

## VITA

Michael Bruce Lister

### EDUCATION

- 2010          Indiana State University, Terre Haute, Indiana  
Psy.D. in Clinical Psychology
- 2006          Indiana State University, Terre Haute, Indiana  
M.S. in Clinical Psychology
- 2004          Carleton University, Ottawa, Ontario, Canada  
B.A. in Psychology
- 2002          Carleton University, Ottawa, Ontario, Canada  
B.A. in Criminology

### PROFESSIONAL EXPERIENCE

- 2009-10      Eastern State Hospital, Williamsburg, Virginia  
Psychology Associate
- 2008- 09      Eastern Virginia Medical School, Norfolk, Virginia  
Psychology Intern
- 2006-08      Hamilton Center, Inc., Terre Haute, Indiana  
Therapist
- 2006-07      Wabash Valley Correctional Facility, Carlisle, Indiana  
Mental Health Professional
- 2005-07      Indiana State University Psychology Clinic, Terre Haute, Indiana  
Graduate Clinician

A COMPARISON OF THE VIOLENCE RISK APPRAISAL GUIDE, PSYCHOPATHY  
CHECKLIST-REVISED, AND CHILD AND ADOLESCENT TAXON SCALE:  
PREDICTIVE UTILITY AND CROSS CULTURAL GENERALIZABILITY

---

A dissertation

Presented to

The College of Graduate and Professional Studies

Department of Psychology

Indiana State University

Terre Haute, Indiana

---

In Partial Fulfillment

of the Requirements for the Degree

Doctor of Psychology

---

by

Michael Bruce Lister

August 2010

Keywords: Child and Adolescent Taxon Scale, Violence Risk Appraisal Guide, Psychopathy  
Checklist Revised

**COMMITTEE MEMBERS**

Committee Chair: Michael J. Murphy, Ph.D.

Professor, Department of Psychology

Indiana State University, Terre Haute, Indiana

Committee Member: Jennifer L. Boothby, Ph.D.

Associate Professor, Department of Psychology

Indiana State University, Terre Haute, Indiana

Committee Member: Kevin P. Bolinsky, Ph.D.

Assistant Professor, Department of Psychology

Indiana State University, Terre Haute, Indiana

**ABSTRACT**

The Violence Risk Appraisal Guide (VRAG) is a widely utilized measure for estimating the risk of violent reoffending among forensic populations. However, completing the VRAG can be a lengthy process as it requires entering scores from a second test, the Psychopathy Checklist Revised (PCL-R), which must be administered separately and requires hours to complete. In order to reduce scoring time, the authors of the VRAG have developed a brief checklist, the Child and Adolescent Taxon Scale (CATS), which can be used in place of the PCL-R as a more efficient method of assessing psychopathy (Quinsey et al., 2004). Previous research has shown the CATS can identify antisocial individuals and yields similar VRAG risk estimates when substituted for the PCL-R (Glover et al., 2002; Quinsey et al., 1998). However, these investigations employed predominantly Caucasian samples, and evidence supporting the validity of the CATS with ethnically diverse populations is presently lacking. This dissertation research addressed these concerns by examining the predictive utility and cross-cultural generalizability of VRAG scores calculated using the CATS with a more racially diverse sample of forensic psychiatric patients. In addition, the utility of the CATS as a stand-alone measure of psychopathy was examined. The relationship between CATS, VRAG, and PCL-R scores was assessed, and the instruments were compared in terms of their ability to predict the length of time African American and Caucasian patients were treated in a maximum security hospital before being approved for a transfer to a less restrictive setting. As expected VRAG probability estimates for recidivism did not differ depending on whether the CATS or the PCL-R was used

as the index of psychopathy. In addition, the CATS showed good concurrent validity with the PCL-R, and no significant race related scoring differences were observed. Finally, the CATS was the only risk assessment measure able to predict the length of time before participants were approved for transfer to a less restrictive setting. Findings are discussed in terms of the implications for clinical-forensic practice.

## **ACKNOWLEDGMENTS**

Without the guidance and support of several individuals this research project would not have been possible. I would like to express thanks to my dissertation chair and committee members for their assistance designing and carrying out this investigation. I also want to recognize Dr. Donna Cairns and the staff at Eastern State Hospital for their advice and encouragement throughout this process. Their support was vital to the implementation and completion of this project. Special thanks are warranted to my parents, Bruce and Pauline Lister, who have been my backbone and heart through life's journeys. Mom and Dad, none of this would have been possible without your easy wisdom, ready encouragement, and unyielding support. Finally, I would also like to thank a very special former postdoctoral student for helping me "find data."

## TABLE OF CONTENTS

<b>COMMITTEE MEMBERS</b> .....	ii
<b>ABSTRACT</b> .....	iii
<b>ACKNOWLEDGMENTS</b> .....	v
<b>LIST OF TABLES</b> .....	ix
<b>Introduction</b> .....	1
<b>Unstructured Clinical Judgment</b> .....	3
<b>Actuarial Assessment</b> .....	5
<b>Violence Risk Appraisal Guide</b> .....	6
<b>Psychopathy and the PCL-R</b> .....	9
<b>Psychopathy and Race</b> .....	14
<b>Psychopathy: Discrete Class or Antisocial Continuum</b> .....	17
<b>The Child and Adolescent Taxon Scale</b> .....	19
<b>Risk Assessment With the CATS</b> .....	22
<b>Summary of Published Research on the CATS</b> .....	30
<b>Research Examining the Forensic Review Panel Process</b> .....	31
<b>The Present Study</b> .....	34
<b>Method</b> .....	41
<b>Design</b> .....	41
<b>Power Analysis</b> .....	41

<b>Participants</b> .....	42
<b>Measures</b> .....	42
<b>Procedure</b> .....	44
<b>Missing Data</b> .....	46
<b>Results</b> .....	47
<b>Descriptive Data and Demographic Characteristics of Participants</b> .....	47
<b>Correlations Between Risk Assessment Instruments and Demographic</b>	
<b>Variables</b> .....	50
<b>Relationships Between Predictor and Criterion Variables</b> .....	53
<b>Comparison of VRAG-C and VRAG-P Scores</b> .....	54
<b>Predicting the Time Spent in Maximum Security</b> .....	57
<b>Discussion</b> .....	62
<b>Regression Results</b> .....	65
<b>Predictive Power of the Demographic Variables</b> .....	69
<b>Implications</b> .....	71
<b>Limitations to the Current Study</b> .....	75
<b>Future Directions</b> .....	77
<b>REFERENCES</b> .....	79
<b>APPENDIX A: DEMOGRAPHIC DATA FORM</b> .....	87
<b>APPENDIX B: VIOLENCE RISK APPRAISAL GUIDE (VRAG) – QUINSEV,</b>	
<b>HARRIS, RICE, &amp; CORMIER, 1998</b> .....	88
<b>APPENDIX C: PSYCHOPATHY CHECKLIST – REVISED (PCL-R 2<sup>nd</sup> ed.) –</b>	
<b>HARE (2003)</b> .....	90



**APPENDIX D: CHILD AND ADOLESCENCE TAXON SCALE (CATS) –**

**HARRIS, RICE, & QUINSEY (1994).....91**

## LIST OF TABLES

Table 1 <i>Range, Measures of Central Tendency, and Variance for Demographic and Predictor Variables</i> .....	48
Table 2 <i>Pearson Correlations for Demographic and Predictor Variables</i> .....	51
Table 3 <i>Pearson Correlations for Demographic and Predictor Variables for Each Racial Group</i> .....	55
Table 4 <i>Summary of Forced Entry Regression Analysis Predicting the Duration of Maximum Security Treatment With the Risk Assessment Measures</i> .....	57
Table 5 <i>Summary of Hierarchical Regression Predicting the Duration of Maximum Security Treatment With Demographic Variables and Risk Assessment Measures</i> .....	59
Table 6 <i>Summary of Hierarchical Regression Predicting the Duration of Maximum Security Treatment With Select Demographic Variables, CATS, and PCL-R</i> .....	61

## CHAPTER 1

### **Introduction**

The insanity defense has been described as criminal law's most controversial issue (Melton, Petrila, Poythress, & Slobogon, 1997). The defense rests on the moral premise that mentally disordered offenders should not be found guilty and punished like other criminals if they did not understand the consequences of their actions or were not able to control their behavior at the time they committed their offense (Melton et al., 1997). According to legal doctrine, individuals judged not guilty by reason of insanity should be handled compassionately and receive proper treatment rather than imprisonment. However, due in large part to media coverage of high profile insanity acquitees such as John Hinckley, who attempted to assassinate President Ronald Reagan, the general public is increasingly dissatisfied with the insanity defense and how they believe insanity acquitees are typically handled by the criminal justice system (Melton et al., 1997). Consequently, the appropriate treatment of individuals found not guilty by reason of insanity has attracted a great deal of judicial, political, and media attention.

It has been suggested that much of the controversy surrounding the treatment of mentally disordered offenders may result from common misunderstandings about the insanity defense and how individuals found not guilty by reason of insanity are managed (Melton et al., 1997). The general public overestimates the percentage of cases where the insanity defense is raised and successfully employed, as well as the dangerousness of mentally disordered individuals (Melton

et al., 1997). It has also been shown that the general public believes those acquitted by reason of insanity are routinely released from custody either immediately or shortly after their trials, when in actuality, insanity acquitees tend to spend more time in custody than offenders found guilty of a similar crime (Melton et al., 1997).

Rather than being immediately released from custody, as is the case with conventional not guilty verdicts, insanity acquitees are usually automatically committed to a forensic psychiatric hospital for a thirty to sixty day period of evaluation (Melton et al., 1997). Following this initial period of hospitalization an evaluation is submitted to the court and a judge decides whether further commitment to a psychiatric hospital is warranted (Melton et al., 1997). Those insanity acquitees who are committed following the initial forensic evaluation are usually hospitalized for a period lasting between six months and two years, after which they undergo subsequent hearings to determine whether they continue to need hospitalization.

The continued hospitalization of an insanity acquittee hinges on whether or not the individual in question continues to be mentally ill and dangerous (Melton et al., 1997). In most jurisdictions there is no limit placed on the length of time that individuals found not guilty by reason of insanity can be hospitalized, provided they continue to meet these criteria. However, as is the case with individuals who require psychiatric treatment more generally, it is generally held that the state should not unduly infringe on the liberties of insanity acquitees and ensure that they receive treatment in the least restrictive setting necessary (Melton et al., 1997). Thus, appropriately managing the treatment of insanity acquitees is inextricably related to the perceived risk associated with their release and their ongoing security needs. The decision-making process about whether an insanity acquittee requires continued hospitalization or whether their current

treatment setting is more restrictive than necessary is largely based on ongoing evaluations of the potential danger they pose to self and others.

The accurate assessment of dangerousness is therefore of central importance to the treatment of individuals adjudicated not guilty by reason of insanity regardless of the specific discharge process employed in their jurisdiction. Consequently, mental health professionals who work with insanity acquittees routinely conduct risk assessments in order to evaluate their patients' potential dangerousness, ongoing security needs, and readiness for discharge. The merit of these assessments rests on the assumption that those conducting the evaluation are able to come to accurate conclusions regarding the danger posed by a given individual. The section that follows will briefly review 1) the accuracy of unstructured clinical and actuarial risk assessment procedures, and 2) the development, strengths, and limitations of two of the most commonly utilized measures for assessing an individual's risk for future violence. Areas in need of further research and the potential benefits of an alternative approach to risk assessment with these measures will also be discussed.

### **Unstructured Clinical Judgment**

The widespread concern about the dangerousness of mentally disordered offenders who are released from custody has drawn attention and research into how mental health professionals working with forensic populations go about assessing an individual's risk for future violence (Vasey, Kotov, Frick, & Loney, 2005). The scientific literature on risk assessment suggests that evaluating dangerousness is neither a simple task nor something that mental health professionals generally do very well. Historically, mental health professionals have tended to mainly rely on unstructured clinical judgment when conducting risk assessments. That is, rather than using structured empirically derived procedures, clinicians estimate risk based upon variables he or she

believes are associated with the potential for future violence (Wrightsmann, 2001). When conducting unstructured assessments, clinicians may base their estimation of the potential for future violence on data such as demographic information, past antisocial behavior, current symptoms, and clinical impressions. Whereas unstructured clinical judgment allows the clinician to take into account a wide range of data and apply these variables on a case-by-case basis, this approach has been criticized for being highly subjective, overly impressionistic, and lacking consistency (Grove & Meehl, 1996).

