrist, psychologist, and/or mental health professional team decisions, 14 used mental health professional team decisions, 13 used the court's decision of competency status, and only two studies relied on psychologist(s) decisions alone (Shields, 2005; Simon, 1987).

The total sample size for the included studies was 26,139; however, the split between competent and incompetent defendants was disproportionate. Most of the total sample consisted of competent participants (n=19,711) as compared to incompetent defendants (n=6,428). The mean total sample size across all included studies was 384.4 and the median was 176, but the sample sizes across studies varied greatly, ranging from 21 total participants to 8,416. The second largest sample size was 1,436 when the aforementioned sample of 8,416 from the study by Warren and colleagues (2006) was removed. The base rate of incompetency ranged from 7% to 70%, but the mean rate of incompetency was 27.5% as per the 59 studies that used non-matched samples. This finding is consistent with the base rate of incompetency found in the previous meta-analysis as well as with the rate described by numerous authors over the years.

In comparison to their competent counterparts, incompetent defendants were predominantly not Caucasian (52.3% versus 43.1%), had a much higher unemployment rate (70.8% versus 58.2%) and a greater percentage were not married (84% versus 77.3%). An investigation of psychiatric variables revealed the greatest discrepancies. Most incompetent defendants were diagnosed with a Psychotic Disorder (66.5%) and had a previous psychiatric

hospitalization (53.4%), and very few had been diagnosed with a Personality Disorder (8.2%). This pattern is different than the competent group, whereby only 22.2% were diagnosed with a Psychotic Disorder, 32.3% had a previous psychiatric hospitalization, and 27.9% were diagnosed with a Personality Disorder. Some similarities, however, were found between these two groups. Most participants in both group samples were males (84.1% incompetent, 81.9% competent) in their early 30s (incompetent mean = 35, competent = 31.8) with a prior arrest history (59.6% incompetent, 63.4% competent). Furthermore, approximately half of both group samples had a current violent criminal charge (50.8% incompetent, 55.1% competent); and, both groups averaged approximately 10 years of education (10.4 years for the incompetent group, 10.5 years for the competent group).

The relations between competency status and eight categorical variables most commonly investigated in the competency arena were examined. The number of studies presenting comparative data on each variable fluctuates and must preface the ultimate findings: ethnicity (n=22); sex (n=18); employment (n=8); marital status (n=10); psychiatric diagnosis (n=25); psychiatric hospitalization history (n=5); competency evaluation history (n=3); and current criminal charge (n=18). These numbers only reflect studies in which the authors presented data on the respective variables; therefore, additional researchers may have collected but not presented such data. Furthermore, a number of studies presented data on the aforementioned variables, but were excluded from the present meta-analysis because they did not utilize a competent comparison group (i.e., 35% of the excluded studies).

Categorical Variables. The second hypothesis of this study, that demographic variables would be associated with incompetency, was supported for three of four variables. A relationship was found for ethnicity, employment status, and marital status. Defendants who were not

Caucasian and not married were approximately one and a half times more likely than their respective counterparts to be found incompetent. The likelihood of being found incompetent was approximately double for unemployed defendants as compared to employed defendants; however, female defendants were essentially equally as likely as male defendants to be found incompetent.

The third hypothesis, that psychiatric and psycholegal variables would be associated with incompetency, was partially supported. Two out of four of these variables were related to incompetency. The most robust finding in this study was that defendants diagnosed with a Psychotic Disorder were approximately eight times more likely to be found incompetent than defendants without such a diagnosis. In addition, the likelihood of being found incompetent was approximately double for those with a previous psychiatric hospitalization versus those without such a history. The other two findings were that those those with a prior competency evaluation were also no more likely to be found incompetent than those without a prior evaluation, and defendants with a current violent criminal charge were nearly one and a half times more likely to be found *competent* than those who were not currently charged with a violent crime (Note: The finding related to current criminal charge was conceptually reversed in direction for ease of interpretation).

Categorical Moderators. To address the fourth hypothesis in this study, six potential moderators were examined for each relationship between competency status and the categorical variables: (i) type of publication; (ii) source of competency decision used for comparison; (iii) type of competent comparison group; (iv) study setting; (v) sample's country of origin; and (vi) recruitment of participants. Unfortunately, moderation could not be examined formally in most cases because of limited variability within moderator groups; however, statistically significant

moderation was found for three of the eight categorical variables (i.e., psychiatric diagnosis, ethnicity, sex), thereby partially supporting the hypothesis.

The effect for psychiatric diagnosis on competency status was moderated by the type of competent comparison group utilized. The cumulative effect size (odds ratio) representing the likelihood of being found incompetent if a defendant was diagnosed with a Psychotic Disorder was 7.96; however, two studies utilizing a pure competent comparison group (i.e., Hoge et al., 1996; 1997a) found the likelihood to be approximately 34 times, while findings from 21 studies using a referred group were more in line with the main finding (i.e., 7.94 times more likely). In addition, one study utilized a mixed competent comparison group and one used a restored group, and the findings were 12.95 and 1.71 times more likely, respectively.

The cumulative odds ratio for ethnicity (1.39) was moderated by the type of competent comparison group used as well, such that studies utilizing a "purely" competent comparison group found non-Caucasian defendants to be 2.33 times more likely to be found incompetent compared to studies using referred groups (1.2 times more likely) and restored groups (1.5 times more likely). Type of recruitment method also moderated the relationship between ethnicity and competency status. Studies employing a prospective recruitment method found non-Caucasian defendants to be 1.77 times more likely to be found incompetent than Caucasian defendants, while the relationship was found to be more neutral by studies utilizing an archival/retrospective method (1.17 times more likely).

Last, the sample's country of origin moderated the relationship between competency status and sex, which was essentially neutral. Female defendants were approximately twice as likely to be found incompetent than males in four Canadian studies (i.e., Crocker, Favreau, & Caulet, 2002; Robertson, Gupton, McCabe, & Bankier, 1997; Roesch, Eaves, Sollner,

Normandin, & Glackman, 1981; Rogers, Gillis, McMain, & Dickens, 1988) as compared to a relatively neutral finding across 14 American studies (1.10 times more likely).

Competency and Traditional Assessment Instruments. The fifth and final hypothesis was that scores on competency assessment measures and traditional measures (i.e., intellectual and personality assessment instruments) would both be correlated with competency decisions, but that larger effect sizes were anticipated for the relationship between scores on competency assessment instruments and such decisions. To test this hypothesis, an additional set of analyses was conducted to investigate the relationship between competency status and scores on traditional and competency assessment instruments. Comparative data on 12 competency measures, the Wechsler intelligence scales, the MMPI and MMPI-2, the BPRS was explored. While numerous studies had incorporated these measures into their designs, few presented sufficient comparative data for effect size coding.

Twelve competency instruments have been developed since the 1970s. The following number of studies and total sample sizes associated with studies using competency instruments reflect all empirical adjudicative competency studies (n=186) initially considered for the present meta-analysis: the Competency Screening Test (CST), 15 studies, total n=1,392; the Competency to Stand Trial Assessment Instrument (CAI), 6 studies, n=331; the Georgia Court Competency Test (GCCT/GCCT-MSH), 10 studies, n=1,126; Interdisciplinary Fitness Interview (IFI/IFI-R), 3 studies, n=174 the Fitness Interview Test (FIT/FIT-R), 10 studies, n=1,924; the Computer-Assisted Determination of Competency to Proceed (CADCOMP), 4 studies, n=378; the Competence Assessment for Standing Trial for Defendants with Mental Retardation (CAST-MR), 6 studies, n=352; the Metropolitan Toronto Forensic Service (METFORS) Fitness Questionnaire (MFQ), 1 study, n=188; the MacArthur Competence Assessment Tool–Criminal

Adjudication (MacCAT-CA), 8 studies, n =1,149; the Mosley Forensic Competency Scale (MFCS), 1 study, n=75; the Evaluation for Competency to Stand Trial–Revised (ECST-R), 5 studies, n=484; and, the Test of Malingered Incompetence (TOMI), 1 study, n=392.

Although much data has been collected on these measures over the years, reports from only eight independent studies presented comparative scores of competent and incompetent defendants for which an effect size was calculable. Moreover, sufficient data was only available for five of these measures (i.e., CST, GCCT-MSH, FIT, MFQ, MFCS), and such data from more than one independent study were only available for the CST and GCCT-MSH.

Two studies employed the CST and found 131 competent defendants (M=26.7, SD=8.3) scored nearly 10 points higher than 26 incompetent counterparts (M=17.0, SD=8.8). This finding is consistent with that of the GCCT-MSH. Four studies utilized the GCCT-MSH and found competent defendants (n=335) scored almost 26 points higher than incompetent defendants (n=102): 81.3 versus 55.6, respectively. The cumulative effect size remained large when data from the eight independent studies on the five measures were combined to investigate the overall difference between incompetent and competent defendants on competency measures. As noted earlier, the FIT, MFQ, and MFCS were used in one included study each; therefore, data on these measures were combined with data on the CST and GCCT-MSH. Incompetent and competent defendants' scores (total n=788) across competency assessment instruments significantly differed and a rather large effect size was found (i.e., d = 1.4). While robust and consistent with the previous results, this overall finding should also be interpreted cautiously because of the few studies and instruments included in its calculation, as well as the relatively small total sample sizes for both incompetent (n=214) and competent defendants (n=574) from which the data were drawn.

Similarly scant calculable effect size data was found with studies utilizing traditional assessment measures. The following descriptive information on traditional instruments includes all empirical adjudicative competency studies initially considered for the present meta-analysis (n=186). The Wechsler scales have been utilized in 36 competency studies (total n = 5,918): one study included the WASI, nine included the WAIS, 19 included the WAIS-R, and seven used the WAIS-III in their designs. The MMPI and its revision were utilized in 13 studies (total n = 3,501), and the BPRS was used in 12 studies (total n = 1,861). The use of other personality, cognitive, and neuropsychological measures was less common (see Table 26). Although many researchers have incorporated the aforementioned traditional measures into their study designs, few have presented comparative data on competent and incompetent defendants.

The main findings related to performance on the Wechsler scales were that competent defendants scored approximately 5-6 points higher than incompetent defendants across all three main indices of cognitive functioning. Three studies found competent defendants (total n=592) to achieve a mean Full Scale IQ score of 87, which is approximately six Full Scale IQ points higher than their incompetent counterparts (total n=366). Four studies found that competent defendants (n=412) scored approximately five Performance IQ points higher than a total of 170 incompetent defendants (85 versus 80). Finally, seven studies presented either Verbal IQ data or Verbal Cognitive Functioning (VCF) data generated from the WAIS-R. The mean VIQ/VCF score for competent defendants (n=729) was 87 as compared to 82 for incompetent defendants (n=409). All three overall effect sizes (i.e., standardized mean difference) are considered small to medium, ranging from 0.32-0.38. While consistent, these results should be interpreted conservatively in light of the few studies included and the relatively small sample sizes overall.

Comparative MMPI data for which effect sizes are calculable have only been presented in two studies (i.e., Maxson & Neuringer, 1970; Sachsenmaier, 1991). Incompetent defendants (total n=153) produced higher scores than their competent counterparts (total n=886) across three scales that have typically been explored in this area: the F scale, and scales 6 and 8. Small to medium effect sizes were found across the scales (i.e., 0.33, 0.39, and 0.33); however, once again, these results should be interpreted cautiously as only two studies are included and the total sample sizes are relatively small.

Only four of the twelve studies utilizing the BPRS presented sufficient data to calculate effect sizes. Incompetent defendants (total n=522) received a mean BPRS total score of approximately 40 compared to a score of 32 for competent defendants (total n=763); however, the group means and the subsequent overall difference score was slightly skewed as a result of Hoge and colleagues' (1996) finding of a 17-point difference between incompetent and competent groups. The total score differences for the remaining three studies ranged from three to six points.

The fifth hypothesis was supported. Scores on competency assessment measures and traditional measures (i.e., intellectual and personality assessment instruments) were related to competency decisions; however, larger effect sizes were found for the relationship between scores on competency assessment instruments and such decisions.

Comparison with Nicholson and Kugler (1991)

The conduct and writing of meta-analyses have become more sophisticated and standardized in practice than they were in the late 1980s when Nicholson and Kugler conducted their study. Nevertheless, it is important to compare and contrast their study with the present meta-analysis in terms of both study design and results. The present study consisted of data from

88 total manuscripts, 68 of which represent independent studies, including 11 dissertations, published between 1967 and 2007. Nicholson and Kugler's meta-analysis was based on 27 independent studies published between 1967 and 1989. In addition, Nicholson and Kugler also did not present their coding forms, manuals, or interrater reliability statistics.

The mean total sample size found in the present study neared 400 (i.e., mean = 384.5) with a median size of 176, while the previous study found a mean total sample size of 272.3. The base rate of incompetency found in the present meta-analysis was 27.5% compared to Nicholson and Kugler's finding of 30.6%. It is not possible to compare the meta-analyses with respect to the descriptive statistics of incompetent and competent subsamples because Nicholson and Kugler only presented such data for total samples rather than bifurcating the incompetent and competent sample data. The reporting of total sample size statistics is not applicable in the context of comparative competency research because such data is reflective of the characteristics of referred defendants rather than that of incompetent and competent defendants. For example, presenting the percent of a total sample of defendants diagnosed with a Psychotic Disorder is misleading without associated subsample statistics. In the present study, 44.4% of the total sample was diagnosed with a Psychotic Disorder; however, when divided, the groups were starkly different. Specifically, 66.5% of incompetent defendants carried such a diagnosis as compared to only 22.2% of their competent counterparts. The interested reader can compare and contrast the descriptive findings from both meta-analyses by referencing Table 3 of the present study and Table 1 of Nicholson and Kugler's meta-analysis.

Nicholson and Kugler found that 86.7% of studies utilized mental health professional decisions as the ultimate competency criterion compared to 77.9% found in the present study. Approximately three-quarters (73.3%) of their studies were conducted in inpatient settings; in the

present meta-analysis, 67.6% of studies were conducted inpatient-only and 16.2% were conducted in a combination of inpatient and outpatient settings. Nicholson and Kugler did not code for the type of competent group used, the sample's country of origin, or the type of recruitment or sampling employed in a study. In addition, the method of calculation for effect sizes was completely different between the meta-analyses. Nicholson and Kugler calculated correlation coefficients and Cohen's *h* statistics, while odds ratios and mean difference statistics were calculated in the present study. Last, Nicholson and Kugler did not perform formal moderation analyses.

Both meta-analyses investigated defendants' background characteristics and their performance on traditional and competency assessment instruments. Nicholson and Kugler coded for six demographic variables (i.e., age, gender, race, marital resources, education, and employment status). These variables were all included in the present study, although effect sizes were not calculated for age or education level because 2 x 2 tables were used in the present study to generate odds ratios. Nicholson and Kugler found statistically significant, yet small relations between findings of incompetency and three of the four remaining demographic variables: female gender (n=12, r=.09), minority race (n=12, r=.09), and marital status/married (n=5, r= -.08). A negative, but non-significant, association was found between being found incompetent and being employed (n=4, r= -.06). Relatively consistent results were found in the present study. The odds ratios (i.e., the odds of being found incompetent if "x") for each of the aforementioned variables neared neutral in the present meta-analysis with the exception of employment status: female (n=18, OR= 1.12), non Caucasian (n=22, OR=1.39), not married (n=10, OR=1.43), and unemployed (n=8, OR=2.07).

Nicholson and Kugler coded for three variables related to legal and psychiatric history (i.e., type of offense, previous legal involvement, previous hospitalization). The correlation between a finding of incompetency and nonviolent offense was virtually non-existent across 12 studies (r=.01), whereas the relationship between incompetency and having no previous legal involvement (n=4, r=.17) and having a previous psychiatric hospitalization (n=5, r=.26) were significant, yet small. Once again, interpretation of the present study's findings are similar, such that the odds of being found incompetent if currently charged with a violent crime approached neutrality (OR= 1.3) across 18 studies, and those with a psychiatric hospitalization history (OR=1.86) were nearly twice as likely to be found incompetent than those without such history.

Nicholson and Kugler also coded for a number of "psychiatric characteristics," including diagnoses of Mental Retardation and Psychosis, as well as eight types of psychiatric symptoms (e.g., disorientation). Diagnosis of Mental Retardation and the eight psychiatric symptom categories were also coded in the present study; however, effect sizes were only calculated for diagnosis of a Psychotic Disorder (yes/no). Like Nicholson and Kugler, the effect size associated with a psychotic diagnosis was the highest in the meta-analysis. Nicholson and Kugler found the association with psychosis and incompetency to be relatively large (r=.45) across 17 studies, and, in the present study, those diagnosed with a Psychotic Disorder were nearly eight times more likely to be found incompetent than those without the diagnosis (n=25).

With respect to psychological test performance, Nicholson and Kugler examined the relationship between scores on four competency assessment instruments and competency status. Moderately sized, negative correlations were found for each measure, such that poor performance was related to incompetency. Effect sizes for the Competency Screening Test (CST) were derived from data from 11 studies (r= -.37), four studies for the Georgia Court

Competency Test (CGGT; r = -.42), two studies for the Competency Assessment Instrument (CAI; r= -.52), and one study for the Interdisciplinary Fitness Interview (IFI; r= -42). The data utilized by Nicholson and Kugler to calculate the aforementioned correlations are unknown. Neither the first author nor the second-coder of the present study was able to replicate Nicholson and Kugler's findings while maintaining the parameters of the inclusion criteria (see *Analysis of Continuous Data* section above); therefore, meaningful comparisons between the findings of the present and previous meta-analyses are not possible. In the present study, sufficient data were only available for five competency instruments (i.e., CST, GCCT-MSH, FIT, MFQ, MFCS), and only the CST and GCCT-MSH had such data from more than one independent study. Competent defendants scored approximately 10 points higher than incompetent defendants on the CST across two studies. Four studies utilized the GCCT-MSH and found competent defendants to score almost 26 points higher than incompetent defendants on average. The FIT, MFQ, and MFCS were used in one included study each and, when data on these measures was combined with data on the CST and GCCT-MSH, the standardized cumulative effect size was relatively large (d=1.4).

Nicholson and Kugler also investigated defendants' performance on IQ[12] and four MMPI scales. The authors reported a small, negative relationship between findings of incompetency and intelligence test scores (r= -16) across eight studies. In the present study, standardized mean difference statistics were calculated to represent the relationship between competency status and Wechsler FSIQ, PIQ, and VIQ/VCF indices. Competent defendants scored approximately 5-6 points greater than their incompetent counterparts across all three indices, and all three effect sizes were small to medium, ranging from 0.32-0.38. Nicholson and Kugler found small, yet significant correlations representing the association between incompetency and four MMPI

scales across five studies; specifically, positive relationships were found for the *F* scale (r=.08), and scales 5 (r=.05), 6 (r=.08), and 8 (r=.08). Effect sizes for the *F* scale and scales 6 and 8 were calculated in the present meta-analysis. Incompetent defendants evidenced higher scores on average than competent defendants, and small to medium standardized effect sizes were found for each scale (i.e., 0.33, 0.39, and 0.33).

The present meta-analysis' results are comparable to those found by Nicholson and Kugler, but odds ratios and (un)standardized difference effect size statistics were used rather than correlation coefficients to provide for clearer and more intuitive interpretation for both researchers and practitioners. It is important to note that while standards of meta-analytic research have improved in the 20 years since Nicholson and Kugler's (1991) study was published, the present study still consists of limitations despite concerted efforts to improve upon the previous meta-analysis, three of which are described in the following section.

Limitations

The present study has built upon the previous meta-analysis (Nicholson & Kugler, 1991) and, in doing so, was able to take advantage of advances in meta-analytic methods and contemporary competency research and theory that have developed over the past 20 years. While the present study provides the field with a timely quantitative review of the competency to stand trial research, limitations remain.

The first limitation is related to moderation analysis. Nicholson and Kugler (1991) addressed potentially moderating variables, but only statistically investigated type of setting (i.e., inpatient vs. outpatient) and date of data collection (i.e., before vs. after the *Jackson* ruling).[13] They presented correlations representing the relationship between a number of patient characteristic variables (i.e., demographics, legal and psychiatric history, and psychiatric status)

and competency status for both sets of moderator categories. *T*-tests were then performed to determine whether the correlations significantly differed between moderators for each variable. Although moderation was more formally investigated in the present meta-analysis (via meta *F*-tests and meta-regression models), these analyses were considerably constrained because of the lack of variability within moderator groups across studies. The moderation analyses should be interpreted cautiously because of the limited primary data available for use in these analyses.

Second, data from some groups was combined for the purposes of comparative analysis (i.e., incompetent versus competent defendants). For example, when applicable, questionably fit groups' data was combined with that of unfit groups, and data from groups whose competency was never in question (e.g., inmates) was combined with that of competent defendant groups.[14] Combining data in this way is not unprecedented (see, for example, Robertson et al., 1997 and Rogers et al., 1998), but it is a method in need of further attention because, for certain variables, questionably fit groups have been found to be more similar to fit groups than to those who were unfit (e.g., McDonald et al., 1991) and competent inmates receiving mental health treatment have been found to be more comparable to incompetent defendants than mentally healthy inmates (e.g., Hoge et al., 1998; Otto et al., 1998; Poythress et al., 1998). Such findings and other possible results may have been missed by combining these data in the present meta-analysis.

Last, some studies included defendants who have had a previous competency evaluation (e.g., Reich & Wells, 1985; Robertson et al., 1997), while others excluded such defendants (e.g., Bluestone & Melella, 1978). The effects of aggregating data from these studies are likely minor, if noteworthy at all, but conceptual concerns remain. In one of the few studies wherein the issue of multiple competency evaluations was investigated, Reich and Wells (1986a) compared those with previous evaluations, or "repeaters," to those who had not been previously evaluated. They

found repeaters were more likely to be diagnosed with Schizophrenia and affective disorders, were less educated, and were found competent less often than their non-repeater counterparts. In addition to potential within-group discrepancies among incompetent defendants, researchers who include repeaters are forced to make a rather arbitrary decision as to which evaluation to include in their data set. For instance, Robertson and colleagues (1997) included data from the initial assessment of two participants with multiple evaluations during their data collection period, but they automatically excluded the initial evaluation data of two other participants which was gathered prior to the data collection period.

General Discussion

This section provides commentary on three areas deemed most relevant to consider for the future of competency research and practice: variables associated with competency status; primary study samples and presentation of data; and the use of traditional and competency assessment instruments. Standards of competency research and guidelines for future competency research are proffered in the fourth subsection.

Variables Associated with Competency Status

Researchers have investigated the relationship between competency status and demographic, psycholegal, criminological, and clinical variables consistently since the 1960s. Eight of the most commonly researched variables were investigated in this meta-analysis: ethnicity; sex; marital status; employment status; psychiatric diagnosis; psychiatric hospitalization history; competency evaluation history; and current criminal charge - most of which do not have a compelling association with findings of incompetency. In fact, only two of the eight variables (i.e., employment and psychiatric diagnosis) evidenced odds ratios above 2.0, such that unemployed defendants were approximately twice as likely to be found incompetent

than those who are employed, and those diagnosed with a Psychotic Disorder were approximately eight times more likely to be found incompetent than those without a psychotic diagnosis. Two of the six other variables, sex and competency evaluation history, produced neutral odds ratios (i.e., no difference in levels of likelihood), but the confidence interval surrounding the effect size for competency evaluation history is quite large and, therefore, to suggest a neutral relationship between that variable and competency status would be inappropriate. Only three studies were included in the analysis of competency evaluation history and their odds ratios differed tremendously: 0.23 (Hoge et al., 1997); 0.40 (Reich & Wells, 1995), and 13.56 (Rosenfeld & Ritchie, 1998). The confidence intervals surrounding the odds ratios of the four remaining variables (i.e., ethnicity, marital status, psychiatric hospitalization history, current criminal charge) all included a neutral odds ratio (i.e., 1.0). These data, coupled with the *fail-safe N* statistics associated with each odds ratio, provide modest support for the relationship between competency status and most of these variables; however, this interpretation is made cautiously in light of the relatively few studies included in some of these analyses (e.g., psychiatric hospitalization and competency evaluation history analyses included 5 and 3 studies, respectively).

An additional issue to consider is that odds ratios were calculated for variables independently in the present study despite their actual interdependence; therefore, conditional, or joint, probabilities were not estimated. In other words, what is the associated likelihood of being found incompetent to stand trial for an unemployed, minority female who has never been married and who has been diagnosed with Schizophrenia? Unfortunately, this type of question cannot actually be answered with any scientific certainty in the present study because it would have required that primary research data be presented in a factorial manner with numerous

variable combinations resulting in many cells. For example, 256 possible defendant characteristic combinations would exist if a researcher were to collect data dichotomously (i.e., yes/no) for the eight categorical variables included in the present study (i.e., 2 x 2 x 2 x 2 x 2 x 2 x 2 x 2). Such data collection would obviously be overwhelming and lead to useless information; however, important information may be gained in the future if some variable combinations were investigated; specifically, it would be useful to first investigate the combination of a Psychotic Disorder diagnosis, unemployment, and a psychiatric hospitalization history, as the largest effect sizes found in this study were associated with these variables. An essential aspect of such an analysis would be to investigate the level of multicollinearity among these variables because of their close association to one another.

Primary Studies' Samples Used and Presentation of Data

Studies were included in the present meta-analysis if they presented comparative data of incompetent and competent defendants on at least one variable. The type of comparison groups used in each study were coded as potential moderators to circumvent problems with aggregating data from dissimilar samples, which could potentially lead to making inappropriate generalizations. Five types of competent comparison groups were identified: (i) referred; (ii) pure; (iii) restored; (iv) study-competent; and, (v) mixed. These groups significantly differed in only two analyses (i.e., ethnicity and psychiatric diagnosis); however, statistical significance was difficult to achieve regardless because so few studies employed non-referred samples. Nevertheless, the type of competent comparison group used in a study can have major implications. Perhaps the most salient example is that of Hoge and colleagues (1996) who utilized an attorney-referred competent comparison group whose competency was never in question (i.e., "pure" competent). The cumulative odds ratio for the relationship between being

diagnosed with a Psychotic Disorder and being found incompetent was approximately eight; however, the odds ratio generated from Hoge and colleagues (1996) was 207.4 as 30 of the 42 incompetent defendants were diagnosed with a Psychotic Disorder, while none of the 42 competent defendants had been given that diagnosis.

The MacArthur studies conducted in the late 1990s further illustrate the significance of the type of competency comparison group used in a study. These studies (i.e., Hoge et al., 1998; Otto et al., 1998; Poythress et al., 1998) incorporated three groups into their sample: a Hospital Incompetent (HI) group comprised of defendants admitted to forensic psychiatric units after being adjudicated incompetent to proceed; a Jail Treated (JT) group consisting of defendants in jail receiving treatment for mental health problems but presumed competent; and a Jail Unscreened (JU) group of randomly selected jail inmates presumed competent. The JT and JU groups were combined in the present meta-analysis because they are competent groups, while the HI group served as the incompetent sample; however, the JT group was more similar to the HI group than to the JU group on a number of variables. For example, the number of previous psychiatric hospitalizations, BPRS scores, the MMPI Psychoticism scale, and psychiatric diagnoses of the JT group were more comparable to the HI group than to the JU group. Studies using "questionably fit" groups elicited similar concerns (e.g., McDonald, Nussbaum, & Bagby, 1991; Nussbaum, Mamak, Tremblay, Wright, & Callaghan, 1998; Rogers et al., 1988). Data from questionably fit groups were combined with that of unfit groups in this meta-analysis because they are conceptually more similar than those deemed competent. As Rogers and colleagues (1988) articulated, combining these groups is justified because questionably fit defendants are typically referred for further assessment and because the "resulting two groups provide a general measure of subjects where (1) patients are clearly fit and (2) where fitness is at

issue" (p. 193). These types of samples pose dilemmas for the meta-analyst with regard to data aggregation; however, future studies investigating the difference between HI and JT groups and those incorporating questionably fit groups may be beneficial in trying to understand "grey area" cases – which are perhaps the most important to consider when attempting to understand the construct of adjudicative competency.

Of the 88 excluded independent studies, 31 did not utilize a competent comparison group; 22 were competency restoration studies; 14 studies met the main inclusion criteria but did not present sufficient data to code; 9 studies used entire samples of participants diagnosed with Mental Retardation; and 12 studies were excluded for other reasons. For example, the samples used by Quinsey, Pruesse, and Fernley (1975) and Ryba (2005) incorporated defendants found not guilty by reason of insanity (NGRI) in addition to those found incompetent to stand trial.

Twenty-two competency restoration studies were excluded from the present meta-analysis because they did not incorporate competent comparison groups and were comprised of participants found incompetent to stand trial had been remanded for competency restoration (e.g., Bennett & Kish, 1990; Carbonell, Heilbrun, & Friedman, 1992; Siegel & Elwork, 1990). Competency restoration studies have typically either investigated predictors of restorability (e.g., Hubbard & Zapf, 2003) or have compared restored defendants to those not restored on treatment outcome variables, such as restoration to competency (yes/no) and/or length of hospitalization (e.g., Mossman, 2007; Nicholson & McNulty, 1992). While 22 of these studies were excluded, three studies utilizing samples comprised of defendants who were all initially found incompetent to stand trial but whose competency was restored were included: Grandjean (2004), Lesser (1990), and Randolph, Hicks, Mason, and Cuneo, (1982). These studies were included in the present meta-analysis because they presented comparative data on incompetent and (restored)

competent defendants and because the restored nature of the competent comparison group was coded as a potentially moderating variable (i.e., under *type of competent comparison group*).

Coding the type of competent comparison group is important to prevent against aggregation of data from dissimilar studies. For example, Randolph and colleagues (1982) employed a sample of 45 male defendants found incompetent to stand trial and data from 39 of them was analyzed in the study. All 39 defendants were recommended competent by the treating psychiatrists and were subsequently evaluated by the court psychiatrists who deemed 10 to be incompetent and 29 competent. The relationship between performance on the Competency Screening Test (CST) and the opinions of the court-appointed evaluating psychiatrists was investigated; however, no actual scores were presented in the study. Rather, a CST cutoff score of 21 (equaling competence) was cross-tabulated with the psychiatrists' ultimate opinion in a 2 x 2 table. It is likely that Nicholson and Kugler (1991) incorporated these data into Table 7 of their meta-analysis, whereby they presented statistics on the proportion of defendants "passing" or "failing" the CST and three additional competency measures (i.e., GCCT, CAI, IFI). As a result, they generated a cumulative effect size without acknowledging that the combined studies employed dissimilar methods.

One study using a study-competent comparison group was included in the present meta-analysis. Randolph and colleagues (1991) utilized a sample of 25 male defendants found incompetent to stand trial. The defendants were re-classified by a psychiatrist as 15 incompetent and 10 competent and their scores on the CST were compared. This study was also included in the previous meta-analysis conducted by Nicholson and Kugler (1991) but, once again, the way in which the groups were bifurcated was not acknowledged. A number of studies excluded from the present meta-analysis used similar methods to classify defendants. The studies conducted by

Barnard and colleagues (1991, 1992), which outlined the development of the *Computer-Assisted Determination of Competency to Proceed* (CADCOMP) provide examples in this regard. In both studies, all defendants were initially found incompetent to stand trial (n=50 and n=99) but their status was re-classified by a forensic psychiatrist and the ratings of two psychiatry fellows and one master's level clinical psychologist based on the psychiatrist's interview. The CADCOMP was subsequently administered to the defendants and its psychometric properties were investigated. While the CADCOMP may be a useful measure for the evaluation of adjudicative competency, its utility is in question because of the samples used in these studies.

Nine of the excluded studies utilized a sample of incompetent defendants diagnosed with Mental Retardation. Diagnosis of Mental Retardation and IQ scores were coded in the present meta-analysis, but these studies were excluded because their data would have confounded the results if aggregated with data from studies incorporating samples with a range cognitive functioning. On the other hand, a few studies explicitly *excluded* defendants diagnosed with Mental Retardation (e.g., Hoge et al., 1997a; Simon, 1987) and they were included in the present study. Such exclusion also presents potential problems with generalizability, albeit much less worrisome, because the majority of defendants whose competency is in question do not have a diagnosis of Mental Retardation. Nevertheless, adjudicative competency research relevant to the developmentally disabled is extremely important; much of which has been pioneered by Dr. Caroline Everington, who developed the *Competence Assessment for Standing Trial for Defendants with Mental Retardation* (CAST-MR; see Everington, 1989; 1990; Everington & Dunn, 1995; Everington, Notario-Smull, & Horton, 2007).

Some studies addressed the issue of malingered incompetency by including bona fide malingerers or coached simulators in their samples. Although data from these studies could have

inappropriately affected the results if aggregated with those from studies not including such samples, the nature of these studies was not enough to exclude them without further consideration. Studies utilizing such samples exclusively were excluded (e.g., Gabel, 2007; Miller, 2004; Vitacco, Rogers, Gabel, & Munizza, 2007); however, studies presenting data on competent and incompetent participants in addition to malingering groups were included (e.g., Gothard et al., 1995; Jackson, Rogers, & Sewell, 1995), and only data from competent and incompetent defendants was coded and meta-analyzed.

Of the 88 excluded studies, 14 met preliminary inclusion criteria but did not present data from which coding was possible. What is arguably more unfortunate is that many included studies presented data not relevant to the comparative study of incompetent and competent defendants per se. Such studies incorporated incompetent and competent samples; however, data from subsamples was presented. For example, Laczko, James, and Alltop (1970) used an archival sample of 421 male pretrial competency defendants from the Forensic Unit of Dorothea Dix Hospital in North Carolina between 1958 and 1964. The only data able to be coded from this study was total sample size, incompetent sample size, competent sample size, percentage deemed incompetent, and design-related variables (e.g., type of setting). The authors presented data pertaining to diagnosis, education level, marital status, military service, previous convictions, previous mental health hospitalizations, occupation, IQ, disposition, and length of stay; however, such data was presented by type of current criminal charge (i.e., murder, assault, sexual, larceny, other) rather than by competency status. Classifications other than competency status such as gender (e.g., Riley, 1998) and psychiatric diagnosis (e.g., Viljoen, Roesch, & Zapf, 2002) have also been used as comparison groups within incompetent and competent samples. As such, the research question is not: *How do incompetent defendants compare to competent defendants on X*

variable(s)?; but rather, *Is there a difference between male and female defendants referred for competency evaluations with respect to their ultimate competency status?* As a result, the population to which one can generalize is not incompetent or competent defendants but rather defendants referred for competency evaluations.

Many authors provided descriptive statistics pertaining to their entire referred sample without presenting comparative descriptive information on incompetent and competent defendants within their samples (e.g., Caldwell, Mandracchia, Ross, & Silver, 2003; Mosley, Thyer, & Larrison, 2001; Mueller, 2007; Nicholson, 1988; Nicholson, Briggs, & Robertson, 1988; Nicholson, Robertson, Johnson, & Jensen, 1988; Ohayon, Crocker, St-Onge, & Caulet, 1998; Reich & Wells, 1985; Robbins, Waters, & Herbert, 1997; Roesch et al., 1997; Sikorski & Benedek, 1977; Webster, Menzies., & Jackson, 1982). Some researchers presented outcome data on different variables for the full sample and the incompetent and competent sub-samples. For example, Shatin & Brodsky (1979) compared low (≤20) and high (≥21) scores on the Competency Screening Test (CST) between incompetent and competent defendants, but only presented Wechsler IQ scores of the entire sample. Information on referred defendants as a population itself is important, and significant differences between referred and non-referred defendants have been found (see Aubrey, 1988; Cooke, Johnston, & Pogany, 1973; Viljoen & Zapf, 2002); however, such information is ancillary when presented in a study wherein differences between incompetent and competent defendants is the focus. Although the research methods and statistical analyses employed in these studies are sound, an abundance of comparative data on incompetent and competent defendants is not presented and, therefore, unavailable to the field.

In some studies data was presented in such a way that it could be misinterpreted when taken out of context. For example, a study conducted by Heller and colleagues (Heller, Traylor, Ehrlich, & Lester, 1981; 1983) is problematic in a number of ways. The authors used competency reports, rather than defendants themselves, as the unit of study; therefore, their sample appears to be comprised of 410 defendants rather than the 386 who actually participated. The language in the 1981 manuscript creates some confusion in this regard: "It is interesting to note that for the sample of 410 pre-trial defendants" (p. 270). The interested reader would have to wait approximately two years for clarification of the actual sample size: "The 410 psychiatric reports involved 386 individuals; 24 defendants were the subject of a reevaluation when they were thought to have regained competency" (p. 417, Heller et al., 1983). Using reports rather than defendants poses problems across all analyses conducted by the authors. For example, Heller and colleagues (1981) indicated: "Of the 410 defendant reports included in the sample, records revealed that 47.8 percent had their intelligence tested using standard intelligence tests" (p. 270). It is impossible to know how many defendants actually had their intelligence tested based on this presentation of information and, as such, interpretation of their between-groups analysis incompetent and competent defendants' IQ scores is compromised. (The measure of IQ itself is noteworthy, as the majority of defendants had their intelligence measured with the Western Personnel Test). Similarly, Johnson and colleagues (1990) administered certain tests (e.g., MMPI) to portions of their sample but did not present the sub-sample sizes.

Perhaps the greatest frustration for a meta-analyst is missing data. No study contains all the data that one may want to extract; however, many authors did not present vital statistics relevant to the area of study. Descriptive data was missing most frequently in studies that employed multivariate or psychometric statistical techniques (e.g., regression analyses: Daniel,

Beck, Herath, Schmitz, & Menninger, 1984; Mueller, 2007; Tillbrook, 2001; factor analysis: Bagby, Nicholson, Rogers, & Nussbaum, 1992). In some cases data were not presented in a manner conducive to coding for the present meta-analysis. For instance, Cooke (1969) and Lesser (1990) presented incompetent and competent group means for a number of MMPI scales without providing the associated standard deviations. Descriptive data and essential statistics should always be presented with future meta-analyses in mind even if outside of the scope of one's study.

Use of Traditional and Competency Assessment Instruments

Effect sizes corresponding to the relationship between competency status and scores on competency measures were larger than those pertaining to scores on traditional measures in the present and the previous meta-analysis; however, the available data in this regard is statistically and conceptually limited.

Nicholson and Kugler (1991) acknowledged one major limitation with respect to the fact that effect sizes associated with competency measures were much larger than those corresponding to traditional measures: none of the primary study authors compared traditional and competency assessment instruments within the same study. Despite proffering an interpretive caveat in this regard (i.e., effect size differences may be attributable to such internal variables as study methodology or sample characteristics), they maintained that the larger correlations associated with competency instruments were meaningful:

> Notably, the larger correlations for forensic instruments were obtained despite the fact that the deck was stacked against them: The correlations between traditional tests and competency were obtained from studies in which forensic examiners had access to the test findings during the course of an evaluation, whereas the correlations between

> forensic devices and competency virtually always came from studies designed to validate the instruments, so that examiners were unaware of the test results. In other words, the forensic instruments were better at predicting an independent criterion than the traditional instruments were at predicting a contaminated criterion. (p.364)

The following are examples of noteworthy limitations and confounds of Nicholson and Kugler's effect size analyses of intelligence tests that they did not address in their manuscript. Heller and colleagues (1981, 1983) administered the Western Personnel Test to the majority of their sample and used the WAIS-R "for those who were either apparently illiterate or those who scored poorly on the Western Personnel Test" (Heller et al., 1981, p, 270). These data should not have been indiscriminately combined with data from studies using WAIS-only protocols. Second, Chellsen (1986) used a sample of 25 offenders diagnosed with Mental Retardation; and, competency status was determined by performance on the LAW-COMP Exam. The author's use of a developmentally disabled sample and their use of an unconventional measure to determine competency status should have precluded the aggregation of this study's data with others investigating the relationship between IQ scores and competency status. More generally, the few number of studies included in each analysis and their associated small sample sizes represent a major limitation (i.e., eight studies were included in the IQ analysis consisting of a total sample size of 882 defendants). In the present meta-analysis, competent defendants scored 5-6 points higher than incompetent defendants across all three main indices of cognitive functioning as measured by the Wechsler Adult Intelligence Scales (i.e., FSIQ, VIQ/VCF, PIQ), but the same limitations were present: three studies (n=958) provided data on Full Scale IQ; four studies (n=582) on PIQ; and seven studies (n=1,138) on either Verbal IQ data or Verbal Cognitive

Functioning (VCF). All three cumulative effect sizes (i.e., standardized mean difference) were considered small to medium, ranging from 0.32-0.38.

With respect to the MMPI, Nicholson and Kugler's analyses were based on data from 1,461 defendants across five studies on four scales (*F*, 5, 6, 8). In contrast, only two studies (i.e., Maxson & Neuringer, 1970; Sachsenmaier, 1991) were included in the present meta-analysis, which presented comparative MMPI/MMPI-2 data from which effect sizes were calculable. Incompetent defendants (total n=153) scored higher than their competent counterparts (total n=886) across three scales that have typically been explored in this area: the *F* scale, and scales 6 and 8. Small to medium effect sizes were found across the scales (i.e., 0.33, 0.39, and 0.33).

Nicholson and Kugler's analyses of the competency assessment instruments were also conducted on few studies with small sample sizes: Competency Screening Test (CST), 11 studies (n=627); Georgia Court Competency Test (GCCT), 4 studies (n=527); Competency Assessment Instrument (CAI), 2 studies (n=90); and, the Interdisciplinary Fitness Interview (IFI), 1 study (n=69). Furthermore, because they did not present the references for the studies used in each analysis, it is impossible to know exactly which studies were included in their calculation of effect sizes. There was one exception: only one study was included in the analysis of the relationship between competency status and the *Interdisciplinary Fitness Interview* (IFI). Nicholson and Kugler found a relatively large effect size (i.e., r = -.42) representing the relationship between a finding of incompetency and IFI scores. This effect size was calculated using data from the only study incorporating the IFI at that time (Golding, Roesch, & Schreiber, 1984; Schreiber, 1982; Schreiber, Roesch, & Golding, 1987).[15] This study was also included in the present meta-analysis, but data on the IFI could not be coded because of numerous conceptual and statistical constraints. First, Nicholson and Kugler reported IFI data on 69

defendants, but this sample size could not be confirmed. Schreiber (1982) and Schreiber et al. (1987) reported collecting IFI data from 78 defendants, and Golding et al. (1984) reported having such data on 77 defendants, but two cases were removed during statistical analyses resulting in a sample size of 75.[16] Second, it is unclear how Nicholson and Kugler calculated an effect size at all. None of the three abovementioned reports presented actual IFI data per se; rather, the results presented across manuscripts included: levels of agreement between the IFI interviewers and a blue ribbon panel (Golding et al., 1984), hospital staff (Schreiber, 1982), and Competency Asessment Instrument (CAI) and Competency Screening Test (CST) scores (Schreiber et al., 1987); the psychometric properties of the IFI, namely interrater reliability statistics. No normative data (e.g., means or standard deviations) related to performance on the IFI were presented in any of the three manuscripts.