Previous research examining the accuracy of unstructured clinical risk assessment indicates that the predictions made by clinicians lack consistency and are no better than predictions made by judges, lay persons, or even chance (Grove & Meehl, 1996; Hilton & Simmons, 2001). The accuracy of unstructured clinical judgment has been shown to be hindered by the tendency among clinicians to unduly stress the importance of extreme scores, inadequately account for base rates, and generally overestimate the likelihood of future violence (Quinsey, Harris, Rice, Cormier, 1998). It has also been suggested that clinicians have difficulty accurately accounting for interrelationships among risk factors, measurement error, and tend to emphasize frequencies rather than probabilities when interpreting data (Quinsey et al., 1998). Research has also shown that the weight given to different risk factors when developing opinions varies across clinicians and according to personal biases (Grove & Meehl, 1996). Furthermore, the amount of clinical training, experience, or the degree of confidence a clinician has in their opinion, has been shown to be unrelated to the accuracy of their predictions (Quinsey et al., 1998). On the basis of these results, many have questioned whether risk assessments based on unstructured clinical judgment should serve as the basis for depriving an individual of their freedom or if they are sufficiently valid to be admissible in court (Quinsey et al., 1998).

However, despite these criticisms, unstructured clinical judgment remains the most common method of conducting risk assessments (Vasey et al., 2005).

### **Actuarial Assessment**

Given the questions surrounding the validity and reliability of predictions based on unstructured clinical judgment, leading researchers have called for greater utilization of actuarial approaches to risk assessment (Grove & Meehl, 1996; Quinsey et al., 1998). Actuarial risk assessment measures systematically incorporate factors that have been shown to be associated with violence and recidivism with forensic populations in general, and assess for these factors among individuals in a standardized manner (Hilton & Simmons, 2001). Actuarial assessments are developed using statistical procedures and employ probabilities to estimate an individual's potential for future violence (Grove & Meehl, 1996; Quinsey et al., 1998). Thus, rather than relying on human judgment, actuarial assessments utilize an empirically derived decision-making process to estimate likelihood of future violence.

A meta-analysis examining the published research comparing unstructured clinical to actuarial predictions have repeatedly found actuarial predictions to be more accurate (Grove et. al, 2000). Overall, actuarial predictions outperformed unstructured clinical predictions by ten percent, and of the 130 studies included in the review, 40% concluded that actuarial approaches yield substantially more accurate predictions (Grove et. al, 2000). Reviewers of the extant body of research on risk assessment have suggested that unstructured clinical assessments are of little value and characterized further debate surrounding the predictive accuracy of clinical versus actuarial risk assessment as unnecessary (Hilton & Simmons, 2001).

Over the past two decades numerous actuarial risk assessment measures have been developed for predicting the likelihood of violent reoffending among forensic psychiatric and

offender populations. However, despite the increasing popularity and superior predictive efficacy of actuarial risk assessment measures, the use of these instruments has yet to become standard clinical practice (Vasey et al., 2005). Moreover, a number of potential drawbacks have been associated with the use of actuarial risk assessments including the added time and financial costs of employing these measures and continued questions surrounding their cross-cultural generalizability (Bolton, 2006; Quinsey et al., 1998). As a result, research continues to be directed towards developing more concise and cross-culturally valid approaches to actuarial risk assessment (Bolton, 2006). This dissertation research adds to this body of scientific literature by examining the predictive utility and cross-culturally generalizability of two measures of psychopathy as well as a more concise approach to calculating one of the most commonly employed actuarial measures, the Violence Risk Appraisal Guide (VRAG).

### **Violence Risk Appraisal Guide**

The VRAG (Harris et al. 1993) is currently regarded as one of the most accurate measures for determining which mentally disordered offenders are at the greatest risk for committing new violent or sexual crimes (Quinsey et al., 2004). The instrument was originally developed from data collected through a series of follow-up studies conducted with maximum security forensic psychiatric patients in Ontario, Canada. These investigations examined the association between various risk factors and instances of violent reoffending in order to develop a means of estimating the likelihood of violent recidivism (Glover et al., 2002; Quinsey et al., 1998). A total of 50 empirically supported risk factors for violent recidivism, psychiatric rehospitalization, and general criminal behavior were included in their investigations. Analyses were conducted to identify those risk factors which had the greatest predictive power and which showed a significant univariate relationship to violence (Quinsey, Harris, Rice, & Cormier,



2006). Ultimately, 12 of the 50 items examined were selected for inclusion in the final instrument.

Items assessed by the VRAG include, but are not limited to, elementary school maladjustment, the presence of a DSM-III personality disorder, separation from biological parents before age 16, and previous conditional release violations (Appendix B; Quinsey et al., 2006). Each of the final items was subsequently assigned a weighting corresponding to its relative predictive power. When totaled, VRAG scores range from -26 to 38 and correspond to 7 and 10 year probability estimates for violent recidivism (Quinsey et al., 1998).

The predictive validity of VRAG scores was assessed using Relative Operating Characteristic (ROC). The ROC was calculated by plotting the hit rate for violent recidivism (true positives) as a function of the false alarm rate (false positives). The area under the plotted curve thus corresponds to the probability of randomly choosing a recidivist scoring higher on the VRAG than a nonrecidivist chosen at random. The area under the ROC curve for VRAG predicted future violence was determined to be .76 for the five year follow-up period examined, indicating a large effect size.

In the time since the initial publication of test development data, a significant body of literature has accrued supporting the use of the VRAG for the prediction of violence with both forensic psychiatric and offender populations (Quinsey, Book, & Skilling, 2004). Research replicating and extending the results from test development has demonstrated that VRAG scores were not only associated with the probability of a subsequent violent act occurring, but also the severity of the act and how rapidly the individual recidivated (Harris, Rice, & Cormier, 2002). According to the authors of the VRAG, the variables associated with subsequent acts of violence were highly consistent across various populations (Quinsey et al., 1998). Studies examining the

predictive utility of the VRAG with sex offender populations and employing ten year follow-up periods have yielded ROC ranges from .73 to .77 (Quinsey et al., 2006). Research has also demonstrated that the VRAG shows promise for assessing the security needs of individuals being held in forensic institutions, as higher VRAG scores have been associated with violent behavior among psychiatric in-patients (Doyle, Dolan, & McGovern, 2002).

The VRAG has been shown to more accurately identify violent recidivists than other leading risk assessment instruments. Quinsey et al. (1998) cite research examining the predictive accuracy of the VRAG, Psychopathy Checklist-Revised (PCL-R) and Historical Clinical Risk Management-20 scale (HCR-20) in a follow-up study examining violent recidivism among discharged forensic psychiatric patients in Belgium. Results indicated ROC areas of .85 for the VRAG and PCL-R, and .80 for the HCR-20, indicating that the VRAG meets or exceeds the predictive accuracy of other widely utilized risk assessment instruments.

In addition to yielding accurate predictions, the VRAG has also demonstrated excellent psychometric characteristics. Scoring the VRAG is relatively straightforward provided the evaluator has access to adequate records, and the measure has been shown to yield valid results even when scored with missing data (Vasey et al., 2005). Multiple investigations have reported excellent interrater reliability with  $r_s > .95$  reported, and the VRAG has demonstrated concurrent validity with other risk assessment measures (Glover et al., 2002; Quinsey et al., 2006).

Of all the risk factors examined by Harris et al. (1993), the presence of psychopathy (a construct which will be described in greater detail below) was the single best predictor of continued violence. Accordingly, a high psychopathy score is the most heavily weighted item when calculating the VRAG. However, psychopathy scores must be calculated separately through the administration of an additional measure. This requirement greatly increases the time

and cost associated with completing the VRAG. Consequently, researchers have proposed reducing the costs associated with VRAG administration by developing a more concise method of assessing psychopathy (Quinsey et al., 1998). This development is not only significant in terms of the VRAG, but the field of risk assessment more generally.

As the following section will illustrate, the development of a parsimonious and efficacious means of assessing psychopathy has been central to both research and clinical practice in forensic psychology for the past three decades. The present investigation seeks to add to this body of literature by assessing the concurrent validity, predictive utility, and cross-cultural generalizability of two separate psychopathy measures that can be employed to score the VRAG or as stand-alone indices of psychopathy. The section below will describe the concept of psychopathy in greater depth, detail how it is commonly measured, and delineate areas of ongoing research surrounding the construct and its measurement. The literature on the widely utilized PCL-R, and a less widely utilized, but potentially more efficient alternative will be discussed in detail in order to provide context and rationale for the research that follows.

### **Psychopathy and the PCL-R**

The concept of psychopathy has become an integral aspect of risk assessment in clinical forensic settings. Psychopathy is a clinical syndrome that is widely described according to a distinct cluster of personality traits and related interpersonal tendencies first described by Hervey Cleckley in 1941 (Cleckley, 1972). Troubled by the disproportionate social and institutional harm caused by a population of criminals who shared a similar clinical profile, he first published his seminal work on the subject, *The Mask of Sanity*, in 1941. Cleckley hoped to facilitate the identification of psychopathic individuals, increase awareness of the syndrome, and spur greater academic and clinical attention (Cleckley, 1972). Unlike the VRAG and other actuarial

measures which were exclusively developed to predict violence, the construct of psychopathy offers a theoretical explanation for the persistent violence evidenced by certain individuals.

Cleckley characterized the *psychopath* as neither mentally defective nor psychotic, and provided 16 specific traits he believed to be diagnostic of the syndrome.

Cleckley felt psychopaths could be distinguished from other criminals by characteristics such as social charm, the absence of nervousness, poverty of affective reactions, egocentricity, and the lack of remorse or shame. According to Cleckley, these personality dimensions predispose psychopathic individuals to persistent antisocial conduct (Cleckley, 1972). Presently, there is mounting research evidence suggesting that diagnostic models that adhere to Cleckley's original description of psychopathic personality disorder are able to classify these individuals with the greatest precision (Hare et al., 1991).

However, while Cleckley's description of the syndrome may better conceptualize the qualitatively distinct antisociality of psychopathic individuals, the validity of diagnoses derived solely from his description has been questioned. This criticism has mainly centered on the limitations inherent to any unguided clinical estimation of personality traits, and is corroborated by the heterogeneous population of individuals that have been diagnosed via Cleckley's descriptions (Skilling, Harris, Rice, & Quinsey, 2002).

The 22-item Hare Psychopathy Checklist (PCL; 1980) and 20-item Psychopathy Checklist Revised (PCL-R; 1991) are regarded as the first measures to successfully operationalize Cleckley's criteria into a psychometrically sound assessment tool (Hare, 1993; Hare et al., 1991; Serin, 1992). The major motivation behind the development of the instrument was to move beyond self-report measures of antisocial traits and develop a structured, user friendly assessment measure based upon the description of psychopathy delineated by Cleckley.

Hare and colleagues first developed and validated the PCL with a sample of 229 Canadian inmates. A team of clinicians familiar with clinical descriptions of psychopathy and experienced in working with forensic populations were provided with the psychopathy checklist and asked to rate the extent that offenders matched the description through detailed interview and file review (Hare, 1993). The rationale behind this two step process was to corroborate and add to the often problematical self-report information gleaned from forensic populations. Interrater reliability for PCL scores ranged from .82 to .93, test-retest coefficients were .89, and they showed concurrent validity to DSM III ASPD diagnoses (Hare, 1985; Schroeder, Schroeder, & Hare, 1983).

The Psychopathy Checklist was modified somewhat as a consequence of subsequent research. Scoring criteria were modified and made more stringent. These revisions led to the development and publication of the PCL-R. Two items included in the original PCL, “previous diagnosis as a psychopath or similar” and “drug and alcohol use not direct cause of antisocial behavior” were eliminated from the revised instrument because they were difficult to accurately assess and had relatively low correlations with total scores (Hare et al, 1990). Despite these changes the revised PCL was determined to measure the same construct as the original scale and demonstrated excellent psychometric characteristics. At present the PCL-R is widely recognized as the “gold standard” for the assessment of psychopathy (Acheson, 2005).