Another example of a study often cited in the literature, but not used in the present meta-analysis, is Rogers, Ustad, Sewell, and Reinhardt's (1996) factor analytic study of the Georgia Court Competency Test (GCCT). While the study was methodologically and statistically sound, and helped delineate the factor structure of the GCCT in a mentally disordered offender sample, the results are not directly generalizable to a sample of defendants whose competency to stand trial is in question. The sample was comprised of 125 mentally disordered offenders housed in the mental health unit of the Tarrant County Jail in Texas. None of the offenders' competency was in question, although the authors "anticipated that this sample would have mentally disordered offenders that were likely to be incompetent to stand trial" (p. 324). If the authors' assumption was correct, their results are difficult to interpret because they did not differentiate between incompetent and competent offenders in their sample (the study was not included in the present meta-analysis because it was not comparative). Rogers and colleagues' findings spoke to

the factor structure of the GCCT when used with mentally disordered jail inmates and not defendants whose competency is in question because psychometric properties are technically characteristics of data, not instruments (Messick, 1995).

Comparative data related to defendants' performance on competency assessment instruments remains scant. Only two studies (n=157) with data on the Competency Screening Test (CST) and four studies (n=437) on the Georgia Court Competency Test-Mississippi State Hospital (GCCT-MSH) were included in the present meta-analysis. Competent defendants scored 10 and 26 points greater than incompetent defendants on these measures, respectively. Data from individual studies presenting comparative data on the FIT, MFQ, and MFCS were combined with the aforementioned data on the CST and GCCT-MSH. Incompetent and competent defendants' scores on competency assessment instruments combined from eight independent studies (total n = 788) resulted in a rather large effect size (i.e., $d = 1.4$).

It was initially surprising that data from so few studies was available for effect size analyses pertaining to traditional and competency assessment instruments in light of the numerous publications in this area over the past 50 years; however, studies are likely to be more heavily scrutinized by meta-analysts than they are by those authoring books/chapters, journal articles, and qualitative reviews. Even when authors of such publications are scrupulous, they usually do not employ formal inclusion criteria for studies incorporated in their literature review. In addition, limitations and relevant interpretive caveats are usually only briefly mentioned and may be overshadowed by the paper's main points (e.g., significant findings).

There is no way to know how many times traditional and competency assessment instruments have actually been incorporated into research studies and not reported, or how much unpublished data has actually been collected; therefore, empirical knowledge is based on

available research. There is insufficient empirical evidence to support the superiority of competency assessment instruments over traditional measures at this time. Nonetheless, insufficient empirical evidence is not equivalent to a lack of empirical evidence, and the conclusions drawn from the present meta-analysis should not be interpreted to mean traditional measures have utility comparable to competency measures for use in competency evaluations.

Nicholson and Kugler's findings with respect to traditional instrument usage were also presented without their associated limitations and caveats:

> Defendants with lower IQ scores and defendants with higher scores on Scales *F,* 5, 6, and 8 of the MMPI were more often judged incompetent. However, these correlations were small at best (all rs < .08), confuting the argument that instruments such as the MMPI can be used for screening purposes in competency evaluation (cf. Maxson & Neuringer, 1970). (p. 363)

Competency assessment instruments, as a class of forensic assessment instruments, have been developed to address specific psycholegal questions related to the psycholegal construct of adjudicative competency; therefore, they are conceptually appropriate for use in competency evaluations. Still, the debate on the use of forensic *versus* traditional assessment instruments is an oversimplified characterization of the often-complicated decisions associated with choosing appropriate assessment measures for evaluations. The previous and presents meta-analyses both found significantly larger effect sizes for the relationship between competency status and competency measures than for the association between competency status and traditional measures, but these findings do not negate the fact that traditional assessment instruments can be useful in competency evaluations for specific reasons.

The reality is the MMPI and MMPI-2 are very robust measures encompassing hundreds of scales that may never be fully explored in the adjudicative competency arena; however, comparative research on the nine validity scales and 10 additional primary clinical scales is needed before definitive recommendations regarding its utility for competency evaluations can be proffered. It is far-reaching to dismiss the potential utility of the MMPI/MMPI-2 in competency evaluations based on data from five studies on four of its scales. Personality measures can be useful in establishing the existence of a mental illness (a threshold issue) or evaluating malingering. The MMPI-2, for example, is constantly evolving and is comprised of a multitude of scales and subscales, which may have utility in this regard. Nevertheless, when measures are used inappropriately, they can be useless or misleading. Thus, the question should not be *if* traditional measures should be used in competency evaluations but rather, *when* and *how*. Unfortunately, edited books devoted to the use of traditional measures in forensic evaluations provide virtually no guidance in this regard. Archer's (2006) book on the forensic uses of clinical assessment instruments includes whole chapters on the MMPI-2, PAI, MCMI-III, PCL-R, and Rorschach; however, the use of these measures in competency evaluations is only addressed in two places: one paragraph in the MMPI-2 chapter authored by Sellbom and Ben-Porath and three paragraphs on the Rorschach in Weiner's chapter. Gacono, Evans, Kaser-Boyd, and Gacono's (2008) recently edited a handbook on forensic Rorschach assessment, which contains a full chapter on its use in trial competency evaluations written by Gray and Acklin. Although the authors presented cogent arguments for using the Rorschach in competency evaluations, they offered no empirical support for its utility in such evaluations – likely because none exists. There are currently no published studies available that present Rorschach data from incompetent and competent defendant samples and only two studies even mention the

incorporation of the Rorschach in their designs (i.e., Laboratory of Community Psychiatry, 1974; Pfeiffer, Einstein, & Dabbs, 1967). It is also noteworthy that only one study mentioned the use of the Thematic Apperception Test (TAT) and Draw-a-Person Test with incompetent/competent samples, but no data were presented (i.e., Vernon, Steinberg, & Montoya, 1999).

Nicholson and Kugler found a small effect size (r = -.16) to represent the relationship between IQ score and competency status; however, they did not provide an explanation nor interpretation of this finding. As a result, it can easily be taken out of context and misinterpreted to support inferiority of traditional measures in the assessment of adjudicative competency. All of the studies included in the present and previous meta-analysis have presented data on three main indices of intellectual functioning: Full Scale IQ, Verbal IQ or VCF and Performance IQ. It is noteworthy that The VIQ and PIQ indices are no longer calculated by the new Wechsler adult intelligence measure (WAIS-IV). In addition, Full Scale IQ score in and of itself provides insufficient information to the competency evaluator, especially without knowing whether it should be interpreted in the first place (i.e., the presence of a significant difference split in scores between verbal and performance indices). The findings of the present and past meta-analysis should not be interpreted as evidence against the use of traditional measures in competency evaluations. In fact, the assessment of cognitive functioning is quite important.

Zapf (1999) argued, "what makes and individual competent is..*cognitive organization*" (p. 78), which she operationalized as an ability to understand, process, and express or communicate information. Zapf argued that cognitive organization was necessary but not sufficient for trial competency, and suggested defendants must not possess any context-specific impairment, or thought processes not based in reality, specific to the context in which the question of competency has arisen, such as a delusional belief regarding the forthcoming

criminal procedures. These abilities comprise the legal standard of competence à la *Dusky*; as such, evaluating cognitive abilities is an essential component of the functional assessment of a defendant's competency (Grisso, 2003; Zapf, 2009) or, as Skeem and Golding (1998) articulated, "the fundamental task for a forensic examiner is to relate any psychopathological or cognitive difficulties to possible impairments in the defendant's psycholegal abilities" (p.358).

Research on the cognitive/neuropsychological assessment of competent and incompetent samples has increased in recent years. Nussbaum and colleagues (1998) administered a cognitive battery in addition to the METFORS Fitness Questionnaire and concluded, "Empirically we have provided initial evidence that the legal fitness concept appears grounded within a cognitive psychological foundation" (p. 59). Nestor and colleagues (1999) administered a comprehensive neuropsychological battery to a sample of 181 patients committed to Bridgewater State Hospital who had undergone competency evaluations between 1987 and 1995. They found the greatest differences between competent and incompetent defendants on tests of memory, particularly verbal memory, "which assess the ability to acquire, encode, retain, and retrieve new verbal information" (p. 407), but found no differences on standardized tests of academic skills. Grandjean (2004) found competent defendants differed from incompetent defendants in four cognitive domains (i.e., verbal memory, verbal comprehension, social judgment, and executive functioning), but not on measures of visual memory, visual spatial skills, or attention. The significance of cognitive abilities in the competency context is further appreciable by the very existence of the literature pertaining to juvenile defendants and those diagnosed with Mental Retardation, wherein such abilities are directly implicated.

The present and previous meta-analysis both found larger effect sizes corresponding to the relationship between competency status and competency assessment instruments than those

associated with traditional measures; however, the data from which these effect sizes have been calculated are limited for a number of reasons and additional research on both classes of measures is needed before empirically-supported conclusions can be made. Nevertheless, some evidence exists for the ability of certain measures and scales to differentiate between competent and incompetent defendants. Competency assessment instruments are useful because they address competence-related abilities directly as per the relevant legal standard (i.e., *Dusky*), but evaluators must be mindful when choosing which measure to use because of the variability in their utility (see Grisso, 2003; Melton et al., 2007; Zapf & Viljoen, 2003). Traditional assessment instruments can be useful in competency evaluations; however, research and commentary to date has not adequately addressed *how* they may be used. Additional research is needed with these measures, particularly those assessing cognitive functioning.

Competency Research Guidelines and Future Directions

As with most meta-analyses, the present study does not represent a final step in the empirical examination of adjudicative competency, but rather a starting point for further investigation. Eagly and Wood (1994) delineated how to use meta-analyses to plan and inform future research by outlining five ways in which a meta-analysis's findings should be critiqued: (i) the level of certainty/discrepancy of findings; (ii) the generalizability of findings to new populations, settings, and research designs; (iii) the unexplained variability across moderators of main-effect findings, as well as moderators identified post hoc and those with ambiguous interpretations; (iv) identifying potential mediator variables; and (v) false certainty from incomplete data analysis, inadequate sample sizes, and/or from the use of an overinclusive definition of the meta-analysis domain. The extent to which the present meta-analysis addresses each of the aforementioned areas is a somewhat subjective question – one which can be

addressed fairly and scientifically so long as this study's overarching goal has been met: to clearly and transparently present its methods and the results. Critical evaluation of primary research studies, qualitative reviews, and the present and previous meta-analysis led to the development of the recommended competency research guidelines and directions for future research presented in this section.

Competency Research Guidelines

Much was learned from conducting this meta-analysis, both with respect to the aggregation of 50 years of data and the empirical investigation of adjudicative competency itself. It is obvious when a primary research study lacks essential information or data during the coding stage of a meta-analysis; indeed, a missing data code is recorded each time this occurs. While missing data is an inherent concern in all research, and the extensiveness of a meta-analysis' coding form virtually ensures the existence of missing data, data should be presented in a consistent fashion across competency studies. Thus, 12 guidelines have been developed for those conducting comparative research with competent and incompetent samples. These guidelines are intended to foster the cohesion and comparability of competency publications, which would greatly benefit future meta-analysts and, subsequently, the field.

1. *Explicitly and clearly present information when publishing multiple manuscripts from the same study.* As per the publication manual of the American Psychological Association (2001), "Piecemeal, or fragmented, publications of several reports of the results of a single study is undesirable unless there is a clear benefit to scientific communication" (p. 352-353). Authors often present their findings in "pieces" for legitimate reasons, including a desire to write papers with different scopes and conceptual trajectories; however, coding a single study from multiple publications

can be challenging for the meta-analyst for various reasons (e.g., the amount of coding decisions is increased, information such as subsample sizes may not be congruent across studies). Thus, authors must explicitly and clearly present methodological information and descriptive and inferential statistics when publishing multiple manuscripts from the same study. In addition, authors should present all available data for the total sample and the incompetent and competent subsamples.

2. *Investigate potential differences between participants and non participants when possible.* Researchers should make efforts to determine whether significant differences exist on competence-related variables between those who participate in their study and those who do not.
3. *Always present means and standard deviations and effect sizes.* This tenet should serve as a reminder to all, but particularly to those who are conducting multivariate statistical analyses as well as those exploring the psychometric properties of a measure. Means and standard deviations for incompetent and competent groups were missing from the majority of studies employing such statistical methods as multiple regression, discriminant function analysis, structural equation modeling, and factor analysis. The meta-analyst will need to extract the means and standard deviations or effect sizes from a primary study; therefore, comparative group data should be presented in addition to the results of more complicated statistical analyses.
4. *Present significant and non-significant findings.* It was not uncommon for authors to mention non-significant findings rather than formally present them. This problem is rooted in the classical statistical approach of null hypothesis testing, which has been highly criticized over the years (see Rindskopf, 1997 for a discussion on testing

"small," not null, hypotheses). Researchers should always present non-significant results, including means and standard deviations, regardless of their statistical approach (i.e., classical or Bayesian). Significant and non-significant findings are equally as important to the meta-analyst whose task is to aggregate findings from numerous studies.

5. *Present data continuously rather than categorically when possible.* Continuous data allows for the use of more statistical analyses to both the primary researcher and the meta-analyst as compared to categorical data. Although some variables are inherently categorical (e.g., sex), others are occasionally presented categorically despite being inherently continuous, such as scores on traditional and competency instruments, age, and level of education. Aggregation of such data is limited because it must be combined in terms of frequencies rather than by means, which in turn, limits many types of statistic analyses.
6. *Use an appropriate competent comparison group.* Studies investigating the characteristics of incompetent samples should incorporate competent comparison groups in their design. Five types of competent comparison groups were coded in the present study: referred defendants, purely competent defendants whose competency was never in question (e.g., inmates), those restored to competency, defendants who were initially deemed incompetent but then classified as competent by the researchers (i.e., study-competent), and a mixed group. The type of competent comparison group used is conceptually important and speaks directly the generalizability of findings to the population of interest. Delineating the differences between incompetent and competent referred defendants is informative to forensic evaluators, while

understanding the differences between incompetent and purely competent groups would be helpful for courts/attorneys with respect to referrals. Research conducted with restored samples can be extremely useful in investigating the psychometric properties of traditional and competency measures in the context of within-subjects research designs (i.e., pre/post). Purely competent participants, such as inmates, can be a useful comparison group because they serve as a matched-sample for relevant competence-related variables. On the other hand, study-competent comparison groups should never be used. Reclassifying defendants' competency status is conceptually problematic for various reasons, not the least of which is determining which classification is appropriate to use.

7. *Code potentially moderating variables rather than using them as exclusion criteria.* Primary studies should be as inclusive as possible and data should be coded rather than serve to exclude participants. For example, defendants diagnosed with Mental Retardation should not be screened out of studies because it hinders investigation of the relationship between cognitive abilities and competence-related abilities; rather, diagnostic classifications and IQ scores should be reported. In addition, the aggregation of studies and comparisons between them is constrained because most researchers do not screen out such defendants.
8. *Perform analyses using all available competency criterions.* Most competency researchers use mental health professionals' decisions as the ultimate criterion of competency status. A concern has traditionally been whether to use the court's decision or that of a mental health professional as the criterion in research, referred to as the "criterion problem" (Golding et al., 1984); however, the "problem" is arguably

more theoretical than practical because the rate of agreement between mental health professionals and the courts with respect to the ultimate competency decision has consistently been found to approach 100% (Zapf, Hubbard, Cooper, Wheeles, & Ronan, 2004) – a reality acknowledged for quite some time (see Vann & Morganroth, 1965). The variation of competency decisions within and between groups of mental health professionals is perhaps a more apropos concern (see Murrie et al., 2008) and competency researchers should conduct statistical analyses using all available competency criterions to investigate potential moderation.[17]

9. *Present data comparatively by competency status in addition to other dichotomies.* Of the 98 excluded studies, 16 met initial inclusion criteria but were excluded because insufficient data were presented. Between-groups analyses were conducted in these studies and data was presented across levels of the categorical variable of interest (e.g., type of charge, sex, diagnosis). Information related to differences between incompetent and competent defendants is lost when data is analyzed and presented in this way; therefore, competency researchers should present data comparatively by competency status regardless of the scope of their research. For instance, a researcher who has investigated differences between male and female defendants referred for competency evaluations should present comparative data based on sex and competency status (even if the presentation of such data is limited to one table).
10. *Incorporate incompetent and competent control groups in malingering studies.* Results from malingered-competency studies can be synthesized with data from other competency studies when the comparative data of competent and incompetent participants is presented in addition to malingering groups (e.g., Gothard et al., 1995;

Jackson et al., 1995). Thus, incorporating incompetent and competent control groups in a malingering study directly improves the quality of the research design and ultimately provides useful information to the field.

11. *Classify defendants as "questionably fit" when possible.* Some studies presented data on a third group – those whose competency was questionable or "grey area" cases. This classification has usually been based on either the decisions of mental health professionals or performance on particular competency assessment measures (e.g., the FIT-R). Data from questionably fit groups were combined with that of unfit groups in the present meta-analysis because too few studies included the classification to investigate meaningful differences; however, it is likely that such differences do exist across levels of particular variables. Researchers should classify defendants as questionably fit when possible; such data can be combined at any point for a comparative analysis of incompetent and competent groups and, perhaps, new information will be gleaned regarding these "grey area" cases.
12. *Present comparative data consistently across staple competency variables.* One of the inherent limitations of a meta-analysis relates to missing data in primary studies. Ten classes of variables pertaining to defendant characteristics should be included in comparative competency research when possible: (i) age; (ii) ethnicity; (iii) sex; (iv) level of education; (v) employment status; (vi) marital status; (vii) psychiatric diagnosis; (viii) psychiatric history; (ix) competency history; and (x) legal history. In addition, nine study-specific variables should be presented: (a) date range of data collection; (b) source of initial competency decision; (c) source of competency decision used for comparison; (d) nature of competent comparison group; (e) names

of and data from traditional and/or competency assessment instruments used; (f) setting of study; (g) sample's country of origin; (h) nature of participant recruitment; and (i) type of sample. The study-level coding form and manual presented in Appendix C and D provide details relevant for the collection and/or coding of the aforementioned variables.

Future Directions

The adjudicative competency literature consists of approximately 200 publications dating back to the early 1960s, but significant gaps in knowledge remain. Three such areas of study to which competency researchers should focus their efforts are: (a) "grey area" cases; (b) defendants' performance on traditional and competency assessment instruments; and (c) case studies.

Psycholegal scholars have recognized the need to focus on defendants whose competency is questionable, or "grey area" cases, for many years. Golding and colleagues (1984) believed doing so would be beneficial towards understanding the construct of competency:

> In exploring the nature of the competency construct, we think that a comparative analysis of different decisional strategies would best occur in a sample with many more unfit and "grey area" cases since these cases are likely to pull for professional differences. (p. 330)

However, few researchers have actually defined questionably competent groups in their studies (e.g., Brown, Felthous, Barratt, Stanford, & Brown, 1994; McDonald et al., 1991; Nussbaum et al., 1998; Rogers et al., 1988) and even when such data has been collected, it was typically combined with that of incompetent groups. Indeed, the present meta-analysis is guilty of such aggregation. Although a dearth of information regarding grey area cases exists as a result, they have been found to differ from incompetent and competent groups on such variables as scores on

the Fitness Interview Test (McDonald et al., 1991). Nevertheless, questionably fit defendants are still thought to be a unique group, distinguishable from their clearly fit and unfit counterparts. Nussbaum and colleagues (1998) characterized such defendants as:

> a relatively heterogeneous group including those [sic] fitness is unable to be determined in a 20 minute screening interview, or alternately, individuals who might meet the technical criteria for fitness, but are acutely mentally ill, in need of treatment, and willing to accept treatment voluntarily during a two week inpatient stay. (p. 51)

It would be beneficial if researchers define questionably competent defendants in the future as a third group to be compared to incompetent and competent defendants. In addition, it may be useful to investigate correctional-psychiatric samples whose competence has not been questioned, such as those used in the MacArthur competency studies (i.e., Hoge et al., 1998; Otto et al., 1998; Poythress et al., 1998). These designs compared three groups: a Hospital Incompetent (HI) group comprised of defendants admitted to forensic psychiatric units after being adjudicated incompetent to proceed; a Jail Treated (JT) group consisting of defendants in jail receiving treatment for mental health problems but presumed competent; and a Jail Unscreened (JU) group of randomly selected jail inmates presumed competent. The JT group was more similar to the HI group than to the JU group on a number of variables (e.g., amount of previous psychiatric hospitalizations, BPRS scores, the MMPI Psychoticism scale, psychiatric diagnoses). Further research is needed in this area, but careful attention must be paid to the reality that inmates with mental illnesses (i.e., Jail Treated) should not necessarily be presumed competent simply because it has never been questioned. There are a number of possible reasons as to why a defendant's competence was not questioned during the pretrial and/or trial stages, not the least of which relates to the fact that most defendants do not actually stand trial. That is, most

defendants engage in a plea bargaining process, which may not elicit concerns regarding their competence-related abilities. As such, prison mental health units undoubtedly consist of a certain number of inmates who would have been incompetent to stand trial if they were evaluated.

A second major line of future research relates to the utility of traditional and competency assessment instruments. Numerous studies have investigated the psychometric properties of the 12 competency instruments developed since the 1970s; however, few have presented comparative data from incompetent and competent defendant groups. In fact, psychometric investigations can be conducted with single sample designs (i.e., incompetent defendants only) and psychometric analyses are often geared toward examining the instrument's performance and/or characteristics (e.g., factor analysis, discriminant function analysis, criterion validity, test-retest reliability) rather than the differential performance between two groups. Determination of a measure's psychometric properties is necessary, but comparing the performance of incompetent and competent defendants on the measure is essential in the investigation of its utility. Comparative performance on traditional measures must be researched more extensively before conclusions can be made with respect to the extent of their utility (or lack thereof) in competency evaluations. An attempt needs to be made to examine traditional measures more holistically rather than based on a few indices or scales. Intellectual assessment measures (e.g., WAIS-IV) and personality assessment instruments (e.g., MMPI-2) tend to be fairly robust measures and, as such, opinions regarding their utility should not be rooted in data derived from only a few studies on a few scales. Research examining group differences on cognitive or neuropsychological measures is perhaps the most promising next step in this area, as adjudicative competency is inherently a cognitive construct.

A third type of research in need of increased attention in the adjudicative competency arena is that of case studies. Case studies have served as a cornerstone for many fields of psychological study and can be as scientific and informative as other types of research methods when conducted properly (e.g., B.F. Skinner's research on operant conditioning). Single-case research designs avoid limitations associated with data aggregation and enables investigation into intrasubject variation and systematic replication (see Hilliard, 1993 for a review). Strupp (1981) argued that most of what was learned about psychotherapy in the 20th century came from "astute and creative clinical observations" (p. 216) and that scientific knowledge is dependent on the ability of researchers to learn about "clinical realities and stay in touch with clinical phenomena" and practicing therapists' ability to develop "a thorough grasp of basic principles of scientific research" (p. 217). Publication of single-case research can help bridge the gap between competency research and practice by facilitating the aforementioned ideals, thereby fostering progress in the field. Unfortunately, few case studies have been published in the competency arena. Bendt, Balcanoff, and Tragellis (1973) presented four competency case studies; however, only one paragraph was devoted to each case. Two recently published manuscripts are more thorough and are useful templates for those authoring competency case studies: Stoops and colleagues (2007) presented a case of a defendant found incompetent based on an intellectual disability diagnosis who was undergoing competency restoration and Pinals, Tillbrook, and Mumley (2006) presented four cases to demonstrate the utility of the MacCAT-CA in competency evaluations.[18]

Conclusion

The present study is a synthesis of 50 years of comparative competency research and represents the current state of knowledge with respect to differences between incompetent and competent defendants on various demographic, psycholegal, criminological, and clinical variables. The major findings of this meta-analysis are that defendants diagnosed with a Psychotic Disorder, those who are unemployed, and those with a psychiatric hospitalization history are most likely to be found incompetent to stand trial. In addition, there is a significantly greater discrepancy in scores between competent and incompetent defendants on competency assessment instruments as compared to traditional psychological measures; however, this finding is based on limited data. Although these findings have been incorporated into competency practice recommendations and standards over the years, many questions remain. Perhaps the greatest strength of this meta-analysis is its stimulation of such questions, which will hopefully inspire research and commentary in this area.

There is room for improvement and growth in all areas of study, and the competency arena is no different; however, that reality should not overshadow the significant advancements in this field since the 1960s. The quality of a meta-analysis is, in part, dependent on the quality of the primary research literature, which is largely dependent on our ability to learn from and improve upon earlier studies. Sustained developments in the adjudicative competency arena are reliant on the continued commitment of authors to conceptualize their work as part of a field of research rather than as isolated investigations (e.g., consideration of if and how their data could be coded by the next meta-analyst) – an integral step toward bridging gaps between research, practice, and policy.

Notes

[1] The terms competency to stand trial, adjudicative competency, and fitness to stand trial are used interchangeably throughout the manuscript.

[2] Nicholson and Kugler (1991) reported the inclusion of 30 studies in their meta-analysis, and therefore, it has been cited as such over the years; however, only 27 independent studies were actually synthesized.

[3] In *Jackson*, the Supreme Court held that indefinitely committing a criminal defendant for an indefinite period of time based solely on a finding of incompetency was unconstitutional.

[4] The term "competent defendants" is used throughout this paper. While the majority of studies conducted in this area have used referred defendants as their sample, a few studies have used competent participants which have included psychiatric patients or inmates whose competence was never questioned.

[5] *Type of competent group* was developed for the purposes of this study. Five types of competent comparison groups were coded in the present study: referred defendants, purely competent defendants whose competency was never in question (e.g., inmates), those restored to competency, defendants who were initially deemed incompetent but then classified as competent by the researchers (i.e., study-competent), and a mixed group. *Recruitment* type was coded as either Archival/Retrospective (i.e., using data that has been previously collected, usually for clinical purposes) or Prospective (i.e., active recruitment of participants for the research study).

[6] Meta F tests and regression analyses were calculated via SPSS macros written by Mark Lipsey and David Wilson (see Lipsey & Wilson, 2001, pp. 208-220).

[7] The statistics presented in Tables 3 and 4 are not weighted estimates because they equal actual percentages. That is, raw frequency data was collected; therefore, no "mean of mean" or "average of percentages" statistics were calculated. Rather, the percentages in the tables were calculated from frequencies and, therefore, do not need to be weighted because the sample size is the Total N. For example, the % Caucasian variable in Table 3 was calculated by counting the number of Caucasian people in the 22 studies that presented such data, and then dividing that number by the total sample size derived from those 22 studies.

[8] Five studies used by Nicholson and Kugler (1991) were not included in the present meta-analysis because they were deemed to not meet inclusion criteria (see Appendix B).

[9] Ames Robey (1965) is credited with developing the first formal measure of competency – a checklist for psychiatrists. Robey's checklist consisted of three sections: Comprehension of Court Proceedings, Ability to Advise Counsel, and Susceptibility to Decompensation while awaiting or standing trial. Each section consisted of eight, seven, and five areas to explore, respectively. These sections were rated either, "OK", "Mental Illness", or "Intellectual Deficiency." Mental Illness included an evaluation of cognition, orientation, apperception and judgment, and (generally) excluded character disorders. Intellectual Deficiency referred to obtaining a Wechsler Adult Intelligence Scale IQ score below 60. Robey's checklist has never

been systematically studied (Roesch & Golding, 1980). Bukatman and colleagues (1971) followed with a series of interview questions designed to assess understanding of the current situation, as well as cooperation and participation in one's own defense. Although these checklists and interview questions are rarely used today, they were instrumental in providing a foundation for the assessment instruments that followed. Following these early efforts at developing checklists/interview questions, forensic psychologists began to develop instruments that were more psychometrically sound, and therefore, more clinically useful.

[10] Two studies (Gothard, Rogers, & Sewell, 1995; Otto et al., 1998; Rogers, Sewell, Grandjean, & Vitacco, 2002) presented data on specific scales of competency measures. Otto and colleagues (1998) reported data on the Understanding, Reasoning, and Appreciation scales of the MacCAT-CA, and Gothard et al. (1995) and Rogers et al. (2002) presented data on the GCCT's Atypical Presentation Scale (APS) developed by Gothard and colleagues in 1995.

[11] Odds Ratios are presented for continuous data whereby standardized mean difference statistics were calculated to maintain continuity throughout the paper (Tables 24, 27-32). The formula provided by Borenstein (2009) was used to convert the *d* statistic to an odds ratio, whereas the odds ratio and its variance are:

$$[\ln(o)] = \frac{\pi d}{\sqrt{3}}$$

$$V_{[\ln(o)]} = \frac{\pi^2 v_d}{3}$$

[12] Nicholson and Kugler (1991) presented effect sizes related to "IQ score" without mentioning which intelligence measure(s) were used in the calculations.

[13] A number of landmark Supreme Court cases have been held since the previous meta-analysis was conducted (e.g., *Godinez*); however, the potential moderator variable, period of data collection, was not formally tested in the present meta-analysis because of practical statistical difficulties related to publication date overlap.

[14] Averaging group means is statistically correct, but doing so with standard deviations is not; therefore, a pooled standard deviation statistic was calculated to represent the "average" of two or more standard deviations.

[15] Two studies have subsequently included the revised version of the IFI (i.e., IFI-R) into their designs (Barnard et al, 1991, 1992; Holmes, 1991).

[16] Fifty-two (n=52) defendants were evaluated with the IFI alone while the remaining defendants were also administered the CAI, but group differences were not explored for potentially moderating methodological effects.

[17] Nicholson and Kugler (1991) reported data on mental health professionals as a whole, but they believed "correlates of competency might vary as a function of the source of the competency determination" (p. 367). Sources of the competency criterion used across studies was delineated

in the present meta-analysis to include mostly psychiatrist(s) decisions (29%); followed by a combination of various individual decisions, or mixed, (25%); a mental health professional "team" (21%); the court (19%); psychologist(s) (3%); and, not reported (3%). Source of competency decision was not found to be a significant moderator in the present meta-analysis; however, limited data was available in this regard.

[18] Authors may also find it useful to view the journal *Clinical Case Studies* for additional case study references.

* Used by Nicholson & Kugler (1991)
† Not used in N & K, but eligible
± Dissertation

Appendix A

Reference List for Included Studies

†Aubrey, M. (1988). Characteristics of competency referral defendants and nonreferred criminal defendants. *Journal of Psychiatry & Law, 16*, 233-245.

Bagby, R. M., Nicholson, R. A., Rogers, R., & Nussbaum, D. (1992). Domains of competency to stand trial: A factor analytic study. *Law and Human Behavior, 16*, 491-507.

Bittman, B. J., & Convit, A. (1993). Competency, civil commitment, and the dangerousness of the mentally ill. *Journal of Forensic Sciences, 38*, 1460-1466.

*Bluestone, H., & Mellela, J. (1978). A study of criminal defendants referred for competency to stand trial in New York City. *Bulletin of the American Academy of Psychiatry & the Law, 7*, 166-178.

Brown, D. L., Felthous, A. R., Barratt, E. S., & Stanford, M. (1994). The incompetent defendant: Support systems help avoid future legal problems. *Journal of Forensic Sciences, 39*, 1057-1068.

Caldwell, R. M., Mandracchia, S. A., Ross, S. A., & Silver, N. C. (2003). Competency to stand trial and criminal responsibility: An examination of racial and gender differences among African American and Caucasian pretrial defendants. *American Journal of Forensic Psychology, 21*, 5-19.

Chaimowitz, G. A., & Ferencz, J. (1999). Cost savings associated with fitness-to-stand-trial assessments in detention centres: A pilot program. *The Canadian Journal of Psychiatry, 44*, 808-810.

±Chatfield, G. G. (1998). *Repeat competence to stand trial evaluations: A marker for severe and persistent mental illness.* Dissertation Abstracts International, 58 (11-A), 4443.

Chellsen, J. A. (1986). Trail competency among mentally retarded offenders: Assessment techniques and related considerations. *Journal of Psychiatry and Law, 14*, 177-185.

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.

*Cooke, G. (1969). The court study unit: Patient characteristics and differences between patients judged competent and incompetent. *Journal of Clinical Psychology, 25*, 140-143.

*Cooke, G., Pogany., E., & Johnston, N. G. (1974). A comparison of blacks and whites committed for evaluation of competency to stand trial on criminal charges. *Journal of Psychiatry and Law, 2*, 319-337.

Cooper, V. G., & Zapf, P. A. (2003). Predictor variables in competency to stand trial decisions. *Law and Human Behavior, 27*, 423-436.

Cox, M. L., & Zapf, P. A. (2004). An investigation of discrepancies between mental health professionals and the courts in decisions about competency. *Law & Psychology Review, 28*, 109-132.

Crocker, A. G., Favreau, O. E., & Caulet, M. (2002). Gender and fitness to stand trial: A 5-year review of remands in Québec. *International Journal of Law and Psychiatry, 25*, 67-84.

†Fitzgerald, J. F., Peszke, M. A., & Goodwin, R. C. (1978). Competency evaluations in Connecticut. *Hospital & Community Psychiatry, 29*, 450-453.

*Golding, S. L. Roesch, R., & Schreiber, J. (1984). Assessment and conceptualization of competency to stand trial: Preliminary data on the Interdisciplinary Fitness Interview. *Law & Human Behavior, 8*, 321-334.

†Goldstein, R. L. (1973). "The fitness factory," part I: The psychiatrist's role in determining competency. *The American Journal of Psychiatry, 130,* 1144-1147.

±Gothard, S. (1994). *The detection of malingering in competency to stand trial evaluations.* Dissertation Abstracts International, 55 (2-B), 591.

Gothard, S., Rogers, R., & Sewell, K. W. (1995). Feigning incompetency to stand trial: An investigation of the Georgia Court Competency Test. *Law and Human Behavior, 19,* 363-373.

Gothard, S., Viglione, D. J., Meloy, J. R., & Sherman, M. (1995). Detection of malingering in competency to stand trial evaluations. *Law and Human Behavior, 19,* 493-505.

±Grandjean, N. R. (2004). *Neuropsychological predictors of incompetency to stand trial in defendants referred for competency restoration.* Dissertation Abstracts International, 66 (7-B), 3947.

†Haney, C. A., & Michielutte, R. (1968). Selective factors operating in the adjudication of incompetency. *Journal of Health and Social Behavior, 9,* 233-242.

Haney, C. A., & Miller, K. S. (1970). Definitional factors in mental incompetency. *Sociology & Social Research, 54,* 520-532.

Haney, C. A., Miller, K. S., & Michielutte, R. (1969). The interaction of petitioner and deviant social characteristics in the adjudication of incompetency. *Sociometry, 32,* 182-193.

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.

*Heller, M. S., Traylor., W. H., Ehrlich, S. M., & Lester, D. (1981). Intelligence, psychosis and competency to stand trial. *Bulletin of the American Academy of Psychiatry & the Law, 9,* 267-274.

*Heller, M. S., Traylor., W. H., Ehrlich, S. M., & Lester, D. (1983). Competent and incompetent defendants referred to a court psychiatric clinic: A clinical comparison. *Journal of Forensic Sciences, 28,* 415-422.

Hoge, S. K., Bonnie, R. J., Poythress, N., Monahan, J., Eisenberg, M., & Feucht-Haviar, T. (1997a). The MacArthur adjudicative competence study: Development and validation of a research instrument. *Law and Human Behavior, 21*, 141-179.

Hoge, S. K., Poythress, N., Bonnie, R., Eisenberg, M., Monahan, J., Feucht-Haviar, T., & Oberlander, L. (1996). Mentally ill and non-mentally ill defendants' abilities to understand information relevant to adjudication: A preliminary study. *Bulletin of the American Academy of Psychiatry & the Law, 24*, 187-197.

Hubbard, K. L., Zapf, P. A., & Ronan, K. A. (2003). Competency restoration: An examination of the differences between defendants predicted restorable and not restorable to competency. *Law and Human Behavior, 27*, 127-139.

Jackson, R. L., Rogers, R., & Sewell, K. W. (2005). Forensic Applications of the Miller Forensic Assessment of Symptoms Test (MFAST): Screening for Feigned Disorders in Competency to Stand Trial Evaluations. *Law and Human Behavior, 29*, 199-210.

*Johnson, W. G. Nicholson., R. A., Service, N. M. (1990). The relationship of competency to stand trial and criminal responsibility. *Criminal Justice and Behaviour, 17*, 169.

*Laczko, A. L. James, J. F., & Alltop, L. B. (1970). A study of four hundred and thirty-five court-referred cases. *Journal of Forensic Sciences, 15*, 311-323.

±Lesser, F. I. (1990). *Formulating a prediction model for regaining competency to stand trial.* Dissertation Abstracts International, 51 (2-B), 993.

*Lipsitt, P. D., Lelos., D., & McGarry, A. L. (1971). Competency for trial: A screening instrument. *American Journal of Psychiatry, 128*, 137-141.

Martell, D. A., Rosner, R., & Harmon, R. B. (1994). Homeless mentally disordered defendants: Competency to stand trial and mental status findings. *Bulletin of the American Academy of Psychiatry & the Law, 22*, 289-295.

±Matthews, B. (1999). *Competency to stand trial and insight in mentally ill defendants.* Dissertation Abstracts International, 59 (8-B), 4473.

*Maxson, L. S., & Neuringer., C. (1970). Evaluating legal competency. *Journal of Genetic Psychology, 117*, 267-273.

McDonald, D. A., Nussbaum, D. S., & Bagby, R. M. (1991). Reliability, validity and utility of the Fitness Interview Test. *The Canadian Journal of Psychiatry, 36*, 480-484.

Mosley, D., Thyer, B. A., & Larrison, C. (2001). Development and preliminary validation of the Mosley Forensic Competency Scale. *Journal of Human Behavior in the Social Environment, 4*, 41-48.

±Mueller, C. L. (2007). *An examination of predictor variables for competency to stand trial determinations in a multicultural setting.* Dissertation Abstracts International, 68 (2-B), 1315.

Nestor, P. G., Daggett, D., Haycock, J., & Price, M. (1999). Competence to stand trial: A neuropsychological inquiry. *Law and Human Behavior, 23*, 397-412.

†Nicholson, R. A. (1988). Validation of a brief form of the Competency Screening Test. *Journal of Clinical Psychology, 44*, 87-90.

*Nicholson, R. A., Briggs, S. R., & Robertson, H. C. (1988). Instruments for assessing competency to stand trial: How do they work? *Professional Psychology: Research and Practice, 19,* 383-394.

*Nicholson, R. A., Robertson, H. C., Johnson, W. G., & Jensen, G. (1988). A comparison of instruments for assessing competency to stand trial. *Law and Human Behavior, 12*, 313-321.

*Nottingham, E. J., & Mattson., R. E. (1981). A validation study of the competency screening test. *Law and Human Behavior, 5*, 329-336.

Nussbaum, D., Mamak, M., Tremblay, H., Wright, P., & Callaghan, J. (1998). The METFORS Fitness Questionnaire (MFQ): A self-report measure for screening competency to stand trial. *American Journal of Forensic Psychology, 16*, 41-65.

Ohayon, M. M., Crocker, A., St-Onge, B., & Caulet, M. (1998). Fitness, responsibility, and judicially ordered assessments. *The Canadian Journal of Psychiatry*, *43*, 491-495.

Otto, R. K., Poythress, N. G., Nicholson, R. A., Edens, J. F., Monahan, J., Bonnie, R. J., Hoge, S. K., & Eisenberg, M. (1998). Psychometric properties of the MacArthur Competence Assessment Tool-Criminal Adjudication. *Psychological Assessment, 10*, 435-443.

*Pfeiffer, E., Einstein, R. B., & Dabbs, E. G. (1967). Mental competency evaluation for the federal courts: I. Methods and results. *Journal of Nervous & Mental Disease, 144*, 320-328.

Poythress, N. G., Hoge, S. K., Bonnie, R. J., Monahan, J., Eisenberg, M., & Feucht-Haviar, T. (1998). The competence-related abilities of women criminal defendants. *Journal of the American Academy of Psychiatry and the Law, 26*, 215-222.

*Randolph, J. J., Hicks., T., & Mason, D. (1981). The Competency Screening Test: A replication and extension. *Criminal Justice and Behavior, 8,* 471-482.

*Randolph, J. J., Hicks., T., Mason, D., & Cuneo, D. J. (1982). The Competency Screening Test: A validation study in Cook County, Illinois. *Criminal Justice & Behavior, 9,* 495-500.

*Reich, J., & Wells, J. (1985). Psychiatric diagnosis and competency to stand trial. *Comprehensive Psychiatry, 26,* 421-432.

±Riley, S. E. (1997). *Gender and competency adjudications in a southwestern urban court system.* Dissertation Abstracts International, 58 (2-A), 0588.

Riley, S. E. (1998). Competency to stand trial adjudication: A comparison of female and male defendants. *Journal of the American Academy of Psychiatry and the Law, 26*, 223-240.

Robertson, R. G., Gupton, T., McCabe, S. B., & Bankier, R. G. (1997). Clinical and demographic variables related to 'fitness to stand trial' assessments in Manitoba. *The Canadian Journal of Psychiatry, 42,* 191-195.

±Roesch, R. M. (1977). Competency to stand trial: An analysis of legal/mental health issues and procedures and a proposal for change. Dissertation Abstracts International, 38 (6-B), 2882.

*Roesch, R. (1978a). A brief, immediate screening interview to determine competency to stand trial: A feasibility study. *Criminal Justice & Behavior, 5,* 241-248.

*Roesch, R. (1978b). Competency to stand trial and court outcome. *Criminal Justice Review, 3,* 45-56.