PCL-R items are typically scored by clinicians through both the administration of a structured interview and systematic review of case or institutional file information (Hare, 1993). To reduce the potential for confounds associated with inferring the various trait dimensions; exemplars illustrating each characteristic were provided, and all items are scored on a 3-point scale according to how well they apply (Hare et al., 1991). The key characteristics assessed by this instrument include both interpersonal traits (e.g. shallow affect, egocentricity, lack of

empathy, and glibbest demeanor), as well as behavior dimensions related to social deviance (e.g. sensation seeking, impulsivity, early behavior problems, adult antisociality (Appendix C). PCL-R scores range from 0-40 according to the degree the individual in question matches the description of a prototypical psychopath, with higher scores reflecting greater psychopathy.

Results from the initial validation study of the PCL-R on 1065 prison inmates and 440 psychiatric patients indicated good interrater reliability for both psychopathic personality ( $r = .73$ ) and social deviance scales ( $r = .83$ ; Hare et al., 1991). In addition, PCL-R scores were found to be positively correlated with a number of related measures including the psychopathic deviate scale from the MMPI, the Antisocial Scale from the Millon Clinical Multiaxial Inventory II, and DSM diagnoses of Antisocial Personality Disorder (ASPD). Hart, Forth and Hare (1989) reported a correlation of .48 between PCL-R psychopathy and DSM-IV derived antisocial personality disorder diagnoses.

The PCL-R also appears to offer incremental validity beyond DSM ASPD in identifying the most persistently antisocial forensic populations. PCL-R based diagnoses of psychopathy have been found to more strongly predict post-release recidivism rates than DSM-III ASPD diagnoses ( $r = .56$  vs.  $.26$ ). Moreover, while 79% of individuals diagnosed with psychopathy also meet the DSM diagnostic criteria for ASPD, only 30% of those diagnosed with ASPD meet the criteria for psychopathy (Hare et al., 1991; Serin, 1992). A follow-up investigation conducted by Hare and colleagues has demonstrated that PCL-R scores are correlated with the revocation of conditional release from prison. Three years after their release date, approximately 75% of psychopaths had been reincarcerated, compared to only 20% of nonpsychopaths.

Similar results demonstrating the effectiveness of the PCL-R in predicting violent recidivism have been consistently demonstrated in multiple empirical investigations employing

the measure (Edens, Poythress, & Lilienfeld, 1999; Hare, 2002; Hemphill, Hare, & Wong, 1998; Tengstrom, Grann, Langstrom, & Kullgran, 2000). Individuals with higher PCL-R scores have not only been shown to reoffend at higher rates, but also tend to commit more violent crimes compared to nonpsychopathic offenders with similar records (Viding, 2004). Research has also shown that PCL-R scores identify prisoners responsible for more violence and rule violations within institutional settings, individuals who demonstrate poor treatment progress, and patients who are less likely to be recommended for discharge by forensic hospital review panels (Edens, Poythress, & Lilienfeld, 1999; Manguno-Mire, Thomson, Bertman-Pate, Burnette, & Thomson, 2007; Morrissey, Mooney, Hogue, Lindsey, & Taylor, 2007; Reiss, Grubin, & Meux, 1999). The demonstrated utility of PCL-R scores in predicting antisocial behavior has encouraged widespread use of the instrument in both North America and European forensic settings (Serin, 1992).

Factor analyses indicate PCL-R items can be grouped into two broad clusters. The first factor has been said to correspond with the interpersonal and affective items (e.g. glibbest presentation, shallow affect) while the second factor tends to reflect exploitive and criminal lifestyle items (e.g. parasitic relationship patterns, criminal versatility; Hare, 2007). More recently, researchers applying confirmatory factor analysis to very large data sets have established a four-factor model also appears to fit the data well. This model clusters PCL-R items according to Interpersonal (e.g. superficial charm) Affective (e.g. shallow affect), Lifestyle (e.g. parasitic relationship patterns), and Antisocial factors (Hare, 2007). According to Hare, the factor structure of the PCL-R has clear clinical implications as descriptions of patients according to their factor and subfacet scores provides a more finely detailed personality profile. Consequently, the most recent edition of the PCL-R incorporates both the two and four factor

descriptions through providing scoring procedures for calculating total, two “factor”, and four “facet” PCL-R scores (Hare & Neumann, 2009).

There is a great amount of research evidence that suggests that the construct of psychopathy, as described by Cleckley and operationalized by Hare, best classifies the persistent and distinct antisocial conduct of psychopathic individuals. Since the publication of the PCL-R a substantial empirical literature on the topic of psychopathy has accrued, and the PCL-R is frequently the instrument of choice for assessing the construct in both basic and applied research. Presently, the PCL-R has virtually become synonymous with the construct of psychopathy and the instrument is by far the most commonly employed means of assessing the construct in both research and institutional settings (Hare, 1993; Hicks, Markon, Patrick, Krueger & Newman, 2004; Viding, 2004).

### **Psychopathy and Race**

The PCL-R was developed using primarily Caucasian samples and the vast majority of initial descriptions of psychopathy were based on case studies of Caucasian patients. Consequently, many have cautioned that the meaning of PCL-R scores and the construct of psychopathy may not generalize to non-Caucasian individuals. While employing any biased psychological instrument can have negative consequences, biases in PCL-R scores could be extremely detrimental given the importance of the decisions made on the basis of risk assessment measures.

These concerns have prompted substantial research into the validity and generalizability of PCL-R scores with non-Caucasian samples. For example, Lynn (2002) suggested base rates for the disorder vary across racial groups. He found the base rates for psychopathy among African American and Native American inmates were higher than among Caucasians.



Moreover, the rate of psychopathy among East-Asian offenders was determined to be less than any of the other ethnic groups examined. However, Lynn's research has since been criticized because he equated psychopathy with antisocial personality disorder, and failed to control for the effects of economic status and intelligence (Skeem, Edens, Sanford, & Colwell, 2003).

Moreover, studies which assess only those individuals who have been assessed as psychopaths with the PCL-R have not found differences in base rates across racial groups (Doninger & Kosson, 2001; Skeem et al., 2003).

While there may not be significant race related differences in base rates for psychopathy as measured by the PCL-R, there is research evidence that the personality dimensions measured by the instrument may vary somewhat across racial groups. First, a number of laboratory findings derived from psychopathic Caucasians do not appear to generalize to psychopathic African Americans. For example, Doninger and Kosson (2001) found that the empathic deficits of Caucasian psychopaths did not replicate with African American subjects. Moreover, Caucasian psychopathic individuals have been shown in laboratory studies to be more resistant to behavior change following punishment; however, this tendency is not as readily observed among African Americans (Thornquist & Zuckerman, 1995). In addition, research employing forensic samples has shown that the pattern of correlation for various dimensions of psychopathy such as impulsivity and a superficially charming, glib presentation differed between African American and Caucasian individuals (Kosson, Smith, & Newman, 1990).

In light of apparent cross-racial differences in key characteristics assessed by the PCL-R, researchers have subjected the instrument to more detailed examinations of its cross-racial generalizability. Cooke, Hart, and Michie (2004) examined the PCL-R for potential racial differences in terms of both the individual items and factor structure using item response theory

and confirmatory factor analysis. PCL-R scores from a total of 359 Caucasian and 356 African American inmates were examined. While subtle differences in item functioning across racial groups were observed, they did not contribute to differences in overall scores or factor structure and were judged to not have any discriminating power. On the basis of their findings the authors suggested that any extant item level differences likely cancel each other out when scores are totaled and concluded that PCL-R scores could be generalized across groups without undue bias.

Toldson (2002) compared key psychometric properties of the PCL-R when used on African American and Caucasian inmates. Similar to Cooke et al. (2004), he found PCL-R item scores to be internally consistent among both Caucasian and African American participants. In addition, total PCL-R scores were associated with the degree of criminal activity and clinical presentations for both groups. However, Toldson also found evidence that some aspects of the PCL-R appear to differ for African American participants. First, he observed that race and PCL-R scores tend to interact, such that African Americans with elevated PCL-R scores received longer sentences than Caucasian offenders with similar scores. While the concern does not indicate a psychometric bias, it does suggest there may be racial biases in how PCL-R scores are employed. In addition, Toldson was unable to replicate the 2-factor structure for PCL-R scores for African American inmates. Although this may partly be due to the relatively small sample size employed ( $N = 251$ ), other investigations have reported problems replicating the PCL-R factor structure with large ethnically diverse samples and further discussion of this finding is warranted.

A number of recent studies employing confirmatory factor analysis have not found strong support for the two factor model for PCL-R scores (Hill et al., 2004; McDermitt et al, 2000 cited in Vitacco, Newmann, & Jackson, 2005). Consequently, there is growing support for three and

four factor PCL-R models. Vitacco et al. (2005) recently explored a four factor model of psychopathy with a sample comprised of 588 Caucasian and 233 African American civil psychiatric patients using confirmatory factor analysis. The four factors assessed correspond with the four “facet scores” (interpersonal, affective, lifestyle, and antisocial factors detailed above) which can be calculated on the most recent version of the PCL-R. Analyses revealed that the four factor model fit the data well and was replicable with male and female, as well as Caucasian and African American participants. Thus, the authors concluded that the four factor model of psychopathy was generalizable across both racial and gender groups and may be a more accurate manner of representing the clusters of PCL-R scores with larger more diverse samples.

Taken together, the psychometric properties of the PCL-R are now well established in a variety of offender and patient populations (Hare & Neumann, 2009). While there may be subtle cross-racial differences in PCL-R profiles, research has consistently shown that these differences do not impact overall scores. Moreover, contradictory findings regarding the generalizability of the factor structure have been addressed through the inclusion of four subfacets (i.e. the four factor model) in the most recent version of the test. However, while the psychometric characteristics of the PCL-R have been shown to be generalizable, potentially damaging biases may exist in how PCL-R scores are employed.

### **Psychopathy: Discrete Class or Antisocial Continuum**

As is the case with behavioral disorders more generally, there exists a great deal of debate surrounding whether psychopathy is best conceptualized as a taxon or dimension. Meehl and Golden (1992) described a taxon as “an entity, type, syndrome, species, disease, or more generally, a non arbitrary class” (p. 127). Accordingly, those who view psychopathy as a taxon

see it as a qualitatively discrete category that distinguishes “psychopaths” from “nonpsychopaths” whereas those who conceptualize it dimensionally see “individuals high in psychopathy” as representing the extreme end of a more quantitatively ordered dimension of antisocial behavior. Historically, psychopaths have been described by clinicians as a discrete class of criminal (e.g. Cleckley) and much of the research on psychopathy refers to the characteristics of the class of individuals who score over 30 on the PCL-R. However, it was not until fairly recently that any research evidence emerged supporting the presupposition that psychopathy is a taxon.

The first published investigation which explored whether psychopathy represented a taxon was conducted by Harris, Rice, and Quinsey (1994). According to the authors, “the simplest procedure to detect a taxon is to examine the distribution of scores on an indicator believed to discriminate between members of a taxon and its complement” (p. 384). Consequently, to find evidence that psychopathy represented a taxon, the distribution of PCL-R scores and numerous additional variables associated with antisociality were examined. Instruments were scored using institutional records from 643 adult mentally disordered offenders. Data were analyzed using multiple taxometric procedures in order to test for the existence of two distinguishable normal distributions, or latent categories, within the dataset (Harris et al., 1994). Similar procedures have been used in previous investigations in order to identify a taxon and its complement within a distribution of scores (Meehl & Golden, 1982). Results revealed strong evidence of a taxon underlying PCL-R factor II scores (Antisocial/Lifestyle) as well as various markers of childhood antisociality. Little support was found for a taxon underlying PCL-R Factor I scores (Interpersonal/Affective) or variables representing adult antisociality.

As part of their analysis Harris et al. (1994) compared the distribution of scores based solely on variables reflecting childhood maladjustment to PCL-R scores. A high degree of agreement between taxometric classifications for individual subjects using either method was observed. On the basis of these findings the authors concluded that childhood behavioral problems were good indicators of psychopathy in later life. Consequently, the eight childhood variables found to most effectively identify members of the psychopathy taxon were employed to develop the eight item Child and Adolescent Taxon Scale (CATS) (Appendix D). Assessing this scale, and the taxon of antisocial individuals it purports to identify, were a major focus of the present investigation.

### **The Child and Adolescent Taxon Scale**

Having developed a scale that can be used to identify individuals with psychopathy, Quinsey et al. (1998) sought to explore whether the CATS could be used as an alternative to the PCL-R when calculating VRAG scores. The authors compared the VRAG ratings scored with the CATS (VRAG-C) to VRAG ratings scored with the PCL-R (VRAG-P) employing a sample of 54 mentally disordered offenders. It was determined that VRAG scores calculated with the PCL-R correlated .98 with VRAG scores calculated with the CATS. Furthermore, replacing the PCL-R with the CATS resulted in nearly identical predictive accuracy performance ( $d = 1.04$ , ROC are = .75).

Harris et al. (1994) proposed that the CATS offered a number of advantages over the PCL-R in identifying psychopathic individuals. Whereas the PCL-R requires scoring 20 separate items across an individual's lifespan, the CATS discriminates taxon members efficiently using just eight items reflecting antisocial and aggressive childhood conduct (Quinsey et al., 1998). Furthermore, scoring the PCL-R typically involves conducting a structured interview that can

take up to two hours to administer. Together the lengthy interview and detailed review of case information needed to score the PCL-R may require several hours to complete. Moreover, the CATS had the added advantage of being a nonrestricted test that can be scored by unlicensed individuals at a lesser cost than the PCL-R. Taken together, scoring the PCL-R may be too demanding and costly for various diagnostic situations and, at least for the purposes of assessing risk for violence, the CATS shows promise as a more parsimonious method of identifying psychopaths (Quinsey et al., 1998). Thus, according to its developers, the CATS offered a number of potential benefits over conducting risk assessments with the PCL-R. However, as the following section will delineate, some controversy surrounds the methodology used to develop the CATS and how the instrument conceptualizes psychopathy.

Recently, there has been increased empirical attention directed at exploring whether a discrete class underlies psychopathy as Harris et al. (1994) observed. With this increased attention has been renewed controversy regarding whether psychopathy is best conceptualized as a taxon (psychopath vs. nonpsychopath) or dimensionally (high versus low psychopathy). Given that the CATS purports to identify those individuals who are members of the psychopathy taxon, this issue merits closer consideration as questions surrounding the taxonicity of psychopathy potentially undermine the instrument's empirical foundation.

A number of recently published investigations have not been able to replicate the findings of Harris et al. (1994) (Edens, Marcuss, Lilienfeld, & Poythress, 2006; Guay, Ruscio, Knight, & Hare, 2007; Walters, Gray, Jackson, Sewell, Rogers, Taylor, & Snowden, 2007). The authors of all three investigations found little support for a discrete class of psychopathic individuals and concluded instead that psychopathy is best represented dimensionally.

These results would appear to undermine the purported ability of the CATS to identify a discrete class of individuals with psychopathy. However, the methodology employed in these studies differed from that which was employed by Harris et al. (1994) in a number of important ways. First, they employed larger, more heterogeneous samples. Second, the taxometric statistics employed in these more recent investigations differed from those employed by Harris et al. (1994). Finally, while Harris et al. (1994) employed multiple measures of psychopathy, the authors of these recent investigations exclusively employed measures of psychopathy that were derived from Hare's conceptualization of the construct (e.g. PCL-R and PCL: SV). It is this last methodological difference that is potentially the most significant, as evidence from other investigations suggests that the discrete class of antisocial individuals identified by Harris et al. (1994) may differ from the psychopathic individuals identified by the PCL-R.

Evidence that the contradictory findings regarding the taxonicity of psychopathy may result from methodological and measurement issues comes from a variety of sources. First, evidence regarding the taxonicity of psychopathy has been shown to be weaker for the interpersonal and affective items of the PCL-R (Factor 1) than for the antisocial and parasitic lifestyle items (Factor 2; Skilling et al., 2002). Second, investigations that have employed child participants or have assessed for childhood behavioral problems have consistently found evidence of a discrete class of antisocial individuals (Guay et al., 2007; Harris et al., 1994; Skilling et al., 2001; Vasey, Kotov, Frick, & Loney, 2005). Third, the base rate for the discrete class of antisocial individuals identified by the CATS, or by assessing for psychopathic tendencies in youth, has been consistently shown to be greater than the estimated based rates of psychopathy (as measured by the PCL-R) in adult samples. Thus, the taxon identified by Harris et al. (1994) and various other researchers appears to differ somewhat from psychopathy as

operationalized by the PCL-R. Consequently, researchers have also referred to the distinct class of antisociality identified in taxonometric research as “life course persistent antisocial behavior” rather than psychopathy (Skilling et al., 2002; Vasey et al., 2005).

In summary, research evidence strongly supports the existence of a taxon underlying psychopathic behavior; however, the evidence is strongest for PCL-R factor 2 items and childhood conduct problems. Further research is needed in order to better clarify the relationship between the taxon assessed by the CATS and psychopathy as measured by the PCL-R. In particular, there is a paucity of research examining the relationship between the CATS and the facets of psychopathy as delineated by the most recent version of the PCL-R. However, given that Factor 2 has been shown to most strongly predict recidivism (Quinsey et al., 1998) and is the exclusive focus of CATS items, the larger debate surrounding whether psychopathy is best conceptualized dimensionally or as a taxon may have little bearing on the utility of the CATS for purposes of risk assessment.

### **Risk Assessment With the CATS**

Skilling, Quinsey, and Craig (2001) attempted to replicate and extend the Harris et al. (1994) investigation using a sample of 1,111 school-age boys. The authors sought to offer further evidence that a taxon underlies persistent antisociality by demonstrating its existence in children. Aside from the desire to demonstrate that a natural class of antisocial individuals could be identified early in life, the investigators also wanted to replicate the Harris et al. (1994) findings with a sample that would yield more generalizable results. Measures utilized in the study included the CATS, a youth version of the psychopathy checklist (PCL:YV) and the DSM-IV conduct disorder diagnosis with criteria scored ordinally as a scale. Taxometric analyses of the distribution of scores for each measure again showed evidence of an underlying taxon. When



each measure was examined separately, base rates for taxon membership were roughly the same (approximately 9% for each dataset). Therefore, analyses of all three measures showed evidence of two distinct groups. Results further revealed that the CATS showed good internal consistency (Alpha = .71) and correlated highly with both PCL-YV total scores and conduct disorder diagnostic criteria. Moreover, seven of the eight CATS items, as compared to 13 of the 18 PCL-YV items, were found to be positive matches with taxon membership. The investigators concluded that measures of problematic behavior in children such as the CATS can identify a distinct class of seriously antisocial boys in early life.

In addition to the Skilling et al. (2001) investigation which replicated findings that a taxon of persistently antisocial individuals could be identified using the CATS, a number of investigations have shown that the CATS can be employed to predict reoffending. Glover et al. (2002) examined the accuracy of several risk assessment measures commonly used to predict recidivism including the VRAG, CATS, and PCL-R. For purposes of comparison, separate VRAG scores were calculated using either the PCL-R (VRAG-P) or the CATS (VRAG-C). Results revealed that most of the predictors examined were significantly correlated with both each other and with recidivism. The mean scores on the CATS, PCL-R, and both versions of the VRAG were significantly higher for recidivists than nonrecidivists. The authors compared the accuracy of each measure by calculating the degree to which scores correlated with recidivism and the probability of recidivists scoring higher than nonrecidivists on each instrument.

Results showed CATS scores were more highly correlated with violent reoffending than PCL-R scores and both measures yielded similar correlations with general reoffending ( $r_s = .19$  and  $.22$  respectively). The probability of an offender scoring higher than a nonoffender on the CATS (.61) was also comparable to the PCL-R (.63). The VRAG was determined to be among

the most accurate of all the measures examined regardless of whether it was calculated with the CATS or the PCL-R. The probability of a recidivist scoring higher than a nonrecidivist on the VRAG-C was .68 as compared to a probability of .72 for the VRAG-P. Finally, interrater reliability for the PCL-R, CATS, and VRAG were all excellent with *rs* for each measure exceeding .90. Taken together, results of the Glover et al. (2002) investigation suggest the CATS can be successfully employed to predict violence and the VRAG is among the most accurate risk assessment measure regardless of whether it is scored with the CATS or the PCL-R as an index of psychopathy.

Bartosh, Garby, Lewis, and Gray (2003) compared the predictive validity of four risk assessment measures commonly employed with sex offender populations. The Static-99, Rapid Risk Assessment for Sex Offender Recidivism (RRASOR), Minnesota Sex Offender Screening Tool (MnSOST-R), and Sex Offender Risk Appraisal Guide (SORAG) were examined. The SORAG (Quinsey et al., 1995) contains most of the same items as the VRAG, but has been adapted to assess factors associated with recidivism among sex offenders. All instruments were scored through reviewing the files of 186 sex offenders released from the Arizona department of corrections. As in the Glover et al. (2002) study, the authors employed the CATS in place of the PCL-R for calculating SORAG totals. ROC analyses were employed to compare scores on risk assessment measures to instances of sexual, violent, or general reoffending. Results indicated the SORAG scored using the CATS was able to predict sexual, violent, and general recidivism. Additionally, the SORAG scored with the CATS was determined to be the strongest overall predictor of violent and general reoffending of all the instruments examined.

In addition to research supporting the use of the CATS in lieu of the PCL-R when assessing risk of sexual or violent recidivism with the VRAG, the CATS has been employed with

other special needs forensic populations. Quinsey et al. (2004) compared various predictors of recidivism including the VRAG-C to rates of reoffending using a sample of 58 men with intellectual disabilities. The authors stated they chose to employ the CATS instead of the PCL-R when calculating VRAG totals because it is a simpler, more concise, and a potentially more appropriate instrument for assessing individuals with mental retardation. Analysis revealed VRAG-C scores exhibited the highest correlation with instances of violent or sexual misbehavior ( $r = .32$ ). Moreover, the predictive accuracy of the VRAG-C as determined by ROC probability was .69. Overall, the VRAG-C proved to be the best predictor of future antisocial behavior.

Folsom and Atkinson (2007) recently examined the predictive utility of the CATS for distinguishing recidivist from nonrecidivist female offenders. A total of 100 Canadian female offenders participated in the study. CATS scores were compared to rates of reoffending (defined as having a conviction of any type) as well as scores on the Level of Service Inventory Revised (LSI-R). The LSI-R is a measure of static and dynamic risk factors that is widely employed in correctional and parole settings. Results indicated that mean CATS scores were significantly higher for recidivists than nonrecidivists, and were significantly associated with both violent as well as general recidivism ( $r_s = .23$  and  $.27$ ). Analyses also revealed the CATS had significant predictive validity with ROC probabilities for violent recidivism of .72. Finally, the CATS demonstrated good concurrent validity, as scores on this measure were highly correlated with LSI-R scores ( $r = .72$ ). On the basis of these findings the authors concluded that the CATS is a reliable and valid measure for distinguishing recidivists from nonrecidivists. However, the authors expressed concern that there may be race-related scoring differences on these instruments and concluded that future research should consider whether the observed results are generalizable across different racial groups.

Taken together, published research investigations employing the VRAG scores calculated with the CATS instead of the PCL-R have consistently been shown to predict both violent and general recidivism in a number of correctional and forensic populations. Research with these populations has also shown the CATS to be a promising stand-alone measure for assessing antisocial tendencies and the likelihood of recidivism. Moreover, the CATS had been employed as a measure of antisociality in numerous laboratory-based research initiatives. Several published investigations have employed the CATS to explore the relationship between antisocial behavior, aggression, and mating strategies (Lalumiere, Chalmers, Quinsey, & Seto, 1996; Quinsey, Book, & Lalumiere, 2001; Seto, Lalumiere, & Quinsey, 1997). The CATS has also been employed in laboratory research examining correlates of psychopathy such as impulsivity and failure to learn from experience in noninstitutional samples (Belmore & Quinsey, 1994). Thus, there is empirical support for using the CATS as a measure of antisocial tendencies in both laboratory and clinical settings.