*Roesch, R. (1979). Determining competency to stand trial: An examination of evaluation procedures in an institutional setting. *Journal of Consulting & Clinical Psychology, 47*, 542-550.

*Roesch, R., Eaves, D., Sollner, R., Normandin, M., & Glackman, W. (1981). Evaluating fitness to stand trial: a comparative analysis of fit and unfit defendants. *International Journal of Law & Psychiatry, 4*, 145-157.

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

Roesch, R., Ogloff, J. R. P., Hart, S. D., Dempster, R. J., Zapf, P. A., & Whittemore, K. E. (1997). The impact of Canadian criminal code changes on remands and assessments of fitness to stand trial and criminal responsibility in British Columbia. *The Canadian Journal of Psychiatry, 42*, 509-514.

*Rogers, R., Gillis, J. R., McMain, S., & Dickens, S. E. (1988). Fitness evaluations: A retrospective study of clinical, criminal, and sociodemographic characteristics. *Canadian Journal of Behavioural Science Revue canadienne des Sciences du comportement, 20*, 192-200.

Rosenfeld, B., & Ritchie, K. (1998). Competence to stand trial: Clinician reliability and the role of offense severity. *Journal of Forensic Sciences, 43*, 151-157.

Rosenfeld, B., & Wall, A. (1998). Psychopathology and competence to stand trial. *Criminal Justice and Behavior, 25*, 443-462.

±Sachsenmaier, S. J. (1991). *The relationship between traditional psychometric test scores and the determination of criminal responsibility and competency to stand trial.* Dissertation Abstracts International, 53 (1-B), 566.

Schreiber, J. (1982). Professional judgment in the assessment of competency to stand trial: Report of an evaluation study. *International Journal of Law and Psychiatry, 5*, 331-340.

Schreiber, J., Roesch, R., & Golding, S. L. (1987). An evaluation of procedures for assessing competency to stand trial. *Bulletin of the American Academy of Psychiatry and the Law, 15*, 187-203.

*Shatin, L., & Brodsky, S. H. (1979). Competency for trial: The Competency Screening Test in an urban hospital forensic unit. *The Mount Sinai Journal of Medicine, 46*, 131-134.

±Shields, C. L. (2005). *Competence to stand trial: Cognitive-related abilities and other factors.* Dissertation Abstracts International, 66 (2-B), 1186.

*Sikorski, C. R., & Benedek., E. P. (1977). Competency to stand trial: The female offender. *Journal of the American Medical Women's Association, 32*, 150-154.

†Simon, M. J. (1987). Use of the Proverbs Test in the assessment of competency to stand trial. *Psychological Reports, 60*, 1166.

Stafford, K. P., & Wygant, D. B. (2005). The role of competency to stand trial in mental health courts. *Behavioral Sciences & the Law, 23*, 245-258.

±Tillbrook, C. E. (2001). *Competency to proceed: A comparative appraisal of approaches to assessment.* Dissertation Abstracts International, 61 (9-B), 5009.

Vernon, M., Steinberg, A. G., & Montoya, L. A. (1999). Deaf murderers: Clinical and forensic issues. *Behavioral Sciences & the Law, 17*, 495-516.

Viljoen, J. L., Zapf, P. A., & Roesch, R. (2003). Diagnosis, current symptomatology, and the ability to stand trial. *Journal of Forensic Psychology Practice, 3*, 23-37.

Warren, J. I., Fitch, W. L., Dietz, P. E., & Rosenfeld, B. D. (1991). Criminal offense, psychiatric diagnosis, and psycholegal opinion: An analysis of 894 pretrial referrals. *Bulletin of the American Academy of Psychiatry & the Law, 19*, 63-69.

Warren, J. I., Murrie, D. C., Stejskal, W., Colwell, L. H., Morris, J., Chauhan, P., et al. (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.

Warren, J. I., Rosenfeld, B., & Fitch, W. L. (1994). Beyond competence and sanity: The influence of pretrial evaluation on case disposition. *Bulletin of the American Academy of Psychiatry & the Law, 22*, 379-388.

*Webster, C. D., Menzies., R. J., & Jackson, M. A. (1982). Clinical assessment before trial. Toronto, Ontario, Canada: Butterworths.

Whittemore, K. E., Ogloff, J. R. P., & Roesch, R. (1997). An investigation of competency to participate in legal proceedings in Canada. *The Canadian Journal of Psychiatry, 42*, 869-875.

*Wildman, R. W., White, P. A., & Brandenburg, C. E. (1990). The Georgia Court Competency Test: The base-rate problem. *Perceptual and Motor Skills, 70*, 1055-1058.

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.

Zapf, P. A., & Roesch, R. (1997). Assessing fitness to stand trial: A comparison of institution-based evaluations and a brief screening interview. *Canadian Journal of Community Mental Health, 16*, 53-66.

Zapf, P. A., & Roesch, R. (1998). Fitness to stand trial: Characteristics of remands since the 1992 Criminal Code amendments. *The Canadian Journal of Psychiatry, 43*, 287-293.

Zapf, P. A., Roesch, R., & Viljoen, J. L. (2001). Assessing Fitness to Stand Trial: The Utility of the Fitness Interview Test (Revised Edition). *The Canadian Journal of Psychiatry, 46*, 426-432.

* Used by Nicholson & Kugler (1991)
† Not used in N & K, but eligible
± Dissertation

Appendix B

Reference List for Excluded Studies

Addington, D. E., & Holley, H. L. (1987). Pre-trial assessment of schizophrenics on remand. *The Canadian Journal of Psychiatry, 32*, 773-776

±Amadeo, A. M. (2005). *The development of a 'restoration to competence' program for patients found incompetent to stand trial.* Dissertation Abstracts International, 66 (3-B), 1706.

±Anderson, S. D. (1999). *The effect of competency restoration training on defendants with mental retardation found not competent to proceed.* Dissertation Abstracts International, 59 (7-B), 3767.

Anderson, S. D., & Hewitt, J. (2002). The effect of competency restoration training on defendants with mental retardation found not competent to proceed. *Law and Human Behavior, 26*, 343-352.

Barnard, G. W., Nicholson, R. A., Hankins, G. C., & Raisani, K. K. (1992). Itemmetric and scale analysis of a new computer-assisted competency assessment instrument (CADCOMP). *Behavioral Sciences & the Law, 10*, 419-435.

Barnard, G. W., Thompson, J. W., Freeman, W. C., & Robbins, L. (1991). Competency to stand trial: Description and initial evaluation of a new computer-assisted assessment tool (CADCOMP). *Bulletin of the American Academy of Psychiatry & the Law, 19*, 367-381.

Beckham, J. C., Annis, L. V., & Bein, M. F. (1986). Don't pass go: Predicting who returns from court as remaining incompetent for trial. *Criminal Justice and Behavior, 13*, 99-109.

±Bennett, A. D. (2006). *The measurement of adjudicative competence: A comparison of three types of competencies in a sample of individuals with mental retardation.* Dissertation Abstracts International, 67 (1-B), 529.

Bennett, G. T., & Kish, G. R. (1990). Incompetency to stand trial: Treatment unaffected by demographic variables. *Journal of Forensic Sciences, 35*, 403-412.

Bendt, R. H., Balcanoff, E. J., & Tragellis, G. S. (1973). Incompetency to stand trial: Is psychiatry necessary? *The American Journal of Psychiatry, 130*, 1288-1289.

Berman, L. M., & Osborne, Y. H. (1987). Attorneys' referrals for competency to stand trial evaluations: Comparisons of referred and nonreferred clients. *Behavioral Sciences & the Law, 5*, 373-380.

±Bertman, L. J. (2000). *Effect of an individualized treatment protocol on competency restoration in pretrial forensic inpatients.* Dissertation Abstracts International, 60 (11-B), 5763.

Bertman, L. J., Thompson, J. W. Jr., Waters, W. F., Estupinan-Kane, L., Martin, J. A., & Russell, L. (2003). Effect of an Individualized Treatment Protocol on Restoration of Competency in Pretrial Forensic Inpatients. *Journal of the American Academy of Psychiatry and the Law, 31*, 27-35.

Bertsch, J. M., Younglove, J. A., & Kerr, M. G. (2002). A pilot study of the Porterville Developmental Center's court competency training program. *Criminal Justice Policy Review, 13*, 65-77.

Buigas, R. A. (1996). *Development of scales for the Computer-Assisted Determination of Competency.* Dissertation Abstracts International, 56 (11-B), 6454.

Carbonell, J. L., Heilbrun, K., & Friedman, F. L. (1992). Predicting who will regain trial competency: Initial promise unfulfilled. *Forensic Reports, 5*, 67-76.

*Chellsen, J. A. (1986). Trail competency among mentally retarded offenders: Assessment techniques and related considerations. *Journal of Psychiatry and Law, 14*, 177-185.

Cooke, G., Johnston, N., & Pogany, E. (1973). Factors affecting referral to determine competency to stand trial. *American Journal of Psychiatry, 130*, 870-875.

Cuneo, D. J., & Brelje, T. B. (1984). Predicting probability of attaining fitness to stand trial. *Psychological Reports, 55*, 35-39.

Daniel, A. E., Beck, N. C., Herath, A., Schmitz, M., & Menninger, K. (1984). Factors correlated with psychiatric recommendations of incompetency and insanity. *Journal of Psychiatry & Law, 12*, 527-544.

Edens, J. F., Poythress, N. G., Nicholson, R. A., & Otto, R. K. (1999). Effects of state organizational structure and forensic examiner training on pretrial competence assessments. *Journal of Behavioral Health Services & Research, 26*, 140-150.

Everington, C. (1989). Demographic variables associated with competence to stand trial referral and evaluation of criminal defendants with mental retardation. *Journal of Psychiatry & Law, 17*, 627-640.

Everington, C. T. (1990). The Competence Assessment for Standing Trial for Defendants with Mental Retardation (CAST-MR): A validation study. *Criminal Justice and Behavior, 17*, 147-168.

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.

Everington, C., Notario-Smull, H., & Horton, M. L. (2007). Can defendants with mental retardation successfully fake their performance on a test of competence to stand trial? *Behavioral Sciences & the Law, 25*, 545-560.

Exner, J. E., Jr. (1980). Diagnosis versus description in competency issues. In F. Wright, C. Bahn, & R. W. Rieber (Eds). *Forensic psychology and psychiatry* (pp. 20-26) –Annals of the New York Academy of Sciences, Vol 347.

±Felchlia, M. A. (1992). *Construct validity of the Competency Screening Test.* Dissertation Abstracts International, 53 (1-B), 604.

±Ficken, C. T. (2004). *Using demographic and clinical variables to predict the length of stay of 'Incompetent to Stand Trial' patients.* Dissertation Abstracts International, 64 (7-B), 3220.

Fontaine, P. A. (1985). The case dispositions of temporarily trial-incompetent criminal defendants. *Journal of Psychiatry & Law, 13,* 435-448.

±Gabel, J. E. (2007). *Consistency of screening measures of malingering among pre-trial adjudicative competence defendants.* Dissertation Abstracts International, 67 (9-B), 5400.

±Gannon, J. L. (1990). *Validation of the Competency Assessment Instrument and elements of competency to stand trial.* Dissertation Abstracts International, 50 (8-B), 3675.

Gold, L. H. (1969). A statistical review of court-appointed psychiatric examinations. *Journal of Forensic Sciences, 14,* 294-301.

Gold, L. H. (1973). Discovery of mental illness and mental defect among offenders. *Journal of Forensic Sciences, 18,* 125-129.

Greenland, C., & Rosenblatt, E. (1972). Remands for psychiatric examination in Ontario, 1969-1970. *The Canadian Psychiatric Association Journal, 17,* 397-401.

Heilbrun, K., Bennett, W. S., Evans, J. H., & Offutt, R. A. (1992). Assessing treatability in mentally disordered offenders: Strategies for improving reliability. *Forensic Reports, 5,* 85-96.

Herbel, B. L., & Stelmach, H. (2007). Involuntary medication treatment for competency restoration of 22 defendants with delusional disorder. *Journal of the American Academy of Psychiatry and the Law, 35*, 47-59.

Ho, T. (1999). Examination of racial disparity in competency to stand trial between White and African American retarded defendants. *Journal of Black Studies, 29*, 771-789.

Ho, T., & Henderson, B. B. (1998). Relationship of psychological, demographic, and legal variables to court decisions of competency to stand trial among mentally retarded criminal defendants. *Journal of Criminal Justice, 26*, 307-320.

Hochstedler, E. (1987). Twice-cursed? The mentally disordered criminal defendant. *Criminal Justice and Behavior, 14*, 251-267.

Hoge, S. K., Poythress, N., Bonnie, R. J., Monahan, J., Eisenberg, M., & Feucht-Haviar, T. (1997b). The MacArthur adjudicative competence study: Diagnosis, psychopathology, and competence-related abilities. *Behavioral Sciences & the Law, 15*, 329-345.

±Holmes, D. A. (1991). *Validation of the IFI with hospital staff judgments of competency to proceed to trial.* Dissertation Abstracts International, 52 (6-B), 3339.

Hubbard, K. L., & Zapf, P. A. (2003). The role of demographic, criminal, and psychiatric variables in examiners' predictions of restorability to competency to stand trial. *International Journal of Forensic Mental Health, 2*, 145-155.

Jacobs, M. S., Ryba, N. L., & Zapf, P. A. (2008). Competence-related abilities and psychiatric symptoms: An analysis of the underlying structure and correlates of the MacCAT-CA and the BPRS. *Law and Human Behavior, 32,* 64-77.

±Jones, B. L. (1995). *Factors related to the competency to proceed to trial of defendants who are mentally retarded.* Dissertation Abstracts International, 55 (8-B), 3565.

±Kennedy, J. D. (2008). Diagnosis, instant offense, substance history, and trial competency: An explorative study. Dissertation Abstracts International, 68 (10-B), 6968.

Ladds, B. (1995). Involuntary electro-convulsive therapy to restore competency to stand trial: A five year study in New York State. *Journal of Forensic Sciences, 40,* 183-187.

Ladds, B., Convit, A., Zito, J., & Vitrai, J. (1993). The disposition of criminal charges after involuntary medication to restore competency to stand trial. *Journal of Forensic Sciences, 38*, 1442-1459.

Lamb, H. R. (1987). Incompetency to stand trial: Appropriateness and outcome. *Archives of General Psychiatry, 44*, 754-758.

Lewis, C. F., Fields, C., & Rainey, E. (2006). A Study of geriatric forensic evaluees: Who are the violent elderly? *Journal of the American Academy of Psychiatry and the Law, 34*, 324-332.

Manguno-Mire, G. M., Thompson, J. W., Shore, J. H., Croy, C. D., Artecona, J. F., & Pickering, J. W. (2007). The use of telemedicine to evaluate competency to stand trial: A preliminary randomized controlled study. *Journal* of the *American Academy* of *Psychiatry* and the *Law*, *35,* 481-489.

McGarry, A. L. (1965). Competency for trial and due process via the state hospital. *American Journal of Psychiatry, 122,* 623-631.

McKee, G. R. (1998). Competency to stand trial in preadjudicatory juveniles and adults. *Journal of the American Academy of Psychiatry and the Law, 26*, 89-99.

*Menzies, R. J., Jackson, M. A., & Glasberg, R. E. (1982). The nature and consequences of forensic psychiatric decision-making. *Canadian Journal of Psychiatry, 27*, 463-470.

*Menzies, R. J., Webster, C. D., Roesch, R., Jensen, F. A. S., & Eaves, D. (1983). The Fitness Interview Test: A semi-structured instrument for assessing competency to stand trial, with a proposal for its implementation. *Medicine & Law, 3,* 151-162.

Miller, H. A. (2004). Examining the Use of the M-FAST With Criminal Defendants Incompetent to Stand Trial. *International Journal of Offender Therapy and Comparative Criminology, 48,* 268-280.

Mossman, D. (2007). Predicting restorability of incompetent criminal defendants. *Journal of the American Academy of Psychiatry and the Law, 35,* 34-43.

Mowbray, C. T. (1979). A study of patients treated as incompetent to stand trial. *Social Psychiatry, 14,* 31-39.

Nicholson, R. A., Barnard, G. W., Robbins, L., & Hankins, G. (1994). Predicting treatment outcome for incompetent defendants. *Bulletin of the American Academy of Psychiatry & the Law, 22,* 367-377.

Nicholson, R. A., & Johnson, W. G. (1991). Prediction of competency to stand trial: Contribution of demographics, type of offense, clinical characteristics, and psycholegal ability. *International Journal of Law and Psychiatry, 14,* 287-297.

Nicholson, R. A., & McNulty, J. L. (1992). Outcome of hospitalization for defendants found incompetent to stand trial. *Behavioral Sciences & the Law, 10,* 371-383.

Nussbaum, D., & Amaral, R. (2001). Internal reliability and item analysis of the METFORS Fitness Questionnaire. *American Journal of Forensic Psychology, 19,* 65-74.

Paramesh, C. R. (1987). The Competency Screening Test: Application to a state maximum security hospital population. *American Journal of Forensic Psychology, 5,* 11-15.

±Patterson, H. R. (2005). *Length of stay in forensic inpatients adjudicated as incompetent to stand trial.* Dissertation Abstracts International, 65 (12-B), 6668.

±Peacock, M. A. (2005). *Mental illness as a mediator of competent behavior in persons with mental retardation.* Dissertation Abstracts International, 66 (2-B), 1181.

±Pierrel, A. S. (1986). *Competency to stand trial and the mentally retarded defendant.* Dissertation Abstracts International, 47 (3-B), 1282.

Quinsey, V. L., Pruesse, M., & Fernley, R. (1975). A follow-up of patients found unfit to stand trial or not guilty because of insanity. *The Canadian Psychiatric Association Journal, 20,* 461-467.

Redding, R. E. (1997). Depression in jailed women defendants and its relationship to their adjudicative competence. *Journal of the American Academy of Psychiatry and the Law, 25*, 105-119.

Redlich, A. D., Silverman, M., & Steiner, H. (2003). Pre-adjudicative and adjudicative competence in juveniles and young adults. *Behavioral Sciences & the Law, 21*, 393-410.

Reich, J., & Wells, J. (1986a). Defendants with repeated competency evaluations. *Journal of Nervous and Mental Disease, 174*, 120-122.

Reich, J., & Wells, J. (1986b). The association between legal charge and psychiatric diagnosis in a competency-to-stand-trial population. *Social Psychiatry, 21*, 206-209.

±Richardson, H. A. (2005). *A community competency restoration program for defendants who do not meet commitment criteria.* Dissertation Abstracts International, 66 (4-B), 2314.

±Roach, R. L. (1994). *Assessment of trial competency in hospitalized criminal defendants.* Dissertation Abstracts International, 54 (11-B), 5952.

Robbins, E., Waters, J., & Herbert, P. (1997). Competency to stand trial evaluations: A study of actual practice in two states. *Journal of the American Academy of Psychiatry and the Law, 25,* 469-483.

*Roesch, R., & Golding, S. L. (1977). *A systems analysis of competency to stand trial procedures: Implications for forensic services in North Carolina* (Pre-final draft). Urbana: University of Illinois Press.

Rogers, R., Jackson, R. L., Sewell, K. W., & Harrison, K. S. (2004). An Examination of the ECST-R as a Screen for Feigned Incompetency to Stand Trial. *Psychological Assessment, 16,* 139-145.

Rogers, R., Jackson, R. L., Sewell, K. W., Tillbrook, C. E., & Martin, M. A. (2003). Assessing dimensions of competency to stand trial: Construct validation of the ECST-R. *Assessment, 10,* 344-351.

Rogers, R., Sewell, K. W., Grandjean, N. R., & Vitacco, M. (2002). The detection of feigned mental disorders on specific competency measures. *Psychological Assessment, 14,* 177-183.

Rogers, R., Ustad, K. L., Sewell, K. W., & Reinhardt, V. (1996). Dimensions of incompetency: A factor analytic study of the Georgia court competency test. *Behavioral Sciences & the Law, 14,* 323-330.

Rosenfeld, B., & Ritchie, K. (1998). Competence to stand trial: Clinician reliability and the role of offense severity. *Journal of Forensic Sciences, 43,* 151-157.

±Ryba, N. L. (2005). *An evaluation of the influence of psychiatric symptoms and cognitive abilities on competence-related abilities.* Dissertation Abstracts International, 65 (12-B), 6672.

Shatin, L. (1979). Brief form of the Competence Screening Test for mental competence to stand trial. *Journal of Clinical Psychology, 35*, 464-467.

Siegel, A. M., & Elwork, A. (1990). Treating incompetence to stand trial. *Law and Human Behavior, 14*, 57-65.

±Simon, A. (1999). *The restoration of competency to stand trail: An analysis of intervention efficacy.* Dissertation Abstracts International, 60 (3-B), 1315.

Smith, S. A. (1996). Comparison of two screening tests of competency to stand trial for defendants with mental retardation. *Psychological Reports, 78*, 234.

Smith, S. A., & Broughton, S. F. (1994). Competency to stand trial and criminal responsibility: An analysis in South Carolina. *Mental Retardation, 32*, 281-287.

Smith, S. A., & Hudson, R. L. (1995). A quick screening test of competency to stand trial for defendants with mental retardation. *Psychological Reports, 76*, 91-97.