Whereas there is a growing literature supporting the use of the CATS as a measure of antisociality, further research is required. Specifically, questions surround the cross-cultural validity of the CATS since the measure was developed and validated using predominantly Caucasian samples (Folsom & Atkinson, 2007). To date, only one published study has sought to specifically examine the utility of employing the CATS as the index of psychopathy in non-Caucasian samples.

Bolton (2006) sought to explore the predictive utility and cross-cultural generalizability of VRAG scores calculated with the PCL-R as the index of psychopathy (VRAG-P) to VRAG probabilities calculated with the CATS (VRAG-C) when used with African American and Caucasian individuals. His investigation employed a retrospective, archival design. The records

of twenty African American and twenty Caucasian male forensic psychiatric patients maintained at Eastern Louisiana Mental Health System- Forensic Division, a psychiatric hospital in eastern Louisiana, were analyzed. The principal investigator reviewed patient records in order to record PCL-R ratings and calculate CATS scores. This data was subsequently employed to score separate VRAG-C and VRAG-P scores for each participant. Bolton employed forensic review panel decisions as the criteria to which scores on risk assessment measures were compared. He proposed that review panel decisions could be employed as an indirect but robust means of operationalizing the risk posed by study participants. Given that the research that follows employed a methodology similar to the Bolton (2006) investigation, the forensic review panel process employed in his analysis merits further description.

In Louisiana, as in other jurisdictions, legislation has been enacted specifying that recommendations related to the ongoing security and supervision requirements of NGRI patients be made by hospital-based forensic review panels (Manguno-Mire et al., 2007). NGRI patients are considered for gradual release by review panels through formal procedures defined by Louisiana statute (Manguno-Mire et al., 2007). These committees regularly meet to evaluate the treatment progress and risk posed by forensic patients, and render recommendation about their readiness for the next step in the discharge process (Manguno-Mire et al., 2007).

The review panel employed in the Bolton (2006) investigation was comprised of the hospital's chief of staff, the psychologists who conducted the risk assessment, as well as the patient's treating psychiatrist, social worker, and forensic aftercare representative. In Louisiana, review panels are convened following a petition from the patient that involves a change in their security status such as a community pass or transfer to a less restrictive setting (Bolton, 2006). Based upon the review of relevant information and an evaluation of the risk associated with the

petition, the review panel renders a decision about whether the request is justified. According to the release process employed in Louisiana, the review panels' recommendations can include; no change in security status, community passes, transfer to a less restrictive hospital, conditional release to a group home, or release to family members (Bolton, 2006; Manguno-Mire et al., 2007).

Bolton assessed the extent that VRAG-C and VRAG-P scores predicted review panel recommendations using multiple regressions. Separate statistics related to the predictive utility of the measures were calculated for African American and Caucasian participants. Bolton found the VRAG yielded equivalent predictions for African American and Caucasian patients regardless of the index of psychopathy employed. No significant differences were found between VRAG-P and VRAG-C on any of the variables assessed. Of all the variables analyzed, "years of education" was the single best predictor of review panel decisions. Results also indicated a trend towards different CAT and PCL-R scores across racial groups, with African Americans tending to score higher on the PCL-R and Caucasians scoring higher on the CATS; however, the sample was too small to yield significant differences.

Bolton concluded that the results of his investigation lend support to the cross-cultural generalizability of VRAG scores calculated with the CATS, but acknowledged that his study was limited by the size of the sample employed. He recommended that future research should seek to replicate his results with a larger patient sample. The Bolton (2006) investigation was also limited in terms of the extent that the CATS and PCL-R were examined as stand-alone indices of psychopathy. Whereas Bolton compared the predictive utility of the VRAG-C to VRAG-P, he did not compare the predictive utility of the PCL-R and CATS when used as independent risk assessment measures. Moreover, the association between the CATS and the PCL-R factor

structure was not explored. Perhaps in recognition of this limitation, Bolton recommended that future research examine the utility of the CATS as a stand-alone index of psychopathy with diverse samples. Finally, the methodology employed by Bolton to quantify review panel decisions may have inadvertently increased the error variance in his analysis. In particular, he suggested that review panel decisions regarding whether or not a patient could be discharged to a group home or other supervised community placement may be dependent upon their personal resources or the availability of funding in their catchment area. For example, when discussing his results Bolton opined that the higher rates of supervised community placement for Caucasian patients may have resulted from this population's greater access to resources, rather than factors associated with risk.

While the findings of the Bolton (2006) investigation are encouraging, further research examining whether the predictive utility of CATS scores will generalize across racial and ethnic groups is required. The Bolton (2006) investigation was limited in that a relatively small sample size was employed. An investigation employing similar methodology with a larger sample size could potentially yield more generalizable results. Moreover, while the Bolton (2006) investigation compared the predictive validity and cross-cultural generalizability of CATS and PCL-R when employed to calculate VRAG scores, the comparison of both indices of psychopathy as stand-alone measures was limited. Directly comparing the concurrent validity and cross-cultural generalizability of the CATS and PCL-R would directly address the concerns expressed by Folsom and Atkinson (2007) about potential race-related scoring differences on this instrument.

### **Summary of Published Research on the CATS**

Overall, the CATS offers a number of potential advantages over the PCL-R when used to calculate VRAG scores or as a stand-alone risk assessment instrument. The CATS is a brief, nonrestricted test that does not require extensive training to be utilized. As a result, the CATS was a much less costly index of psychopathy compared to the PCL-R. Research has demonstrated that when used in lieu of the PCL-R to score the VRAG, the CATS appeared to generate comparable probability estimates for future violence (Bolton, 2006). In addition, the CATS had also been shown to be associated with recidivism and violent behavior in multiple investigations. Thus, for a number of reasons, the CATS promises to be a valuable risk assessment tool for mental health professionals when used either as part of the VRAG or as a stand-alone measure of psychopathy.

Whereas the CATS offered a number of potential advantages over existing indices of psychopathy, further research is needed. In particular, the cross-cultural generalizability of the CATS has been questioned. The initial dataset used to develop the CATS, as well as the majority of the published research employing the scale, have utilized predominantly Caucasian samples from Canadian institutions. Whereas the Bolton (2006) investigation offers some initial support for the cross-cultural generalizability of the CATS when used in lieu of the PCL-R to calculate VRAG scores, his research was exploratory in nature and employed a relatively small sample. Additional research is needed in order to confirm his results. Moreover, evidence supporting the validity of the CATS when administered as a stand-alone risk assessment measure with ethnically diverse populations is nearly altogether lacking.

The dissertation research that follows was designed to address these concerns and build upon the Bolton (2006) investigation by examining the predicative utility and cross-cultural



generalizability of the CATS with a larger participant sample. A methodology similar to that used by Bolton (2006) was employed. In particular, review panel decision-making once again served as a criterion against which risk assessment instruments were compared. In order to provide context for the methodology, analysis, and discussion that follows, the scientific literature examining the review panel process for NGRI patients will be briefly reviewed.

### **Research Examining the Forensic Review Panel Process**

There is currently a scarcity of published investigations examining the review panel process for forensic psychiatric patients. According to Manguno-Mire et al. (2007), only two research articles on the topic had been published prior to their investigation. Moreover, even when literature examining similar tribunal processes employed outside the United States is considered, only a handful of studies on the topic have been published. As much of the research to date has been exploratory in nature, the variables examined tend to differ substantially between investigations. Nevertheless, certain patient characteristics have repeatedly been shown to exert influence on the forensic review panel process.

Perhaps the most detailed exploration of the forensic review panel process comes by way of two related investigations conducted by Hilton and Simmons (2001) and McKee, Harris, and Rice (2007). These studies examined various patient characteristics associated with whether or not NGRI aquitees were approved for transfer from a maximum security hospital to a less restrictive setting. Archival data from the same Canadian facility was employed in both cases, and the participant samples were composed primarily of Caucasian individuals. Hilton and Simmons (2001) examined review panel decisions rendered from 1992 through 1999. Among the patient characteristics included in their analysis, physical attractiveness and PCL-R scores were correlated with the review panel decisions. Specifically, more attractive patients and those

with lower PCL-R scores were more likely to be approved for transfer to a less restrictive setting. Their results did not find a correlation between VRAG scores and review panel decisions.

McKee et al. (2007) conducted a follow-up investigation by analyzing review panel hearings conducted between 2000 and 2003. The authors were interested in replicating and extending the Hilton and Simmons (2001) investigation. In particular, they wanted to explore whether clinical opinions and review panel decisions had become more aligned with the scores on actuarial risk assessment instruments as a result of their increased use. Again these authors found PCL-R scores to be among the best predictors of more restrictive review panel decisions. Moreover, VRAG scores were moderately correlated with the treating clinicians' recommendations at the hearing. On the basis of their findings, the researchers concluded that clinicians were placing greater emphasis on empirically valid risk factors as compared to the findings of the Hilton and Simmons (2001) investigation. However, their findings also suggested that clinical recommendations continue to be based in part on factors that have little or no relationship with future violence.

The only published examination of forensic review panels in America was conducted by Manguno-Mire et al. (2007). These researchers reviewed patient records from Eastern Louisiana Mental Health System, Forensic Division, the same facility where Bolton (2006) conducted his research. Multiple demographic and treatment related variables were analyzed in terms of their relationship to review panel decisions. The authors sought to examine whether review panel judgments were informed by empirically supported data. According to their results, two factors were found to be significantly related to review panel decisions. Panel members were less likely to recommend release for individuals with higher PCL-R scores and who committed their crimes at an earlier age. The authors concluded that even though review panel members are presented

with multiple sources of clinical information (such as summaries of the patient's recent behavior, medical records, and social worker evaluations of discharge readiness), clinicians appear to place an emphasis on risk assessment instruments when rendering decisions.

Finally, although not the primary focus of the study, factors associated with the forensic review panel process in England were detailed in an investigation conducted by Morrissey et al. (2007). As with the present dissertation research, these investigators sought to examine the predictive validity of a risk assessment measure, by comparing PCL-R scores to review panel decisions. The specific variable examined was whether or not the review panel approved the transfer of intellectually disabled patients out of a maximum security setting during the first two years since admission. Once again, PCL-R scores were found to be negatively correlated with decisions to transfer patients to a less restrictive setting. That is, those participants with higher PCL-R scores were less likely to be approved for transfer.

In summary, available research has shown review panel decisions to be correlated with PCL-R scores, such that those participants with the highest psychopathy ratings are the least likely to be approved for transfer to less restrictive settings. Few other variables have been repeatedly associated with the review panel process; however, this is likely a consequence of the limited amount of literature available and the lack of overlap between published investigations in terms of the variables examined. Due to the limited research on the review panel process in America, investigators have turned to related areas of research in the hopes of better understanding the review panel process (e.g. Manguno-Mire et al., 2007).

One such related area of inquiry is the literature on conditional discharge. In brief, conditional discharge refers to the process whereby forensic patients are permitted to return to the community so long as they agree to abide by certain conditions. If the terms of the

conditional discharge are not followed, release can be revoked and the patient returned to the hospital.

Investigations examining the revocation of conditional release have found that Caucasian patients, women, and those individuals with more education are less likely to have their conditional release revoked (Callaghan & Silver, 2001; Monson, Gunnin, Fogel, & Kyle, 2001). Consequently, Manguno-Mire et al. (2007) included these variables in their analysis and found evidence of similar race-related differences in which patients were conditionally discharged by the forensic review panel. Similarly, the Bolton (2006) investigation also found conditional discharge to be influenced by subtle race-related differences. Moreover, consistent with the conditional release literature, years of education were found to be the best overall predictor of review panel decisions in Bolton's analysis. Thus, it appears race and educational attainment are also potential predictors of review panel decisions in addition to the above described PCL-R scores, physical attractiveness, and age-related variables. Consequently, these demographic variables were included in the present investigation as potential predictors of security-related review panel decisions.