Stoops, R., Hess, J., Scott, T., Tolan, J., Gallucci, G., & Brierley-Bowers, P. (2007). Training competency to stand trial in an individual with intellectual disability and behavioral health concerns. *Mental Health Aspects of Developmental Disabilities, 10*, 47-52.

Ustad, K. L., Rogers, R., Sewell, K. W., & Guarnaccia, C. A. (1996). Restoration of competency to stand trial: Assessment with the Georgia Court Competency Test and the Competency Screening Test. *Law and Human Behavior, 20*, 131-146.

Vann, C. R., & Morganroth, F. (1965). The psychiatrist as judge: A second look at the competence to stand trial. *University of Detroit Law Journal, 43*, 1-12.

Vernon, M., Steinberg, A. G., & Montoya, L. A. (1999). Deaf murderers: Clinical and forensic issues. *Behavioral Sciences & the Law, 17*, 495-516.

Viljoen, J. L., Roesch, R., & Zapf, P. A. (2002). An examination of the relationship between competency to stand trial, competency to waive interrogation rights, and psychopathology. *Law and Human Behavior, 26*, 481-506.

Viljoen, J. L., & Zapf, P. A. (2002). Fitness to stand trial evaluations: A comparison of referred and non-referred defendants. *International Journal of Forensic Mental Health, 1*, 127-138.

Vitacco, M. J., Rogers, R., Gabel, J., & Munizza, J. (2007). An evaluation of malingering screens with competency to stand Trial patients: A known-groups comparison. *Law and Human Behavior, 31*, 249-260.

Wall, B. W., Krupp, B. H., & Guilmette, T. (2003). Restoration of competency to stand trial: A training program for persons with mental retardation. *Journal of the American Academy of Psychiatry and the Law, 31*, 189-201.

±Whittemore, K. E. (2001). *Releasing the mentally disordered offender: Disposition decisions for individuals found unfit to stand trial and not criminally responsible.* Dissertation Abstracts International, 61 (7-B), 3866.

*Wildman, R. W, II, Batchelor, E. S., Thompson, L., Nelson, F. R., Moore, J. T, Patterson, M. E., & de Laosa, M. (1978). *The Georgia Court Competency Test: An attempt to develop a rapid, quantitative measure of fitness for trial.* Unpublished manuscript, Forensic Services Division, Central State Hospital, Milledgeville, GA.

Wygant, D. B., Sellbom, M., Ben-Porath, Y. S., Stafford, K. P., Freeman, D. B., Heilbronner, R. L. (2007). The relation between symptom validity testing and MMPI-2 scores as a function of forensic evaluation context. *Archives of Clinical Neuropsychology, 22*, 489-499.

±Zapf, P. A. (1999). *An investigation of the construct of competence in a criminal and civil context: A comparison of the fit, the MacCAT-CA, and the MacCAT-T. (adjudicative competence, treatment competence, Macarthur competence assessment tool).* Dissertation Abstracts International, 60 (6-B), 3012.

Zapf, P. A., & Roesch, R. (2001). A comparison of MacCAT-CA and the FIT for making determinations of competency to stand trial. *International Journal of Law and Psychiatry, 24*, 81-92.

Zapf, P. A., & Roesch, R. (2005). An Investigation of the Construct of Competence: A Comparison of the FIT, the MacCAT-CA, and the MacCAT-T. *Law and Human Behavior, 29,* 229-252.

Appendix C

Study-Level Coding Form

Bibliographic reference:
_____
_____
_____

___ ___ ___ ___ 1) Study ID Number [id]

___ 2) Type of publication [pubtype]

1 journal article
2 book chapter
3 book
4 conference presentation
5 dissertation
6 abstract
7 other (specify)

___ ___ ___ ___ 3) Publication year (all four digits; -99 if unknown) [pubyear]

___ 4) Coder [coder]

1 Gianni
2 Grazyna

Sample Descriptors

(-99 if missing and -88 if not applicable)
***NOTE:** indicate "missing" if information is discussed and not presented*

_____ 5) Total sample size [n]

_____ 6) Incompetent group sample size [n_ist]

_____ 7) Competent group sample size [n_cst]

___ ___ ___ ___ 8) Percentage of sample Incompetent to stand trial [ist]

Age

___ ___ ___ ___ 9a) Mean age of the Incompetent group [age_ist]

___ ___ ___ ___ 9b) Standard Deviation of age of the Incompetent group [sdage_ist]

___ ___ ___ ___ 9c) Mean age of the Competent group [age_cst]

___ ___ ___ ___ 9d) Standard Deviation of age of the Competent group [sdage_cst]

Ethnicity

___ ___ ___ ___ 10a) Number Caucasian in Incompetent group [cauc_ist]

___ ___ ___ ___ 10b) Number African American in Incompetent group [aa_ist]

___ ___ ___ ___ 10c) Number Hispanic in Incompetent group [hisp_ist]

___ ___ ___ ___ 10d) Number Other in Incompetent group [other_ist]

___ ___ ___ ___ 10e) Number Caucasian in Competent group [cauc_cst]

___ ___ ___ ___ 10f) Number African American in Competent group [aa_cst]

___ ___ ___ ___ 10g) Number Hispanic in Competent group [hisp_cst]

___ ___ ___ ___ 10h) Number Other in Competent group [other_cst]

Sex

Yes No N/R 11a) Includes Females [inclfem]

___ ___ ___ ___ 11b) Number Male in Incompetent group [male_ist]

___ ___ ___ ___ 11c) Number Female in Incompetent group [female_ist]

__ ___ ___ ___ 11d) Number Male in Competent group [male_cst]

___ ___ ___ ___ 11e) Number Female in Competent group [female_cst]

Education

___ ___ ___ ___ 12a) Mean years of education in Incompetent group [educ_ist]

___ ___ ___ ___ 12b) Standard Deviation of years of education of the Incompetent group [sdeduc_ist]

___ ___ ___ ___ 12c) Mean years of education in Competent group [educ_cst]

___ ___ ___ ___ 12d) Standard Deviation of years of education of the Competent group [sdeduc_cst]

Employment Status

___ ___ ___ ___ 13a) Number Unemployed in Incompetent group [unemp_ist]

___ ___ ___ ___ 13b) Number Unemployed in Competent group [unemp_cst]

Marital Status

___ ___ ___ ___ 14a) Number Single/Never Married in Incompetent group [single_ist]

___ ___ ___ ___ 14b) Number Married in Incompetent group [mar_ist]

___ ___ ___ ___ 14c) Number Separated/Divorced in Incompetent group [sepdiv_ist]

___ ___ ___ ___ 14d) Number Single/Never Married in Competent group [single_cst]

___ ___ ___ ___ 14e) Number Married in Competent group [mar_cst]

___ ___ ___ ___ 14f) Number Separated/Divorced in Competent group [sepdiv_cst]

Psychiatric Diagnosis

___ ___ ___ ___ 15a) Number of Incompetent group with a Psychotic Disorder [psych_ist]

___ ___ ___ ___ 15b) Number of Incompetent group with Mental Retardation [mr_ist]

___ ___ ___ ___ 15c) Mean IQ score for Incompetent group [IQ_ist]

___ ___ ___ ___ 15d) Standard Deviation of IQ of the Incompetent group [sdIQ_ist]

___ ___ ___ ___ 15e) Number of Incompetent group with a Mood Disorder [mood_ist]

___ ___ ___ ___ 15f) Number of Incompetent group with a Personality Disorder [person_ist]

___ ___ ___ ___ 15g) Number of Incompetent group with a Substance Use Disorder [subuse_ist]

___ ___ ___ ___ 15h) Number of Incompetent group with Other Disorder (specify) [otherdis_ist]

___ ___ ___ ___ 15i) Number of Competent group with a Psychotic Disorder [psych_cst]

___ ___ ___ ___ 15j) Number of Competent group with Mental Retardation [mr_cst]

___ ___ ___ ___ 15k) Mean IQ score for Competent group [IQ_cst]

___ ___ ___ ___ 15L) Standard Deviation of IQ of the Competent group [sdIQ_cst]

___ ___ ___ ___ 15m) Number of Competent group with a Mood Disorder [mood_cst]

___ ___ ___ ___ 15n) Number of Competent group with a Personality Disorder [person_cst]

___ ___ ___ ___ 15o) Number of Competent group with a Substance Use Disorder [subuse_ist]

___ ___ ___ ___ 15p) Number of Competent group with Other Disorder (specify) [otherdis_cst]

Psychiatric Symptoms

___ ___ ___ ___ 16a) Number of Incompetent group with Disorientation [disorient_ist]

___ ___ ___ ___ 16b) Number of Incompetent group with Impaired Judgment [impaired_ist]

___ ___ ___ ___ 16c) Number of Incompetent group with Delusions [delusion_ist]

___ ___ ___ ___ 16d) Number of Incompetent group with Hallucinations [halluc_ist]

___ ___ ___ ___ 16e) Number of Incompetent group with Impaired Memory [memory_ist]

___ ___ ___ ___ 16f) Number of Incompetent grp w/ Impaired Thought/Communication [thought_ist]

___ ___ ___ ___ 16g) Number of Incompetent group with Disturbed Behavior [behave_ist]

___ ___ ___ ___ 16h) Number of Incompetent group with Affective Disturbance [affect_ist]

___ ___ ___ ___ 16i) Number of Competent group with Disorientation [disorient_cst]

___ ___ ___ ___ 16j) Number of Competent group with Impaired Judgment [impaired_cst]

___ ___ ___ ___ 16k) Number of Competent group with Delusions [delusion_cst]

___ ___ ___ ___ 16L) Number of Competent group with Hallucinations [halluc_ist]

___ ___ ___ ___ 16m) Number of Competent group with Impaired Memory [memory_cst]

___ ___ ___ ___ 16n) Number of Competent grp w/ Impaired Thought/Communication [thought_cst]

___ ___ ___ ___ 16o) Number of Competent group with Disturbed Behavior [behave_cst]

___ ___ ___ ___ 16p) Number of Competent group with Affective Disturbance [affect_cst]

Psychiatric and Incompetency History

___ ___ ___ ___ 17a) Number Incompetent group w/ Previous Psych. Hospitalization [psyhosp_ist]

___ ___ ___ ___ 17b) Number Incompetent group w/ Other Psychiatric Care (specify) [othercare_ist]

___ ___ ___ ___ 17c) Number Competent group w/ Previous Psych. Hospitalization [psyhosp_cst]

___ ___ ___ ___ 17d) Number Competent group w/ Other Psychiatric Care (specify) [othercare_cst]

___ ___ ___ ___ 18a) Number Incompetent group w/ Previous Competency Evaluation [eval_ist]

___ ___ ___ ___ 18b) Number Competent group w/ Previous Competency Evaluation [eval_cst]

Legal History

___ ___ ___ ___ 19a) Number Incompetent group with a Prior Arrest History [arrest_ist]

___ ___ ___ ___ 19b) Number Incompetent group with a Prior Violent Crime Charge [prviol_ist]

___ ___ ___ ___ 19c) Number Incompetent group with a Current Violent Crime Charge [curviol_ist]

___ ___ ___ ___ 19d) Number Competent group with a Prior Arrest History [arrest_cst]

___ ___ ___ ___ 19e) Number Competent group with a Prior Violent Crime Charge [prviol_cst]

___ ___ ___ ___ 19f) Number Competent group with a Current Violent Crime Charge [curviol_cst]

Competency

___ 20) Source of *initial* competency decision [decision]

1 court decision
2 psychiatrist's decision
3 psychologist's decision
4 mixed (based on various individual decisions)
5 other mental health professional's decision (specify)
6 mental health professional "team" decision
7 "blue-ribbon" panel (of experts) decision
8 via assessment instrument score (specify)
9 competency determined during study
10 other (specify)
11 none (referred)
-99 not reported
-88 not applicable

___ 21) Source of competency decision used for comparison [dv_decision]

1 court decision
2 psychiatrist's decision
3 psychologist's decision
4 mixed (based on various individual decisions)
5 other mental health professional's decision (specify)
6 mental health professional "team" decision
7 "blue-ribbon" panel (of experts) decision
8 via assessment instrument score (specify)
9 competency determined during study
10 other (specify)
-99 not reported
-88 not applicable

___ 22) Type of competent comparison group [type_cst]

1 pure competent
2 referred competent
3 study competent
4 test competent
5 restored competent
6 mixed

___ 23a) Competency assessment instrument used [instrument]

1 CAST-MR
2 CST
3 CAI
4 CADCOMP
5 ECST
6 ECST-R
7 FIT
8 FIT-R
9 GCCT
10 GCCT-MSH
11 IFI
12 IFI-R
13 MacCAT-CA
14 MFQ
15 MFCS
16 Other (specify)
17 None
-99 not reported
-88 not applicable (n/a)

___ 23b) Second competency assessment instrument used [scdinstrument]

1 CAST-MR	10 GCCT-MSH
2 CST	11 IFI
3 CAI	12 IFI-R
4 CADCOMP	13 MacCAT-CA
5 ECST	14 MFQ
6 ECST-R	15 MFCS
7 FIT	16 Other (specify)
8 FIT-R	17 None
9 GCCT	-99 not reported
	-88 not applicable (n/a)

___ 24) Intelligence assessment instrument used [intelligence]

1 WAIS	4 None
2 Stanford Binet	-99 not reported
3 Other (specify)	

___ 25) Personality assessment instrument used [personality]

1 MMPI	4 None
2 Rorschach	-99 not reported
3 Other (specify)	

Research Design Descriptors

___ ___ 26) Setting of study [setting]

1 inpatient setting	4 other (specify)
2 outpatient setting	-99 missing
3 mixed	-88 not applicable

___ ___ 27) Sample's country of origin [country]

1 United States	3 other (specify)
2 Canada	-99 missing

___ ___ 28) Research Design [design]

1 one-group (correlational)	4 three-groups, comparison
2 one-group, pretest–posttest	5 other (specify)
3 two-groups, comparison	-99 cannot determine
	-88 not applicable

___ ___ 29) Recruitment of Participants [recruit]

1 prospective
2 archival/retrospective
3 simulation/analog
4 other (specify)

___ ___ 20) Type of Sample [sampletype]

1 matched sample
2 random sample
3 other (specify)

31) Archival Coding (describe)

_____

_____

___ ___ 32) Overall Coder Confidence Rating of Research Design Descriptors [confidence]

1 very low (little basis)
2 low (guess)
3 moderate (weak inference)
4 high (strong inference)
5 very high (explicitly stated)

Appendix D

Study-Level Coding Manual

Bibliographic reference: Write a complete citation in (approx.) APA form.

1) Study ID Number. Assign a unique identification number to each study. If a report presents two independent studies (i.e., two independent studies with different participants), then add a decimal to the study ID number to distinguish each study within a report and code each independent study separately.

2) What type of publication is the report? If two separate reports are being used to code a single study, code the type of the more formally published report (e.g., book or journal article).

1 journal article	4 conference presentation
2 book chapter	5 dissertation
3 book	6 abstract
	7 other (specify)

3) What is the publication year? Use all four digits (e.g., 1995) or "-99" if date is unknown. If two separate reports are being used to code a single study, code the publication year of the more formally published report.

4) Coder: Specify coder (i.e., Gianni or Grazyna)

Sample Descriptors

5) Total sample size.

6) Incompetent group sample size.

7) Competent group sample size.

8) Percentage of sample deemed Incompetent to stand trial. Specify the approximate or exact percentage of defendants in the sample who have been classified or deemed as Incompetent to stand trial using the best information available. If missing enter "-99"; if it is not applicable, enter "-88".

Age

9a) Mean age of the Incompetent group. Specify the approximate or exact mean age of the incompetent group using the best information available. Estimate mean age from age ranges if they are reported instead. If missing, enter "-99"; if it is not applicable, enter "-88".

9b) Standard Deviation of age for Incompetent group. If missing, enter "-99"; if it is not applicable, enter "-88".

9c) Mean age of the Competent group. Specify the approximate or exact mean age of the incompetent group using the best information available. Estimate mean age from age ranges if they are reported instead. If missing enter "-99"; if it is not applicable, enter "-88".

9d) Standard Deviation of age for Competent group. If missing, enter "-99"; if it is not applicable, enter "-88".

Ethnicity

10a) Number of Caucasian defendants in Incompetent group. Write in exact number. If missing enter "-99"; if it is not applicable, enter "-88".

10b) Number of African American defendants in Incompetent group. Write in exact number. If missing enter "-99"; if it is not applicable, enter "-88".

10c) Number of Hispanic defendants in Incompetent group. Write in exact number. If missing enter "-99"; if it is not applicable, enter "-88".

10d) Number of Other defendants in Incompetent group. Write in exact number. If missing enter "-99"; if it is not applicable, enter "-88".

10e) Number of Caucasian defendants in Competent group. Write in exact number. If missing enter "-99"; if it is not applicable, enter "-88".

10f) Number of African American defendants in Competent group. Write in exact number. If missing enter "-99"; if it is not applicable, enter "-88".

10g) Number of Hispanic defendants in Competent group. Write in exact number. If missing enter "-99"; if it is not applicable, enter "-88".

10h) Number of Other defendants in Competent group. Write in exact number. If missing enter "-99"; if it is not applicable, enter "-88".

Sex

11a) Includes Females. Circle Yes, No, or N/R (not reported) if the sample includes females.

11b) Number of Male defendants in Incompetent group. Write in exact number. If missing enter "-99"; if it is not applicable, enter "-88".

11c) Number of Female defendants in Incompetent group. Write in exact number. If missing enter "-99"; if it is not applicable, enter "-88".

11d) Number of Male defendants in Competent group. Write in exact number. If missing enter "-99"; if it is not applicable, enter "-88".

11e) Number of Female defendants in Competent group. Write in exact number. If missing enter "-99"; if it is not applicable, enter "-88".

Education

12a) Mean years of education in Incompetent group. Specify the approximate or mean number of years of education of the sample. Code the best information available. Estimate the mean years of education from grade level if necessary. If missing enter "-99"; if it is not applicable, enter "-88".

12b) Standard Deviation of years of education for Incompetent group. If missing enter "-99"; if it is not applicable, enter "-88".

12c) Mean years of education in Competent group. Specify the approximate or mean number of years of education of the sample. Code the best information available. Estimate the mean years of education from grade level if necessary. If missing enter "-99"; if it is not applicable, enter "-88".

12d) Standard Deviation of years of education for Competent group. If missing, enter "-99"; if it is not applicable, enter "-88".

Employment Status

13a) Number of Unemployed defendants in Incompetent group. Specify the approximate or mean number of defendants in the Incompetent group who are Unemployed. If missing, enter "-99"; if it is not applicable, enter "-88".

13b) Number of Unemployed defendants in Competent group. Specify the approximate or mean number of defendants in the Competent group who are Unemployed. If missing, enter "-99"; if it is not applicable, enter "-88".

Marital Status

14a) Number Single/Never Married in Incompetent group. Specify the approximate or mean number of defendants in the Incompetent group who were Single or Never Married at the time of the study. If missing, enter "-99"; if it is not applicable, enter "-88".

14b) Number Married in Incompetent group. Specify the approximate or mean number of defendants in the Incompetent group who were Married at the time of the study. If missing, enter "-99"; if it is not applicable, enter "-88".

14c) Number Separated/Divorced in Incompetent group. Specify the approximate or mean number of defendants in the Incompetent group who were Separated or Divorced at the time of the study. If missing, enter "-99"; if it is not applicable, enter "-88".

14d) Number Single/Never Married in Competent group. Specify the approximate or mean number of defendants in the Competent group who were Single or Never Married at the time of the study. If missing, enter "-99"; if it is not applicable, enter "-88".

14e) Number Married in Competent group. Specify the approximate or mean number of defendants in the Competent group who were Married at the time of the study. If missing, enter "-99"; if it is not applicable, enter "-88".

14f) Number Separated/Divorced in Competent group. Specify the approximate or mean number of defendants in the Competent group who were Separated or Divorced at the time of the study. If missing, enter "-99"; if it is not applicable, enter "-88".

Psychiatric Diagnosis

15a) Number of Incompetent group with a Psychotic Disorder. Specify the approximate or mean number of individuals in the Incompetent group diagnosed with a Psychotic Disorder. If missing, enter "-99"; if it is not applicable, enter "-88".

15b) Number of Incompetent group with Mental Retardation. Specify the approximate or mean number of individuals in the Incompetent group diagnosed with Mental Retardation. If missing, enter "-99"; if it is not applicable, enter "-88".

15c) Mean IQ score for Incompetent group. Specify the approximate or mean IQ score for the Incompetent group. Code the best information available. Estimate the mean IQ score if classificatory ranges are provided. If missing, enter "-99"; if it is not applicable, enter "-88".

15d) Standard Deviation of IQ for Incompetent group. If missing, enter "-99"; if it is not applicable, enter "-88".

15e) Number of Incompetent group with a Mood Disorder. Specify the approximate or mean number of individuals in the Incompetent group diagnosed with a Mood Disorder. If missing, enter "-99"; if it is not applicable, enter "-88".

15f) Number of Incompetent group with a Personality Disorder. Specify the approximate or mean number of individuals in the Incompetent group diagnosed with a Personality Disorder. If missing, enter "-99"; if it is not applicable, enter "-88".

15g) Number of Incompetent group with a Substance Use Disorder. Specify the approximate or mean number of individuals in the Incompetent group diagnosed with a Substance Use Disorder. If missing, enter "-99"; if it is not applicable, enter "-88".