### **The Present Study**

As detailed above, the CATS offers the promise of being a more concise and less costly index of psychopathy compared to the PCL-R, however, evidence supporting its cross-cultural generalizability and validity when used without the VRAG is presently limited. The current study was designed to address these concerns by examining the cross cultural generalizability and predictive utility of the CATS. The CATS was examined as both a stand-alone measure and as part of the VRAG. Methodology similar to that used by Bolton (2006) was employed in order to substantiate and build upon his findings with a larger racially mixed sample of Caucasian and

African American forensic psychiatric patients. Consistent with Bolton (2006), a retrospective archival design was employed to examine the equivalence of VRAG, CAT, and PCL-R scores across racial groups. The concurrent validity of the CATS with the PCL-R was also examined. Finally, scores on all three measures were also compared in terms of their ability to predict forensic review panel decision-making.

Unlike Bolton (2006), the present investigation employed a dataset from a psychiatric hospital in Virginia rather than Louisiana. Due to differences in the treatment process for NGRI patients in Virginia, the research that follows differed from the Bolton investigation in terms of the methodology employed to compare risk assessment measures to the review panel process. Consequently, the section that follows will detail the review panel procedures for NGRI patients in the state of Virginia, and explain how this process was utilized to assess the predictive utility of the CATS, VRAG, and PCL-R.

Similar to the above described review panel process in Louisiana, forensic review panel decisions in the state of Virginia are an integral part of the treatment process for NGRI patients. In both jurisdictions, it is incumbent upon forensic psychiatric hospitals to render decisions regarding the level of supervision needed to properly manage the risk posed by NGRI patients. Moreover, as in Louisiana, Virginia employs a graduated release process. The procedures utilized in Virginia are based on a “demonstration” model of clinical risk management and mandate that patients work through a step-by-step reduction in supervision in order to secure a discharge (*Guidelines for the Management*, 2003). Accordingly, NGRI patients in Virginia must demonstrate their ability to effectively follow the rules at the more restrictive level before progressing to a privilege that involves less supervision. The prearranged series of hospital privileges for NGRI offenders in Virginia are as follows:

1. Transfer from the maximum security forensic unit of Central State Hospital to a civil unit of a state operated mental health facility
2. Escorted grounds privileges at the treating facility (i.e. being able to leave the confines of their assigned ward and building if accompanied by facility staff)
3. Unescorted grounds privileges
4. Community visits, with staff escort (up to forty-eight hours)
5. Unescorted community visits, not overnight
6. Unescorted community visits, overnight, but not over forty-eight hours
7. Trial visits for greater than forty-eight hours
8. Conditional release
9. Release without conditions

The procedures in place in Virginia specify that forensic review panels must periodically evaluate and render decisions about whether NGRI patients are ready for the next privilege. NGRI patients are permitted to request a new privilege every thirty days. If the request is approved by their treatment team, an application packet is compiled and submitted for panel review. Moreover, review panels are obligated to review whether each patient can be managed at a less restrictive level of privilege every three months.

Most privilege requests are evaluated exclusively by the Internal Forensic Privileging Committee at the hospital where the patient resides. This committee is comprised of five members that must include the hospital director, forensic coordinator, medical director or psychiatrist, and licensed clinical psychologist (if the hospital's forensic coordinator is not clinical psychologist). The other committee members are appointed by the hospital director (*Guidelines for the Management*, 2003).

In addition to approval by the Internal Forensic Privileging Committee, in the state of Virginia certain NGRI privileges require the approval of a second independent review panel, the Forensic Review Panel. This second administrative body provides oversight regarding privileging decisions that involve the most substantial decreases in the patient's level of supervision. The Forensic Review Panel is comprised of at least seven members appointed by the Commissioner of the Bureau of Behavioral Health and Health Facilities and must include a minimum of two psychiatrists and two licensed clinical psychologists. The NGRI privileges that must be approved by the Forensic Review Panel include a transfer from the maximum security unit at Central State Hospital, initial unescorted community privileges, and release into the community (with or without conditions).

Both review panels base their evaluations of whether a patient is ready for less restrictive hospital privileges "explicitly" on risk assessment criteria (*Guidelines for the Management*, 2003). The perceived risk for future violence on which these decisions are based is dependent upon a clinical risk assessment that is conducted shortly after patients are admitted to the maximum security unit at Virginia's Central State Hospital. This risk assessment, entitled the Analysis of Aggressive Behavior, involves systematically assessing patients according to known statistical correlates for future aggressive behavior (*Guidelines for the Management*, 2003).

Although the review panel process in Virginia is similar to the review panel process in Louisiana in that risk is central to all panel decisions and mechanisms for gradual release are employed in both jurisdictions, key differences in the process do not allow for the Bolton (2006) methodology to be employed in the present study. In particular, the outcome of a single review panel hearing in Virginia does not lead to an array of possible decisions that correspond to the patient's perceived risk as is the case in Louisiana. Consequently, the present dissertation

research focused on review panel decisions related to a single NGRI privilege. In doing so, the present study employed a methodology similar to the above described Morrissey et al. (2007) investigation of the PCL-R's predictive validity. Specifically, the length of time participants spent in maximum security before being approved by the Forensic Review Panel for transfer to a less restrictive setting, served as the criterion against which scores on the CATS, PCL-R, and VRAG were compared. While the PCL-R is occasionally administered as part of the above described Analysis of Aggressive Behavior, the CATS and the VRAG are not. Thus, review panel members in Virginia are not provided with scores on these three risk assessment measures prior to making decisions. The PCL-R, CATS, and VRAG were therefore scored separately and retrospectively compared to the length of time participants spent in maximum security in order to examine their utility.

The time required for NGRI patients to be approved for transfer out of maximum security was chosen as the criterion against which scores on risk assessment measures were compared for a number of reasons. First, as described above, this privilege involves a significant decrease in supervision for NGRI patients and, therefore, requires the approval of the larger Forensic Review Panel. Second, perhaps more than any other privilege in the Virginia system, the criteria used to assess whether a patient can be transferred out of maximum security focus most specifically on the patient's ongoing risk and security needs. Review panel decisions about whether a patient is ready for transfer to a civil facility hinge upon the risk factors identified in their Analysis of Aggressive Behavior (described above) and whether the patient has been free of any recent violent behavior, is a low escape risk, and is not currently under any special precautions for being a risk to self or others (M. Fahey, personal communication, July 17, 2009). In contrast, decisions related to whether a patient is ready for later privileges in the Virginia system are



dependent upon an increasingly broad array of clinical factors (e.g. insight into their psychiatric symptoms and need for treatment). Also, concerns that were identified as possible extraneous variables affecting discharge decisions in the Bolton (2006) investigation (e.g. differential access to financial resources or the availability of supervised group home in the patient's catchment area), have no bearing on decisions related to the patient's readiness for transfer out of maximum security. Finally, while it can require multiple requests and years before higher risk offenders are transferred to a less restrictive setting, the vast majority of patients are eventually approved. In contrast, the highest risk offenders may not have graduated to later privileges (e.g. conditional discharge) within the time period for which records employed in the present investigation were available.

Previous investigations of the forensic review panel process have utilized similar methodology. For instance, both Hilton and Simmons (2007) and McKee et al. (2007) examined the forensic review panel process by comparing the characteristics of those patients that were approved for transfer out of maximum security to those who were not. Moreover, Morrissey et al. (2007) examined the predictive validity of the PCL-R by examining whether it was able to predict which patients would be approved for transfer out of maximum security during their first two years of treatment. Thus, the time required for NGRI patients to be approved by the Forensic Review Panel for transfer out of maximum security was chosen as the criterion against which scores on risk assessment measures were compared because it is an established and robust indicator of their perceived risk and security needs.

To summarize, the present research was designed to add to the existing literature on the CATS by addressing a number of complimentary goals. First, the equivalence of VRAG actuarial risk predictions scored with the CATS (VRAG-C) and the PCL-R (VRAG-P) was

examined. The concurrent validity of the taxon scale as a stand-alone risk assessment measure was also assessed by comparing CATS and PCL-R scores. Questions surrounding the cross-cultural validity of the CATS were addressed by assessing for any race-related scoring differences in a racially mixed sample of Caucasian and African American participants. Finally, the predictive validity of VRAG, CATS, and PCL-R was assessed by exploring whether scores on these measures were able to predict the length of time before participants were approved for transfer out of a maximum security treatment setting. It was hypothesized that:

- 1) Whether calculated with the PCL-R or the CATS as the index of psychopathy, higher VRAG scores would predict longer periods of hospitalization in a maximum security setting for both Caucasian and African American patients.
- 2) Whether calculated with the PCL-R or the CATS as the index of psychopathy, VRAG ratings would not significantly differ for the overall sample or within racial groups.
- 3) Scores on the CATS and the PCL-R would be highly correlated and provide similar ratings of psychopathy.
- 4) Higher CATS and PCL-R scores would predict longer periods of hospitalization in a maximum security setting for both Caucasian and African American patients.

## CHAPTER 2

### Method

#### Design

The present research was retrospective and employed a correlational research design. Participant's scores on risk assessment measures were calculated and then retrospectively compared to the period of time they were treated at a maximum security forensic hospital before being approved for transfer to a less restrictive setting. Regression analyses were employed to examine the extent that participants' scores on risk assessment measures accounted for variance in the length of time they spent in maximum security confinement.

#### Power Analysis

The present study is an extension of the Bolton (2006) investigation. Using similar methodology, he found a moderate effect size for the association between PCL-R scores and review panel decisions. Previous investigations employing the CATS, VRAG, and PCL-R have repeatedly found moderate to large effect sizes for their ability to predict an individual's risk for violent behavior. Assuming a similar medium effect size, an alpha level of .05, and total of six predictors in the regression equation, Cohen (1992) recommends that a minimum of 97 participants be employed in order to minimize the chances of Type II error. While the sample employed in the present investigation approached this recommended number, a sample of 94 participants was ultimately obtained.

## **Participants**

The present research was a retrospective archival study. Data was obtained from records maintained at Eastern State Hospital, a civil psychiatric hospital operated by the Virginia Department of Behavioral Health and Developmental Services that provides inpatient assessment and treatment to psychiatric patients. The Eastern State Hospital patient population is predominantly African American, male, and of lower socioeconomic status.

The records of adult male forensic psychiatric patients treated at Eastern State Hospital were reviewed in order to gather demographic data, score risk assessment measures, and record the number of months each participant was treated in a maximum security setting before they were approved for transfer to a less restrictive setting. The patient charts of 47 African American and 47 Caucasian male patients found Not Guilty by Reason of Insanity (NGRI), and whose records indicate a history of violent behavior, were selected for data collection. Female patients, individuals who were not identified as Caucasian or African American, and patients whose records did not indicate a history of violence or a felony NGRI offense were excluded from the analysis.

## **Measures**

**VRAG.** The Violence Risk Appraisal Guide (Appendix B) is a 12-item instrument, scored from collateral data, designed to provide 7- and 10-year probabilities of violent recidivism. Each item on the VRAG (e.g. History of alcohol problems, Marital history) is assigned a weight, and the sum of those weights provides the total score. One of the most heavily weighted items on the VRAG is the patients' score on a measure of psychopathy. This item can be scored using either the PCL-R or the CATS as the index of psychopathy. Total VRAG scores range from -26 to 38 and correspond to 7- and 10-year probability estimates for

violent recidivism. The instrument has excellent interrater reliability ( $r > .95$ ) and concurrent validity with other risk assessment measures (Quinsey et al., 1998; Quinsey et al., 2006). VRAG scores have repeatedly been shown to effectively distinguish recidivists from nonrecidivists among various forensic populations and to be associated with institutional violence among psychiatric inpatients (Doyle et al., 2002)

**PCL-R.** The Psychopathy Checklist Revised is a 20-item checklist designed to identify psychopathic individuals. Items are scored based on an interview and/or a review of collateral data. Items (e.g. lack of remorse or guilt, shallow affect) are scored on a 3-point Likert-type scale (0, 1, or 2) and are summed to provide a total score that ranges from 0 to 40. Total scores can be converted into weights and factored into the VRAG. In addition, separate scores can be calculated for the interpersonal and affective versus criminal lifestyle factors as well as the four facet scales (Interpersonal, Affective, Lifestyle, and Antisocial). The PCL-R can be scored by record review alone when employed for research. Published investigations have shown the correlation between the standard PCL-R administration (i.e. file review plus interview) and file review only scoring to be very high ( $r = .93$ ; Wong, 1998). Whether scored with or without a clinical interview, the PCL-R has demonstrated excellent interrater ( $r_s = .82$  to  $.93$ ) and test-retest ( $r = .89$ ) reliability. The PCL-R has repeatedly been shown to predict general and violent recidivism, poor treatment progress, instances of aggression or misconduct within institutional settings, and review panel decisions about the release of forensic psychiatric inpatients (Bolton, 2006; Edens et al., 1999; Edens & Campbell, 2007; Hare, 1985; Manguno-Mire et al., 2007; Morrissey et al., 2007). The principal investigator has participated in a multiday professional workshop on the PCL-R conducted by the instrument's author that included instruction on scoring the instrument through file review.

**CATS.** The Child and Adolescent Taxon Scale is a brief 8-item instrument designed to provide an index of adult psychopathy through assessing a series of childhood and adolescent behavioral dimensions. The CATS yields total scores that can range from 0 to 8 which can be converted into weights and factored into the VRAG. Published investigations employing the CATS found the measure to have good internal consistency (Alpha = .71) and excellent interrater reliability ( $r > .90$ ; Skilling et al., 2001; Glover et al., 2002). The CATS has been shown to predict violent and general recidivism, demonstrates concurrent validity with other established risk assessment measures, and has been associated with psychopathic behaviors such as impulsivity, aggression, and failure to learn from experience in non-forensic populations (Belmore & Quinsey, 1994; Glover et al., 2002; Skilling et al., 2002).

A demographic record form was completed for each participant. This form was designed to gather treatment related data and variables that have been associated to review panel decisions in previous investigations. The variables recorded included the participants' age, age at index offense, race, years of education, length of hospitalization, number of previous hospitalizations, the number of months between admission as an NGRI patient and their approval for transferred to a civil psychiatric hospital, and the date the data was recorded (Appendix A). Although attractiveness has been associated with review panel decisions in previous investigations, this variable was not included in the record form as it could not be coded from available archives.

## **Procedure**

Patient files were reviewed in order to score all assessment measures, gather demographic data, and code criterion variables. A randomly ordered list of male NGRI patients treated at Eastern State Hospital was generated and employed to sort through patient records for individuals who met inclusion criteria. Individuals who were not charged with felony offenses,

did not have a history of violence, and were not identified as Caucasian or African American were excluded from data collection. Once a patient file was excluded, the principal investigator proceeded to the next patient on the list.

The records of all individuals who met inclusion criteria were then reviewed in detail in order to retrospectively score the VRAG, PCL-R, CATS, and complete a demographic data form. Separate VRAG scores were then computed by entering weighted values from the CATS (VRAG-C) and PCL-R (VRAG-P). Once the files of 50 African American participants were coded, only the records of Caucasian individuals were reviewed in order to ensure the sample was composed of an equal number of participants from each racial group. Ultimately, a total of 47 Caucasian participants met inclusion criteria. Consequently, data coded from the final three African American participants (i.e. 48 through 50) were excluded from the analysis.

A total of 117 patient files were reviewed. Eighteen participants were excluded from data collection due to not having a history of violence, identifying as belonging to a racial group other than Caucasian or African American, and being found NGRI for a misdemeanor offense. Two additional participants were excluded because their files were missing or unclear regarding required data (e.g. the number of months spent in maximum security before being approved for transfer). Finally, as described above, three African American participants were excluded from the analysis in order to ensure equal number of participants from each racial group.

In order to protect confidentiality, all participants were assigned a study related identification number so that names and other sources of identifying information were not associated with study data. These identification numbers were affixed to all demographic record forms and test summary sheets. No interviews, tests, or other procedures that required direct

contact with patients were performed and no unique identifying information was coded as part of this research.

### **Missing Data**

Consistent with the methodology employed in the Bolton (2006) investigation, when records did not provide sufficient information to score one of the items from the CATS or VRAG, this researcher assumed the patient did not possess the risk factor and scored it in the negative direction. For example the second item on the CATS reads “Teenage alcohol problem” and is scored either “yes” or “no.” If the record being examined did not have the information necessary to confirm or deny the presence of a teenage alcohol problem, it was assumed that the patient did not have a teenage alcohol problem and the item was scored “no.” When insufficient information was available to code an item on the PCL-R, this item was prorated according to procedures delineated by the PCL-R technical manual (Hare, 2003).



## CHAPTER 3

### Results

The present study was designed to examine the cross-cultural generalizability and predictive utility of the CATS, a concise and potentially more efficient measure for assessing psychopathy compared to the PCL-R. The research design employed allowed for the CATS to be examined as both a stand-alone measure and as part of the VRAG, a widely employed actuarial risk assessment scale. The concurrent validity of the CATS with the most widely employed measure of psychopathy, the PCL-R, was also examined. This chapter presents study results as they pertain to characteristics of the participant sample, the hypothesized relationships between risk assessment instruments, and the relationship between these variables and the length of time participants were hospitalized in maximum security before being approved for transfer to a less restrictive setting. Data were summarized and descriptive statistics examined. Significant relationships between demographic variables and scores on risk assessment measures were identified by calculating Pearson's correlation coefficients. Finally, multiple regression analyses were completed to determine the extent to which predictor variables accounted for variance in the criterion variable.

#### **Descriptive Data and Demographic Characteristics of Participants**

Descriptive statistics for participants' demographic data, scores on risk assessment measures, and the length of time required for them to be approved for transfer out of maximum

security are summarized in Table 1. Demographic variables include the participants' age, age at index offense, years of education, number of prior hospitalizations, and the total number of months they had been hospitalized at the time that data was collected. Examination of Table 1 reveals that participants ranged in age from 21 to 76 and the mean age was 45.30 ( $SD = 13.37$ ). The participants' education levels ranged from the fourth grade to seven years of college. The mean years of formal education was 11.22 ( $SD = 2.57$ ). Participants also differed considerably with regard to the age at which they were found NGRI for the index offense. The ages recorded for this variable ranged from 18 to 69, with a mean of 36.14 ( $SD = 11.99$ ).

Table 1

*Range, Measures of Central Tendency, and Variance for Demographic and Predictor Variables*

Variable	Mean	<i>SD</i>	Median	Range
Age	45.30	13.37	45.00	21-76
Education	11.22	2.57	12.00	4-18
Age at Index Offense	36.14	11.99	33.00	18-69
Months Hospitalized	76.14	79.15	47.00	7-322
Months Before Transfer	30.80	50.59	11.00	4-243
Previous Hospitalizations	7.29	6.57	5.00	0-27
PCL-R	15.82	7.87	15.80	0-33
CATS	3.43	2.20	3.00	0-8
VRAG-P	4.06	11.11	5.00	-21-27
VRAG-C	5.28	11.11	5.00	-18-27

Descriptive data related to the participants' treatment histories revealed considerable variability among participants for these indicators. The number of months that participants had been hospitalized as NGRI patients varied from as few as seven months of treatment to 322 months, with a mean of 76.14 ( $SD = 79.15$ ) months of treatment. The large amount of variance for this variable was anticipated as there are no set guidelines regarding the length of time NGRI patients are committed for psychiatric treatment, so long as they continue to be considered mentally ill and dangerous by the court of jurisdiction. Similarly, where some participants had no previous psychiatric admissions, others had as many as 27. The mean number of prior hospitalizations for participants was 7.29 ( $SD = 6.57$ ). Finally, the length of time participants were held in maximum security before being approved for transfer to a civil psychiatric hospital ranged from four months to 243, with a mean of 30.80 months ( $SD = 50.60$ ). This variance was also expected as the length of time before participants were approved for transfer is again reflective of the case specific nature of risk assessment and related decision-making with forensic psychiatric patients. NGRI patients in Virginia are treated in a maximum security unit until such time that the Forensic Review Panel deems their level of risk to be manageable in a less restrictive setting.

The means and standard deviations for participant scores on risk assessment instruments are also presented in Table 1. The mean scores on risk assessment instruments were compared to those published in previous investigations using independent means  $t$ -tests. The mean PCL-R score for participants was 15.82 ( $SD = 7.87$ ). This score is significantly lower than the mean score of 21.5 ( $SD = 6.9$ ) for forensic psychiatric patients reported in the PCL-R technical manual  $t(1338) = 7.62, p < .05$  (Hare, 2003), but did not significantly differ from the mean score of 13.53 ( $SD = 6.69$ ) reported in the Bolton (2006) investigation  $t(132) = 1.60, p > .05$ . In terms of

the VRAG, the mean participant scores for the VRAG-P and VRAG-C were 4.06 ( $SD = 11.11$ ) and 5.28 ( $SD = 11.11$ ) respectively. The mean participant score for the VRAG-C was significantly lower than the mean VRAG score of 9.7 ( $SD = 9.8$ ) for forensic psychiatric patients in a recently published investigation (Snowden, 2009),  $t(144) = 2.23, p < .05$ . Lastly, the mean CATS score for the sample 3.43 ( $SD = 2.20$ ) was significantly higher than the mean of 2.00 ( $SD = 1.20$ ) reported by Bolton (2006),  $t(132) = 3.87, p < .05$ .

### **Correlations Between Risk Assessment Instruments and Demographic Variables**

Several hypotheses regarding the relationships among risk assessment instruments were presented in the introduction. It was hypothesized that VRAG scores would be similar regardless of whether the CATS or the PCL-R was used as the index of psychopathy. Consistent with this prediction, the scores from the VRAG-C and VRAG-P were determined to be strongly correlated ( $r = .98, p < .01$ ). This result is consistent with the correlation of .98 between VRAG-C and VRAG-P scores reported by Quinsey et al. (1998). The seven-year probability estimates for recidivism generated from VRAG-C and VRAG-P scores were also compared using Pearson's correlation coefficients. As anticipated, the recidivism estimates were also strongly correlated ( $r = .96, p < .01$ ).

Beyond the hypothesis that the PCL-R and the CATS would yield similar VRAG ratings, it was also predicted that the scores on these two instruments would be highly correlated when used independent of the VRAG. Once again, results were consistent with this prediction as PCL-R and CATS scores were determined to be highly correlated ( $r = .60, p < .01$ ) as reflected in Table 2.

Table 2

*Pearson Correlations for Demographic and Predictor Variables*

Variable	VRAG-C	VRAG-P	CATS	PCL-R	Months Before Transfer
VRAG-C	--	.98**	.79**	.72**	.09
VRAG-P	--	--	.73**	.78**	.05
CATS	--	--	--	.60**	.23*
PCL-R	--	--	--	--	-.03
Factor 1: Interpersonal/Affective	.49**	.58**	.37**	.84**	-.03
Factor 2: Antisocial/Lifestyle	.75**	.78**	.67**	.91**	-.01
Facet 1: Interpersonal	.44**	.51**	.26*	.67**	-.03
Facet 2: Affective	.38**	.46**	.36**	.73**	-.03
Facet 3: Lifestyle	.63**	.67**	.52**	.85**	-.22*
Facet 4: Antisocial	.68**	.69**	.67**	.76**	.22*
Age	-.53**	-.50**	-.38**	-.31**	.22*
Age at Index Offense	-.56**	-.51**	-.47**	-.28**	-.25**
Years of Education	-.31**	-.27**	-.46**	-.31**	-.49**
Months Hospitalized	-.04	-.08	.11	-.10	.71**
Previous Hospitalizations	.10	.14	.02	.18	.14

*Note.* \* $p < .05$ ; \*\* $p < .01$

The correlation between CATS scores and the factor and facet scores of the PCL-R was calculated in order to further examine the relationship between the two instruments. Consistent with previous research (Skilling et al., 2002), the CATS was found to be more highly correlated

with PCL-R Factor 2, Social Deviance ( $r = .67, p < .01$ ) than with Factor 1, Interpersonal/Affective ( $r = .37, p < .01$ ). The relationship between the CATS and Social Deviance dimension of the PCL-R was then assessed by computing correlations for the facet subscales of Factor 2. Results indicated a slightly stronger correlation for PCL-R Facet 4, Antisocial ( $r = .67, p < .01$ ) than for Facet 3, Lifestyle ( $r = .52, p < .01$ ). Thus, the CATS appears to be most strongly correlated with PCL-R items that are associated with persistent criminality.