15h) Number of Incompetent group with Other Disorder (specify). Specify the approximate or mean number of individuals in the Incompetent group diagnosed with Other Disorder. Specify the disorder. If missing, enter "-99"; if it is not applicable, enter "-88".

15i) Number of Competent group with a Psychotic Disorder. Specify the approximate or mean number of individuals in the Competent group diagnosed with a Psychotic Disorder. If missing, enter "-99"; if it is not applicable, enter "-88".

15j) Number of Competent group with Mental Retardation. Specify the approximate or mean number of individuals in the Competent group diagnosed with Mental Retardation. If missing, enter "-99"; if it is not applicable, enter "-88".

15k) Mean IQ score for Competent group. Specify the approximate or mean IQ score for the Competent group. Code the best information available. Estimate the mean IQ score if classificatory ranges are provided. If missing, enter "-99"; if it is not applicable, enter "-88".

15L) Standard Deviation of IQ for Competent group. If missing, enter "-99"; if it is not applicable, enter "-88".

15m) Number of Competent group with a Mood Disorder. Specify the approximate or mean number of individuals in the Competent group diagnosed with a Mood Disorder. If missing, enter "-99"; if it is not applicable, enter "-88".

15n) Number of Competent group with a Personality Disorder. Specify the approximate or mean number of individuals in the Competent group diagnosed with a Personality Disorder. If missing, enter "-99"; if it is not applicable, enter "-88".

15o) Number of Competent group with a Substance Use Disorder. Specify the approximate or mean number of individuals in the Competent group diagnosed with a Substance Use Disorder. If missing, enter "-99"; if it is not applicable, enter "-88".

15p) Number of Competent group with Other Disorder (specify). Specify the approximate or mean number of individuals in the Incompetent group diagnosed with Other Disorder. Specify the disorder. If missing, enter "-99"; if it is not applicable, enter "-88".

Psychiatric Symptoms

16a) Number of Incompetent group with Disorientation. Specify the approximate or mean number of individuals in the Incompetent group with symptoms of Disorientation. If missing, enter "-99"; if it is not applicable, enter "-88".

16b) Number of Incompetent group with Impaired Judgment. Specify the approximate or mean number of individuals in the Incompetent group with symptoms of Impaired Judgment. If missing, enter "-99"; if it is not applicable, enter "-88".

16c) Number of Incompetent group with Delusions. Specify the approximate or mean number of individuals in the Incompetent group with symptoms of Delusions. If missing, enter "-99"; if it is not applicable, enter "-88".

16d) Number of Incompetent group with Hallucinations. Specify the approximate or mean number of individuals in the Incompetent group with symptoms of Hallucinations. If missing, enter “-99”; if it is not applicable, enter “-88”.

16e) Number of Incompetent group with Impaired Memory. Specify the approximate or mean number of individuals in the Incompetent group with symptoms of Impaired Memory. If missing, enter “-99”; if it is not applicable, enter “-88”.

16f) Number of Incompetent group w/ Impaired Thought/Communication. Specify the approximate or mean number of individuals in the Incompetent group with symptoms of Impaired Thought or Communication. If missing, enter “-99”; if it is not applicable, enter “-88”.

16g) Number of Incompetent group with Disturbed Behavior. Specify the approximate or mean number of individuals in the Incompetent group with symptoms of Disturbed Behavior. If missing, enter “-99”; if it is not applicable, enter “-88”.

16h) Number of Incompetent group with Affective Disturbance. Specify the approximate or mean number of individuals in the Incompetent group with symptoms of Affective Disturbance. If missing, enter “-99”; if it is not applicable, enter “-88”.

16i) Number of Competent group with Disorientation. Specify the approximate or mean number of individuals in the Competent group with symptoms of Disorientation. If missing, enter “-99”; if it is not applicable, enter “-88”.

16j) Number of Competent group with Impaired Judgment. Specify the approximate or mean number of individuals in the Competent group with symptoms of Impaired Judgment. If missing, enter “-99”; if it is not applicable, enter “-88”.

16k) Number of Competent group with Delusions. Specify the approximate or mean number of individuals in the Competent group with symptoms of Delusions. If missing, enter “-99”; if it is not applicable, enter “-88”.

16L) Number of Competent group with Hallucinations. Specify the approximate or mean number of individuals in the Competent group with symptoms of Hallucinations. If missing, enter “-99”; if it is not applicable, enter “-88”.

16m) Number of Competent group with Impaired Memory. Specify the approximate or mean number of individuals in the Competent group with symptoms of Impaired Memory. If missing, enter “-99”; if it is not applicable, enter “-88”.

16n) Number of Competent group w/ Impaired Thought/Communication. Specify the approximate or mean number of individuals in the Incompetent group with symptoms of Impaired Thought or Communication. If missing, enter “-99”; if it is not applicable, enter “-88”.

16o) Number of Competent group with Disturbed Behavior. Specify the approximate or mean number of individuals in the Competent group with symptoms of Disturbed Behavior. If missing, enter "-99"; if it is not applicable, enter "-88".

16p) Number of Competent group with Affective Disturbance. Specify the approximate or mean number of individuals in the Competent group with symptoms of Affective Disturbance. If missing, enter "-99"; if it is not applicable, enter "-88".

Psychiatric and Incompetency History

17a) Number of Incompetent group with a Previous Psychiatric Hospitalization. Specify the approximate or mean number of individuals in the Incompetent group who have had a previous psychiatric hospitalization. If missing, enter "-99"; if it is not applicable, enter "-88".

17b) Number of Incompetent group with Other Psychiatric Care (specify). Specify the approximate or mean number of individuals in the Incompetent group who have had Other psychiatric care. Specify the type of care received. If missing, enter "-99"; if it is not applicable, enter "-88".

17c) Number of Competent group with a Previous Psychiatric Hospitalization. Specify the approximate or mean number of individuals in the Competent group who have had a previous psychiatric hospitalization. If missing, enter "-99"; if it is not applicable, enter "-88".

17d) Number of Competent group with Other Psychiatric Care (specify). Specify the approximate or mean number of individuals in the Competent group who have had Other psychiatric care. Specify the type of care received. If missing, enter "-99"; if it is not applicable, enter "-88".

18a) Number of Incompetent group with a Previous Competency Evaluation. Specify the approximate or mean number of individuals in the Incompetent group who have had a previous competency evaluation. If missing, enter "-99"; if it is not applicable, enter "-88".

18b) Number of Competent group with a Previous Competency Evaluation. Specify the approximate or mean number of individuals in the Competent group who have had a previous competency evaluation. If missing, enter "-99"; if it is not applicable, enter "-88".

Legal History

19a) Number of Incompetent group with a Prior Arrest History. Specify the approximate or mean number of individuals in the Incompetent group who have been Previously Arrested. If missing, enter "-99"; if it is not applicable, enter "-88".

19b) Number of Incompetent group with a Prior Violent Crime Charge. Specify the approximate or mean number of individuals in the Incompetent group who have been Previously Charged with a Violent Crime. If missing, enter "-99"; if it is not applicable, enter "-88".

19c) Number of Incompetent group with a Current Violent Crime Charge. Specify the approximate or mean number of individuals in the Incompetent group who have been Currently Charged with a Violent Crime. If missing, enter "-99"; if it is not applicable, enter "-88".

19d) Number of Competent group with a Prior Arrest History. Specify the approximate or mean number of individuals in the Competent group who have been Previously Arrested. If missing, enter "-99"; if it is not applicable, enter "-88".

19e) Number of Competent group with a Prior Violent Crime Charge. Specify the approximate or mean number of individuals in the Competent group who have been Previously Charged with a Violent Crime. If missing, enter "-99"; if it is not applicable, enter "-88".

19f) Number of Competent group with a Current Violent Crime Charge. Specify the approximate or mean number of individuals in the Competent group who have been Currently Charged with a Violent Crime. If missing, enter "-99"; if it is not applicable, enter "-88".

Competency

20) Source of *initial* competency decision (to qualify for research)

1 court decision
2 psychiatrist's decision
3 psychologist's decision
4 mixed (based on various individual decisions)
5 other mental health professional's decision (specify)
6 mental health professional "team" decision
7 "blue-ribbon" panel (of experts) decision
8 via assessment instrument score (specify)
9 competency determined during study
10 other (specify)
11 none (referred)
-99 not reported
-88 not applicable

21) Source of competency decision used for comparison

1 court decision
2 psychiatrist's decision
3 psychologist's decision
4 mixed (based on various individual decisions)
5 other mental health professional's decision (specify)
6 mental health professional "team" decision
7 "blue-ribbon" panel (of experts) decision
8 via assessment instrument score (specify)
9 competency determined during study
10 other (specify)
-99 not reported
-88 not applicable

22) Type of competent comparison group

1 pure competent
(e.g., a jail group or any group whereby there was NEVER a question of competency)
2 referred competent
(i.e., a group referred for a competency evaluation and found competent)
3 study competent
(e.g., initially found incompetent by the court, but reclassified as competent for the purpose of the research based on an evaluation (during research) for the purpose of the research)
4 test competent
(e.g., initially deemed incompetent by the court, but reclassified as competent based on competency measure for the purpose of the research)
5 restored competent
(e.g., initially found incompetent by the court, but restored and found competent)
6 mixed
-select if competent groups were combined for the purposes of this meta-analysis

23a) Competency assessment instrument used

1 CAST-MR
2 CST
3 CAI
4 CADCOMP
5 ECST
6 ECST-R
7 FIT
8 FIT-R
9 GCCT
10 GCCT-MSH
11 IFI
12 IFI-R
13 MacCAT-CA
14 MFQ
15 MFCS
16 Other (specify)
17 None
-99 not reported
-88 not applicable (n/a)

23b) Second competency assessment instrument used

1 CAST-MR
2 CST
3 CAI
4 CADCOMP
5 ECST
6 ECST-R
7 FIT
8 FIT-R
9 GCCT
10 GCCT-MSH
11 IFI
12 IFI-R
13 MacCAT-CA
14 MFQ
15 MFCS
16 Other (specify)
17 None
-99 not reported
-88 not applicable (n/a)

24) Intelligence assessment instrument used

1 WAIS	4 None
2 Stanford Binet	-99 not reported
3 Other (specify)	

25) Personality assessment instrument used

1 MMPI	4 None
2 Rorschach	-99 not reported
3 Other (specify)	

Research Design Descriptors

26) Setting of study. Select the code that best describes where the study was conducted. If missing, enter "-99"; if it is not applicable, enter "-88".

1 inpatient setting
2 outpatient setting
3 mixed
4 other (specify)

27) Sample's country of origin. Select the code that best describes from which country the sample was drawn. If missing enter "-99".

1 United States	3 other (specify)
2 Canada	-99 missing

28) Research design. Select the code that best describes how the study was conducted with respect to the groups included. If missing, enter "-99"; if it is not applicable, enter "-88".

1 one-group (correlational)	4 three-groups, comparison
2 one-group, pretest–posttest	5 other (specify)
3 two-groups, comparison	

29) Recruitment of Participants. Select the code that best describes how Participants were Recruited for the study. If missing, enter "-99"; if it is not applicable, enter "-88".

1 prospective	3 simulation/analog
2 archival/retrospective	4 other (specify)

30) Type of Sample. Select the code that best describes how the samples were derived for comparison.

1 matched sample (e.g., matched in sample sizes)
2 random sample (e.g., all evaluations in a particular center for a set period of time)
3 other (specify)

31) Archival Coding. If data was collected in an archival or retrospective fashion, describe how it was coded.

32) Overall Coder Confidence Rating of Research Design Descriptors. Select the code that best describes your level of confidence in rating the aforementioned research design descriptors.

1 very low (little basis)
2 low (guess)
3 moderate (weak inference)
4 high (strong inference)
5 very high (explicitly stated)

Appendix E

Continuous Outcome Coding Form

__ __ __ __ 1) Study ID number

__ __ __ __ 2) Effect size sequence number [es_numb]

Dependent Measure Descriptors

__ __ 3) Type of outcome measure [outcome]

1 competency decision (i.e., incompetent/competent)
2 competency measure
3 Demographic variable
4 Psychiatric variable
5 Criminological variable
6 Personality measure
7 Intelligence measure
8 other (specify)
-99 missing
-88 not applicable

4) Outcome descriptor: _____

Effect Size Data

__ 5) Type of effect size data [es_type]

1 means and standard deviations
2 *t*-value or *F*-value
3 chi-square (df =1)
4 frequencies or proportions, dichotomous
5 frequencies or proportions polychotomous
6 other (specify)

__ __ __ __ 6) Page number(s) where the data for this effect size was found [page]

__ 7) Raw difference in direction of which group? [rawdiff]

1 incompetent group
2 neither (exactly equal)
3 competent group
-99 not reported

Sample Size

__ __ __ 8a) Incompetent group sample size
__ __ __ 8b) Competent group sample size

Means and Standard Deviations

__ __ __ __ 9a) Incompetent group mean [ist_mean]
__ __ __ __ 9b) Competent group mean [cst_mean]
__ __ __ __ 10a) Incompetent group standard deviation [ist_sd]
__ __ __ __ 10b) Competent group standard deviation [cst_sd]

Significance Tests

__ __ __ __ 11a) *t*-value [tvalue]
__ __ __ __ 11b) *F*-value (df for the numerator must equal 1) [Fvalue]
__ __ __ __ 11c) Chi-square value (df = 1) [chivalue]

__ __ __ __ 12) p-value [pvalue]

Calculated Effect Size

__ __.__ __ 13) Effect size [effectsize]

__ 14) Confidence rating in effect size computation [es_confidence]

1 highly estimated (have *N* and crude p-value only, such as $p < .10$, and must reconstruct via rough *t*-test)
2 moderate estimation (have complex but relatively complete statistics, such as multifactor ANOVA, as basis for estimation)
3 some estimation (have unconventional statistics and must convert to equivalent *t*-values or have conventional statistics but incomplete, such as exact *p*-level)
4 slight estimation (must use significance testing statistics rather than descriptive statistics, but have complete statistics of conventional sort)
5 no estimation (have descriptive data such as means, standard deviations, frequencies, proportions, etc. and can calculate the effect size directly)

Appendix F

Continuous Outcome Coding Manual

For each effect size, code the following information. Note that studies will have different numbers of effect sizes, and hence, different numbers of effect size level coding forms.

1) Study ID number. ID number of the study from which effect size is coded.

2) Effect size number. Assign each effect size within a study a unique number. Number multiple effect sizes within a study sequentially, e.g., 1, 2, 3, 4, etc.

Dependent Measure Descriptors

3) Type of outcome measure. Select the code that best describes the dependent variable. If missing enter "-99"; if it is not applicable, enter "-88".

1 competency decision (i.e., incompetent/competent)
2 competency measure
3 Demographic variable
4 Psychiatric variable
5 Criminological variable
6 Personality measure
7 Intelligence measure
8 other (specify)
-99 missing
-88 not applicable

4) Outcome descriptor. Write in a description of an outcome variable.

Effect Size Data

5) Type of effect size data.

1 means and standard deviations
2 *t*-value or *F*-value
3 chi-square (df =1)
4 frequencies or proportions, dichotomous
5 frequencies or proportions, polychotomous
6 other (specify)

6) Page number(s) where the data for this effect size was found.

7) Raw difference in direction of which group?

1 incompetent group
2 neither (exactly equal)
3 competent group
-99 not reported
-88 not applicable

Sample Size

8a) Incompetent group sample size (write in appropriate number)
8b) Competent group sample size (write in appropriate number)

When means and standard deviations are reported or can be estimated:

9a) Incompetent group mean (write in value for the mean, if available)
9b) Competent group mean (write in value for the mean, if available)
10a) Incompetent group standard deviation (write in the value for the *sd*, if available)
10b) Competent group standard deviation (write in the value for *sd*, if available)

When significance test information is reported:

11a) *t*-value (write in the value, if available)
11b) *F*-value (df for the numerator must equal 1) (write in the value, if available)
11c) Chi-square value (df = 1) (write in the value, if available)

12) p-value (If significant but not reported, code .049; if not sig. & not reported, code 0)

Calculated Effect Size

13) Effect size using the Excel effect size determination program or calculated by hand or by ***xxxx*** program. Report to two decimals with an algebraic sign in front: Plus if difference favors incompetent and minus if difference favors competent; Use -99.99 if NA

14) Confidence rating in effect size computation.

1 highly estimated (have *N* and crude p-value only, such as $p < .10$, and must reconstruct via rough *t*-test)
2 moderate estimation (have complex but relatively complete statistics, such as multifactor ANOVA, as basis for estimation)
3 some estimation (have unconventional statistics and must convert to equivalent *t*-values or have conventional statistics but incomplete, such as exact *p*-level)
4 slight estimation (must use significance testing statistics rather than descriptive statistics, but have complete statistics of conventional sort)
5 no estimation (have descriptive data such as means, standard deviations, frequencies, proportions, etc. and can calculate the effect size directly)

References

Ackerman, M. J. (1999). *Essentials of forensic psychological assessment*. New York, NY: John Wiley & Sons, Inc.

Altman, D. G. (1991). *Practical statistics for medical research.* London: Chapman and Hall.

American Psychological Association. (2001). *Publication manual of the American Psychological Association* (5th Edition). Washington, DC: Author.

Archer, R. P. (2006). *Forensic uses of clinical assessment instruments.* Mahwah, NJ, US: Lawrence Erlbaum.

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

Aubrey, M. (1988). Characteristics of competency referral defendants and nonreferred criminal defendants. *Journal of Psychiatry & Law, 16*, 233-245.

Bagby, R. M., Nicholson, R. A., Rogers, R., & Nussbaum, D. (1992). Domains of competency to stand trial: A factor analytic study. *Law and Human Behavior, 16*, 491-507.

Bendt, R. H., Balcanoff, E. J., & Tragellis, G. S. (1973). Incompetency to stand trial: Is psychiatry necessary? *The American Journal of Psychiatry, 130,* 1288-1289.

Bennett, G. T., & Kish, G. R. (1990). Incompetency to stand trial: Treatment unaffected by demographic variables. *Journal of Forensic Sciences, 35*, 403-412.

Blackstone, W. (1783). *Commentaries on the laws of England* (9th ed.). London: W. Strahan.

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

Bonnie, R. J. (1993). The competence of criminal defendants: Beyond *Dusky* and *Drope*. *University of Miami Law Review, 47,* 539-601.

Bonnie, R. J., & Grisso, T. (2000). Adjudicative competence and youthful offenders. In: T. Grisso & R. G. Schwartz, (Eds.), *Youth on trial: A developmental perspective on juvenile justice* (pp. 73-103). Chicago, IL: University of Chicago Press.

Borenstein, M. (2009). Effect sizes for continuous data. In H. Cooper & L. V. Hedges (Eds.), *The handbook of research synthesis, Second Edition* (pp. 221-235). New York: Russell Sage.

Borenstein, M., Hedges, L., Higgins, J., & Rothstein, H. (2005). Comprehensive Meta-analysis Version 2, Biostat, Englewood NJ.

Borum, R, & Grisso, T. (1995). Psychological test use in criminal forensic evaluations. *Professional Psychology: Research and Practice, 26,* 465-473.

Brown, D. L., Felthous, A. R., Barratt, E. S., & Stanford, M. (1994). The incompetent defendant: Support systems help avoid future legal problems. *Journal of Forensic Sciences, 39,* 1057-1068.

Bukatman, B. A., Foy, J. L., & Degrazia, E. (1971). What is competency to stand trial? *American Journal of Psychiatry, 127,* 1225-1229.

Caldwell, R. M., Mandracchia, S. A., Ross, S. A., & Silver, N. C. (2003). Competency to stand trial and criminal responsibility: An examination of racial and gender differences among African American and Caucasian pretrial defendants. *American Journal of Forensic Psychology, 21,* 5-19.

Carbonell, J. L., Heilbrun, K., & Friedman, F. L. (1992). Predicting who will regain trial competency: Initial promise unfulfilled. *Forensic Reports, 5,* 67-76.

Cohen, J. (1977). *Statistical power analysis for the behavioral sciences* (Rev. ed.). New York: Academic Press.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Erlbaum.

Colwell, K., Colwell, L. H., Perry, A. T., Wasieleski, D., & Billings, T. (2008). The Test of Malingered Incompetence (TOMI): A forced-choice instrument for assessing cognitive malingering in competence to stand trial evaluations. *American Journal of Forensic Psychology, 26,* 17-42.

Cooke, G. (1969). The court study unit: Patient characteristics and differences between patients judged competent and incompetent. *Journal of Clinical Psychology, 25,* 140-143.

Cooke, G., Johnston, N., & Pogany, E. (1973). Factors affecting referral to determine competency to stand trial. *American Journal of Psychiatry, 130,* 870-875.

Cooper, D., & Grisso, T. (1997). Five year research update (1991-1995): Evaluations for competence to stand trial. *Behavioral Sciences and the Law, 15,* 167-180.

Cooper, H., & Hedges, L. V. (1994). Potentials and limitations of research synthesis. In H. Cooper and L. V. Hedges (Eds). *The handbook of research synthesis (pp.521-529).* New York: Russell Sage Foundation.

Crocker, A. G., Favreau, O. E., & Caulet, M. (2002). Gender and fitness to stand trial: A 5-year review of remands in Québec. *International Journal of Law and Psychiatry*, *25,* 67-84.

Daniel, A. E., Beck, N. C., Herath, A., Schmitz, M., & Menninger, K. (1984). Factors correlated with psychiatric recommendations of incompetency and insanity. *Journal of Psychiatry & Law, 12*, 527-544.

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

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

Eagly, A. H., & Wood, W. (1994). Using research synthesis to plan future research. In H. Cooper & L. V. Hedges (Eds.), *The handbook of research synthesis* (pp. 485-500). New York: Russell Sage.

Everington, C. (1989). Demographic variables associated with competence to stand trial referral and evaluation of criminal defendants with mental retardation. *Journal of Psychiatry & Law, 17,* 627-640.

Everington, C. T. (1990). The Competence Assessment for Standing Trial for Defendants with Mental Retardation (CAST-MR): A validation study. *Criminal Justice and Behavior, 17,* 147-168.

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.

Everington, C., Notario-Smull, H., & Horton, M. L. (2007). Can defendants with mental retardation successfully fake their performance on a test of competence to stand trial? *Behavioral Sciences & the Law, 25*, 545-560.

Favole, R. J. (1983). Mental disability in the American criminal process: A four issue survey. In J. Monahan & H. J. Steadman (Eds.), *Mentally disordered offenders: Perspectives from law and social science* (pp. 247-295). New York: Plenum.

Fleiss, J. L. (1981). *Statistical methods for rates and proportions (2nd ed.).* New York: Wiley.

Fleiss, J. L., Levin, B. A., & Paik, M. C. (2003). *Statistical methods for rates and proportions (3rd Edition).* New York: John Wiley and Sons.

Frith's Case, 22 How. St. Tr. 307 (1790).

Gabel, J. E. (2007). *Consistency of screening measures of malingering among pre-trial adjudicative competence defendants.* Dissertation Abstracts International, 67 (9-B), 5400.

Gacono, C. B., Evans, F. B., Kaser-Boyd, N., & Gacono, L. A. (2008). *The handbook of forensic Rorschach assessment.* New York, NY, US: Routledge/Taylor & Francis.

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

Golding, S. L. (1992). Studies of incompetent defendants: Research and social policy implications. *Forensic Reports, 5,* 77-83.

Golding, S. L. Roesch, R., & Schreiber, J. (1984). Assessment and conceptualization of competency to stand trial: Preliminary data on the Interdisciplinary Fitness Interview. *Law & Human Behavior, 8,* 321-334.

Goldstein, A. M. (2003). *Handbook of psychology: Forensic psychology (vol. 11).* Hoboken, NJ: John Wiley & Sons, Inc.

Goldstein, A. M. (2007). Forensic psychology: Toward a standard of care. In A. M. Goldstein (Ed.), *Forensic psychology: Emerging topics and expanding roles* (pp. 3-41). Hoboken, NJ: John Wiley & Sons, Inc.

Gothard, S., Rogers, R., & Sewell, K. W. (1995). Feigning incompetency to stand trial: An investigation of the Georgia Court Competency Test. *Law and Human Behavior, 19,* 363-373.

Grandjean, N. R. (2004). *Neuropsychological predictors of incompetency to stand trial in defendants referred for competency restoration.* Dissertation Abstracts International, 66 (7-B), 3947.

Grisso, T. (1986). *Evaluating competencies: Forensic assessments and instruments*. New York, NY: Kluwer Academic/Plenum Publishers.

Grisso, T. (1992). Five-year research update (1986-1990): Evaluations for competence to stand trial. *Behavioral Sciences and the Law, 10,* 353-369.

Grisso, T. (2003). *Evaluating competencies: Forensic assessments and instruments (2nd ed.).* New York, NY: Kluwer Academic/Plenum Publishers.

Haddock, C. K., Rindskopf, D., & Shadish, W. R. (1998). Using odds ratios as effect sizes for meta-analysis of dichotomous data: A primer on methods and issues. *Psychological Methods, 3,* 339-353.

Hedges, L. V. & Olkin, I. (1985). *Statistical methods for meta-analysis.* New York: Academic Press.

Heilbrun, K. (2001). *Principles of forensic mental health assessment*. New York, NY: Kluwer Academic/Plenum Publishers.

Heilbrun, K., Marczyk, G. R., & DeMatteo, D. (2002). *Forensic mental health assessment: A casebook.* New York, NY: Oxford University Press.

Heller, M. S., Traylor., W. H., Ehrlich, S. M., & Lester, D. (1981). Intelligence, psychosis and competency to stand trial. *Bulletin of the American Academy of Psychiatry & the Law, 9,* 267-274.

Heller, M. S., Traylor., W. H., Ehrlich, S. M., & Lester, D. (1983). Competent and incompetent defendants referred to a court psychiatric clinic: A clinical comparison. *Journal of Forensic Sciences, 28,* 415-422.

Hilliard, R. B. (1993). Single-case methodology in psychotherapy process and outcome research. *Journal of Consulting and Clinical Psychology, 61,* 373-380.

Hoge, S. K., Bonnie, R. J., Poythress, N., Monahan, J., Eisenberg, M., & Feucht-Haviar, T. (1997a). The MacArthur adjudicative competence study: Development and validation of a research instrument. *Law and Human Behavior, 21*, 141-179.

Hoge, S. K., Poythress, N., Bonnie, R., Eisenberg, M., Monahan, J., Feucht-Haviar, T., & Oberlander, L. (1996). Mentally ill and non-mentally ill defendants' abilities to understand information relevant to adjudication: A preliminary study. *Bulletin of the American Academy of Psychiatry & the Law, 24*, 187-197.

Hoge, S. K., Poythress, N., Bonnie, R. J., Monahan, J., Eisenberg, M., & Feucht-Haviar, T. (1997b). The MacArthur adjudicative competence study: Diagnosis, psychopathology, and competence-related abilities. *Behavioral Sciences & the Law, 15*, 329-345.

Hubbard, K. L., & Zapf, P. A. (2003). The role of demographic, criminal, and psychiatric variables in examiners' predictions of restorability to competency to stand trial. *International Journal of Forensic Mental Health, 2*, 145-155.

Jackson v. Indiana, 402 U.S. 715 (1972).

Jackson, R. L., Rogers, R., & Sewell, K. W. (2005). Forensic Applications of the Miller Forensic Assessment of Symptoms Test (MFAST): Screening for Feigned Disorders in Competency to Stand Trial Evaluations. *Law and Human Behavior, 29*, 199-210.

Johnson, W. G. Nicholson., R. A., Service, N. M. (1990). The relationship of competency to stand trial and criminal responsibility. *Criminal Justice and Behaviour, 17*, 169.

Kline, R. B. (2004). Nonparametric effect size indexes. In R. B. Kline *Beyond significance testing: Reforming data Analysis methods in behavioral research (pp. 143-162).* Washington, DC: American Psychological Association.

Laboratory of Community Psychiatry, Harvard Medical School (1974). *Competency to stand trial and mental illness.* New York: Jason Aronson, Inc.

Laczko, A. L. James, J. F., & Alltop, L. B. (1970). A study of four hundred and thirty-five court-referred cases. *Journal of Forensic Sciences, 15*, 311-323.

Landis, J., & Koch, G. (1977). The measurement of observer agreement for categorical data. *Biometrics, 33*, 159-174.

Lesser, F. I. (1990). *Formulating a prediction model for regaining competency to stand trial.* Dissertation Abstracts International, 51 (2-B), 993.

Maxson, L. S., & Neuringer., C. (1970). Evaluating legal competency. *Journal of Genetic Psychology, 117*, 267-273.

McDonald, D. A., Nussbaum, D. S., & Bagby, R. M. (1991). Reliability, validity and utility of the Fitness Interview Test. *The Canadian Journal of Psychiatry, 36*, 480-484.

Melton, G., B., Petrila, J., Poythress, N. G., & Slobogin, C. (1997). *Psychological evaluations for the courts: A handbook for mental health professionals and lawyers* (2nd ed). New York, NY: Guilford Press.

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

Messick, S. (1995). Validity of psychological assessment: Validation of inferences from persons' responses and performances as scientific inquiry into score meaning. *American Psychologist, 50*, 741-749.

Miller, H. A. (2004). Examining the Use of the M-FAST With Criminal Defendants Incompetent to Stand Trial. *International Journal of Offender Therapy and Comparative Criminology, 48,* 268-280.

Mosley, D., Thyer, B. A., & Larrison, C. (2001). Development and preliminary validation of the Mosley Forensic Competency Scale. *Journal of Human Behavior in the Social Environment, 4,* 41-48.

Mossman, D. (2007). Predicting restorability of incompetent criminal defendants. *The Journal of the American Academy of Psychiatry and the Law, 35,* 34-43.

Mueller, C. L. (2007). *An examination of predictor variables for competency to stand trial determinations in a multicultural setting.* Dissertation Abstracts International, 68 (2-B), 1315.

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 and the Law, 21,* 329-350.

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

Nestor, P. G., Daggett, D., Haycock, J., & Price, M. (1999). Competence to stand trial: A neuropsychological inquiry. *Law and Human Behavior, 23,* 397-412.

Nicholson, R. A. (1986). Correlates of commitment status in psychiatric patients. *Psychological Bulletin, 100,* 241-250.

Nicholson, R. A. (1988). Validation of a brief form of the Competency Screening Test. *Journal of Clinical Psychology, 44,* 87-90.

Nicholson, R. A., Briggs, S. R., & Robertson, H. C. (1988). Instruments for assessing competency to stand trial: How do they work? *Professional Psychology: Research and Practice, 19,* 383-394.

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

Nicholson, R. A., & McNulty, J. L. (1992). Outcome of hospitalization for defendants found incompetent to stand trial. *Behavioral Sciences & the Law, 10*, 371-383.

Nicholson, R. A., & Norwood, S. (2000). The quality of forensic psychological assessments, reports, and testimony: Acknowledging the gap between promise and practice. *Law and Human Behavior, 24,* 9-44.

Nicholson, R. A., Robertson, H. C., Johnson, W. G., & Jensen, G. (1988). A comparison of instruments for assessing competency to stand trial. *Law and Human Behavior, 12*, 313-321.

Nunnally, J. C. (1978). *Psychometric theory (2nd ed.).* New York: McGraw Hill.

Nussbaum, D., Mamak, M., Tremblay, H., Wright, P., & Callaghan, J. (1998). The METFORS Fitness Questionnaire (MFQ): A self-report measure for screening competency to stand trial. *American Journal of Forensic Psychology, 16*, 41-65.

Ohayon, M. M., Crocker, A., St-Onge, B., & Caulet, M. (1998). Fitness, responsibility, and judicially ordered assessments. *The Canadian Journal of Psychiatry*, *43*, 491-495.

Otto, R. K., Poythress, N. G., Nicholson, R. A., Edens, J. F., Monahan, J., Bonnie, R. J., Hoge, S. K., & Eisenberg, M. (1998). Psychometric properties of the MacArthur Competence Assessment Tool-Criminal Adjudication. *Psychological Assessment, 10*, 435-443.

Pfeiffer, E., Einstein, R. B., & Dabbs, E. G. (1967). Mental competency evaluation for the federal courts: I. Methods and results. *Journal of Nervous & Mental Disease, 144*, 320-328.

Pinals, D. A., Tillbrook, C. E., & Mumley, D. L. (2006). Practical application of the MacArthur Competence Assessment Tool-Criminal Adjudication (MacCAT-CA) in a public sector forensic setting. *Journal of the American Academy of Psychiatry and the Law, 34,* 179-188.

Pirelli, G. (2008). The use of arbitrary metrics in competence to stand trial assessment instruments. Unpublished doctoral thesis.

Poythress, N. G., Hoge, S. K., Bonnie, R. J., Monahan, J., Eisenberg, M., & Feucht-Haviar, T. (1998). The competence-related abilities of women criminal defendants. *Journal of the American Academy of Psychiatry and the Law, 26,* 215-222.

Quinsey, V. L., Pruesse, M., & Fernley, R. (1975). A follow-up of patients found unfit to stand trial or not guilty because of insanity. *The Canadian Psychiatric Association Journal, 20,* 461-467.

Randolph, J. J., Hicks., T., & Mason, D. (1981). The Competency Screening Test: A replication and extension. *Criminal Justice and Behavior, 8*, 471-482.

Randolph, J. J., Hicks, T., Mason, D., & Cuneo, D. J. (1982). The Competency Screening Test: A validation study in Cook County, Illinois. *Criminal Justice & Behavior, 9,* 495-500.

Reich, J., & Wells, J. (1985). Psychiatric diagnosis and competency to stand trial. *Comprehensive Psychiatry, 26*, 421-432.

Reich, J., & Wells, J. (1986a). Defendants with repeated competency evaluations. *Journal of Nervous and Mental Disease, 174*, 120-122.

Riley, S. E. (1998). Competency to stand trial adjudication: A comparison of female and male defendants. *Journal of the American Academy of Psychiatry and the Law, 26*, 223-240.

Rindskopf, D. M. (1997). Testing "small," not null, hypotheses: Classical and Bayesian approaches. In L. L. Harlow, S. A. Mulaik, and J. H. Steiger (Eds.), *What if there were no significance tests?, pp. 319-332.* Mahwah, NJ, US: Lawrence Erlbaum Associates Publishers.

Robbins, E., Waters, J., & Herbert, P. (1997). Competency to stand trial evaluations: A study of actual practice in two states. *Journal of the American Academy of Psychiatry and the Law, 25,* 469-483.

Robertson, R. G., Gupton, T., McCabe, S. B., & Bankier, R. G. (1997). Clinical and demographic variables related to 'fitness to stand trial' assessments in Manitoba. *The Canadian Journal of Psychiatry, 42*, 191-195.

Robey, A. (1965). Criteria for competency to stand trial. A checklist for psychiatrists. *American Journal of Psychiatry, 122*, 616-623.

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

Roesch, R., Eaves, D., Sollner, R., Normandin, M., & Glackman, W. (1981). Evaluating fitness to stand trial: a comparative analysis of fit and unfit defendants. *International Journal of Law & Psychiatry, 4*, 145-157.

Roesch, R., Ogloff, J. R. P., Hart, S. D., Dempster, R. J., Zapf, P. A., & Whittemore, K. E. (1997). The impact of Canadian criminal code changes on remands and assessments of fitness to stand trial and criminal responsibility in British Columbia. *The Canadian Journal of Psychiatry, 42*, 509-514.

Roesch, R., Zapf, P. A., Golding, S. L., & Skeem, J. L. (1999). Defining and assessing competency to stand trial. In A. K. Hess & I. B. Weiner (Eds.), *The handbook of forensic psychology* (2nd ed., pp. 327-349). New York: John Wiley & Sons, Inc.

Rogers, R., Gillis, J. R., McMain, S., & Dickens, S. E. (1988). Fitness evaluations: A retrospective study of clinical, criminal, and sociodemographic characteristics. *Canadian Journal of Behavioural Science Revue canadienne des Sciences du comportement, 20*, 192-200.

Rogers, R., Sewell, K. W., Grandjean, N. R., & Vitacco, M. (2002). The detection of feigned mental disorders on specific competency measures. *Psychological Assessment, 14*, 177-183.

Rogers, R., Ustad, K. L., Sewell, K. W., & Reinhardt, V. (1996). Dimensions of incompetency: A factor analytic study of the Georgia court competency test. *Behavioral Sciences & the Law, 14*, 323-330.

Rosenthal, R. (1979). The "file drawer problem" and tolerance for null results. *Psychological Bulletin, 86,* 638-641.

Rosenthal, R. (1994). Parametric measures of effect size. In H. Cooper and L. V. Hedges (Eds). *The handbook of research synthesis (pp.231-260).* New York: Russell Sage Foundation.

Ryba, N. L. (2005). *An evaluation of the influence of psychiatric symptoms and cognitive abilities on competence-related abilities.* Dissertation Abstracts International, 65 (12-B), 6672.

Ryba, N. L., Cooper, V. G., & Zapf, P. A. (2003). Juvenile competence to stand trial evaluations: A survey of current practices and test usage among psychologists. *Professional Psychology: Research and Practice, 34,* 499-507.

Sachsenmaier, S. J. (1991). *The relationship between traditional psychometric test scores and the determination of criminal responsibility and competency to stand trial.* Dissertation Abstracts International, 52 (1-B), 566.

Sánchez-Meca, J., Marín-Martínez, F., & Chacón-Mascoso, S. (2003). Effect-size indices for dichotomized outcomes in meta-analysis. *Psychological Methods, 8,* 448-467.

Sandercock, P. (1989). The odds ratio: A useful tool in neurosciences. *Journal of Neurology, Neurosurgery, and Psychiatry, (52),* 817-820.

Schreiber, J. (1982). Professional judgment in the assessment of competency to stand trial: Report of an evaluation study. *International Journal of Law and Psychiatry, 5*, 331-340.

Schreiber, J., Roesch, R., & Golding, S. L. (1987). An evaluation of procedures for assessing competency to stand trial. *Bulletin of the American Academy of Psychiatry & Law, 15,* 187-203.

Schumacker, R. E. (2005). Effect size and confidence intervals in general linear models for categorical data analysis. *Multiple Linear Regression Viewpoints, 31,* 42-45.

Shatin, L., & Brodsky, S. H. (1979). Competency for trial: The Competency Screening Test in an urban hospital forensic unit. *The Mount Sinai Journal of Medicine, 46,* 131-134.

Shields, C. L. (2005). *Competence to stand trial: Cognitive-related abilities and other factors.* Dissertation Abstracts International, 66 (2-B), 1186.

Siegel, A. M., & Elwork, A. (1990). Treating incompetence to stand trial. *Law and Human Behavior, 14*, 57-65.

Sikorski, C. R., & Benedek., E. P. (1977). Competency to stand trial: The female offender. *Journal of the American Medical Women's Association, 32*, 150-154.

Simon, M. J. (1987). Use of the Proverbs Test in the assessment of competency to stand trial. *Psychological Reports, 60*, 1166.

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

Skeem, J., Golding, S. L., & Emke-Francis, P. (2004). Assessing adjudicative competency: Using legal and empirical principles to inform practice. In W. T. Donohue & E. R. Levensky (Eds.). *Forensic psychology: A handbook for mental health and legal professionals* (pp. 175-211). New York: Academic Press.

Stone, A. (1975). *Mental health and law: A system in transition*. (DHEW Pub. No. ADM 75-176). Rockville, MD: National Institute of Mental Health.

Stoops, R., Hess, J., Scott, T., Tolan, J., Gallucci, G., & Brierley-Bowers, P. (2007). Training competency to stand trial in an individual with intellectual disability and behavioral health concerns. *Mental Health Aspects of Developmental Disabilities, 10*, 47-52.

Strupp, H. H. (1981). Clinical research, practice, and the crisis of confidence. *Journal of Consulting and Clinical Psychology, 49*, 216-219.

Tillbrook, C. E. (2001). *Competency to proceed: A comparative appraisal of approaches to assessment.* Dissertation Abstracts International, 61 (9-B), 5009.

United States v. Lawrence, 26 F. Cas. 887 D.C. Cir. (1899).

Vann, C. R., & Morganroth, F. (1965). The psychiatrist as judge: A second look at the competence to stand trial. *University of Detroit Law Journal, 43,* 1-12.

Viljoen, J. L., Roesch, R., & Zapf, P. A. (2002). An examination of the relationship between competency to stand trial, competency to waive interrogation rights, and psychopathology. *Law and Human Behavior, 26,* 481-506.

Viljoen, J. L., & Zapf, P. A. (2002). Fitness to stand trial evaluations: A comparison of referred and non-referred defendants. *International Journal of Forensic Mental Health, 1*, 127-138.

Vitacco, M. J., Rogers, R., Gabel, J., & Munizza, J. (2007). An evaluation of malingering screens with competency to stand Trial patients: A known-groups comparison. *Law and Human Behavior, 31,* 249-260.

Warren, J. I., Murrie, D. C., Stejskal, W., Colwell, L. H., Morris, J., Chauhan, P., et al. (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.

Webster, C. D., Menzies., R. J., & Jackson, M. A. (1982). Clinical assessment before trial. Toronto, Ontario, Canada: Butterworths.

Winick, B. J. (1985). Restructuring competency to stand trial. *UCLA Law Review, 32,* 921-985.

Winick, B. J. (1996). Incompetency to proceed in the criminal process: Past, present, and future. In D. B. Wexler & B. J. Winick (Eds.), *Law in a therapeutic key: Developments in therapeutic jurisprudence* (pp. 77-111). Durham, NC: Carolina Academic Press.

Youtsey v. United States, 97 F. 937, 940 (6th Cir.1899).

Zapf, P. A. (1999). *An investigation of the construct of competence in a criminal and civil context: A comparison of the fit, the MacCAT-CA, and the MacCAT-T. (adjudicative competence, treatment competence, Macarthur competence assessment tool).* Dissertation Abstracts International, 60 (6-B), 3012.

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.

Zapf, P. A., & Roesch. R. (2009). *Guide to best practices for forensic mental health assessments: Competency to stand trial*. New York: Oxford.

Zapf, P. A., & Viljoen, J. L. (2003). Issues and considerations regarding the use of assessment instruments in the evaluation of competency to stand trial. *Behavioral Sciences and the Law, 21*, 351-367.