A number of demographic variables were found to be related to scores on risk assessment instruments (Table 2). Age and age at index offense were negatively correlated with scores on all risk assessment instruments, indicating that younger participants tended to have higher risk assessment scores. There was a strong negative correlation between age and VRAG-C ( $r = -.53, p < .01$ ) and VRAG-P ( $r = -.50, p < .01$ ) scores, and moderate negative correlations for CATS ( $r = -.38, p < .01$ ) and PCL-R scores ( $r = -.31, p < .01$ ). For age at index offense, the correlations were  $-.56$  and  $-.51$  for VRAG-C and VRAG-P respectively,  $-.47$  for the CATS, and  $-.28$  for the PCL-R ( $ps < .01$ ). Thus, as with chronological age, those participants who committed index offenses at a younger age tend to have higher scores on risk assessment measures.

In addition to the negative relationship between the age variables and risk assessment scores, years of education was also negatively correlated with scores on the VRAG-C ( $r = -.31, p < .01$ ), VRAG-P ( $r = -.27, p < .01$ ), CATS ( $r = -.46, p < .01$ ), and PCL-R ( $r = -.46, p < .01$ ). Thus, participants with fewer years of education tend to have higher scores on each of the risk assessment measure examined. The correlations among risk assessment scores and demographic variables are presented in Table 2.

## Relationships Between Predictor and Criterion Variables

In addition to examining the interrelations among risk assessment instruments, a central focus of the present investigation was to examine the extent that these measures would predict the perceived risk of NGRI acquittees. It was hypothesized that higher scores on the VRAG, CATS, and PCL-R would be associated with longer periods of confinement in maximum security before the Forensic Review Panel approved a transfer to a less restrictive setting. For the total sample, results indicated a statistically significant positive relationship between CATS scores and the number of months participants were hospitalized in a maximum security setting ( $r = .23$ ,  $p < .05$ ), but no significant relationships were found for VRAG-C, VRAG-P, or PCL-R total or factor scores. Thus, of the three scales, the CATS alone was correlated the length of time participants were confined in maximum security before being approved for transfer.

Although no significant relationship was found for PCL-R total or factor scales scores and the time before participants were approved for transfer, an interesting pattern emerged for Facets 3 and 4. Specifically, Facet 4, Antisocial, was positively correlated with the length of time participants were treated in a maximum security setting ( $r = .22$ ,  $p < .05$ ), while results for Facet 3, Lifestyle, showed a negative correlation ( $r = -.22$ ,  $p < .05$ ). Thus, PCL-R items that directly reflect criminal behavior (e.g. criminal versatility, revocation of conditional release) have the strongest association with the length of time before participants were approved for transfer. Perhaps the simplest explanation for the relationship between Facet 4 and the length of time before participants were approved for transfer is that these PCL-R items directly correspond with criminal record data. That is, those individuals with highest Facet 4 scales would tend to also have the lengthiest criminal records, and therefore would naturally be perceived by review panel members as being higher risk.

The negative relationship between Facet 3, Lifestyle, and the length of time required for participants to be transferred to a less restrictive setting was unexpected. Perhaps some of the items that comprise this facet (e.g. parasitic lifestyle, lack of realistic long term goals, irresponsibility) are less likely to contribute to behavioral problems within high-security environments. That is, the expression of these behaviors may be more evident and destructive in the community, as compared to the limited timeframe and high degree of supervision that would have characterized the participants' treatment in the maximum security hospital. Still, reasons for the negative association are unclear and may require more detailed examination in the future.

To assess the extent that race contributed to variation in the criterion variable, the sample was divided according to race and separate Pearson's correlation coefficients were computed. Results indicate a number of correlations differed across racial groups (Table 3). For Caucasian participants, the correlation between CATS scores and the duration of confinement before being approved for transfer out of a maximum security was slightly higher than results for the overall sample ( $r = .29, p < .05$ ). Unlike the results for the total sample, no significant correlation was found between the antisocial facet of the PCL-R and the length of time participants were treated in a maximum security setting. When African American participants were examined, none of the risk assessment instruments were significantly correlated with the length of time before they were approved for transfer to a less restrictive setting.

### **Comparison of VRAG-C and VRAG-P Scores**

Turning to the comparison of the VRAG-C and VRAG-P, a paired samples *t*-test was employed to determine if scores differed depending on whether the CATS or PCL-R was employed as the index of psychopathy. The mean total score for the VRAG-P, 4.06 (SD = 11.09) was significantly lower than the mean of 5.28 (11.11) for the VRAG-C,  $t(93) = 5.58$ ,



$p < .01$ . However, when the seven-year recidivism estimates derived from the measures were

Table 3

*Pearson Correlations for Demographic and Predictor Variables for Each Racial Group*

Variable	VRAG-C	VRAG-P	CATS	PCL-R	Months Before Transfer
African Americans					
VRAG-C	--	.98**	.79**	.75**	-.03
VRAG-P	--	--	.74**	.69**	-.06
CATS	--	--	--	.60**	.15
PCL-R	--	--	--	--	-.02
Facet 4: Antisocial	.66**	.65**	.64**	.73**	.19
Age	-.63**	-.60**	-.44**	-.36*	.22
Age at Index Offense	-.67**	-.64**	-.49**	-.38**	-.19
Years of Education	-.10	-.15	-.36*	-.23	-.56**
Caucasians					
VRAG-C	--	.98**	.80**	.74**	.17
VRAG-P	--	--	.74**	.81**	.13
CATS	--	--	--	.63**	.29*
PCL-R	--	--	--	--	.03
Facet 4: Antisocial	.70**	.73**	.71**	.78**	.25
Age	-.42**	-.38**	-.33*	-.25	.22
Age at Index Offense	-.47**	-.40**	-.46**	-.18	.30*
Years of Education	-.42**	-.37**	-.56**	-.37*	.44*

Note. \* $p < .05$ ; \*\* $p < .01$

compared using a paired samples *t*-test, the difference between the mean VRAG-C, .37 (*SD* = .19) and VRAG-P, .36 (*SD* = .20) probability estimates was not significant.

To determine if African American and Caucasian participants scored differently on either the VRAG-C or VRAG-P, an independent samples *t*-test was employed. On the VRAG-C, the mean score for African American participants was 6.21 (*SD* = 10.57), while the mean for Caucasians was 4.34 (*SD* = 11.67). With an  $\alpha = .05$ , the difference between the means was not significant,  $t(92) = .82$ . On the VRAG-P, African American participants had a mean score of 5.40 (*SE* = 10.30) while the mean for Caucasians was 2.71 (*SD* = 11.80). This difference was also not significant,  $t(92) = 1.18, p > .05$ .

Next, an independent samples *t*-test was computed to determine if African American and Caucasian participants scored differently on the CATS or PCL-R when they are used as stand-alone measures. The mean score for Caucasians on the CATS 3.49 (*SD* = 2.25) was not significantly different from the mean of 3.36 (*SD* = 2.19) calculated for African American participants  $t(92) = .28, p > .05$ . The difference between mean scores on PCL-R for Caucasians 14.75 (*SD* = 8.16) and African Americans 16.88 (*SD* = 7.50) was also not significant,  $t(92) = 1.32, p > .05$ .

A final independent samples *t*-test was conducted to assess if African American and Caucasian participants differed in terms of how long they were perceived as requiring confinement in a maximum security before being approved for transfer to a less restrictive setting. The mean number of months spent in maximum security for African American patients 29.5 (*SD* = 43.90) did not significantly differ from the mean number of months for Caucasian patients 32.11 (*SD* = 56.95)  $t(92) = .25, p > .05$ .

### Predicting the Time Spent in Maximum Security

The relationship between risk assessment measures and the time required for patients to be approved for transfer out of maximum security was evaluated through a series of regression analyses. The first regression examined the time before participants were approved for transfer as a function of the CATS, VRAG-C, and PCL-R scores. The VRAG-P was not included in the analysis due to its strong correlation with the VRAG-C. For this preliminary analysis, predictor variables were entered using simultaneous forced entry and yielded a significant model,  $F(3, 90) = 3.22, p < .05$ , that explained 10% of the variance in the length of time required for participants to be approved for transfer. Of the three measures entered, only CATS scores made a significant contribution to the model ( $\beta = .44, t = 2.68, p < .05$ ) as reflected in Table 4.

Table 4

#### *Summary of Forced Entry Regression Analysis Predicting the Duration of Maximum Security*

##### *Treatment With the Risk Assessment Measures*

Variable	<i>B</i>	<i>SE B</i>	$\beta$
CATS	9.60	3.36	.42**
PCL-R	-1.38	1.04	-.21
VRAG-P	-.39	.86	-.09

*Note.*  $R^2 = .10$ ; \*\* $p < .01$

Next, a hierarchical regression analysis was conducted in order to examine the extent that the variance in the time participants spent in maximum security could be explained by both the risk assessment measures and demographic variables. The demographic variables selected for inclusion were either significantly correlated with the length of time participants spent in

maximum security or had repeatedly been shown to predict review panel decisions in previous investigations (e.g. Bolton, 2006; Manguno-Mire et al., 2007; Monson et al., 2001). They included race, years of education, and age at index offense. Once again all three risk assessment measures were included in the regression. The demographic predictors were entered in the first step of the regression and scores from the CATS, PCL-R, and VRAG-C were entered in the second block.

When regressed onto the length of time before participants were approved for transfer out of maximum security, the demographic variables yielded a significant model,  $F(3, 88) = 8.40, p < .01$  (Table 5). Approximately 20% of the variance in the time required before being approved for transfer out of maximum security was explained by the demographic variables comprising the model. It should be noted that although the beta values for years of education ( $\beta = -.40, t = -4.17, p < .01$ ) and age at index offense ( $\beta = -.19, t = -1.99, p = .05$ ) were significant, the beta for race was not ( $\beta = -.10, t = 1.03, ns$ ).

When the CATS, PCL-R, and VRAG-C scores were entered as a separate block, the risk assessment measures explained an additional 4% of the variance in the dependent variable. However, this increase was not significant,  $F_{\Delta}(3, 85) = .26, ns$ . As can be seen in Table 5, years of education was the only predictor that yielded a significant beta when both demographic and risk assessment variables were included in the regression. This is likely due to the relatively strong correlations between the predictors analyzed.

Table 5

*Summary of Hierarchical Regression Predicting the Duration of Maximum Security Treatment With Demographic Variables and Risk Assessment Measures*

Variable	<i>B</i>	<i>SE B</i>	$\beta$
Step 1			
Race	-9.17	8.91	-.10
Years of Education	-7.66	1.84	-.40**
Age at Index Offense	-.75	.38	-.19*
Step 2			
Race	-.585	9.11	-.05
Years of Education	-8.20	2.16	-.43**
CATS	2.55	3.40	.12
PCL-R	1.43	.96	.24
VRAG-P	-.04	.83	.01

*Note.*  $R^2 = .22$  for Step 1;  $\Delta R^2 = .04$  for Step 2; \* $p < .05$ ; \*\* $p < .01$ .

A more parsimonious hierarchical multiple regression was computed in order to better understand the predictive utility of select demographic and risk assessment variables. The predictors employed in first step of the regression included the demographic variables of years of education and age at index offense. CATS and PCL-R scores were then separately entered in blocks two and three. Race and VRAG-P scores were not included in the analysis as they were shown to have the lowest betas in the first regression. Once again the number of months before participants were approved for transfer out of maximum security was used as the dependent

variable. The demographic variables once again yielded a significant model,  $F(2, 89) = 12.06, p < .01$  and explained 20% of the variance in the time before participants were approved for transfer. The addition of CATS scores did not explain any additional variance  $\Delta R^2 < .01, F_{\Delta}(3, 88) = .01, ns$ . However, the addition of PCL-R scores in the final block of the regression was significant  $F_{\Delta}(4, 87) = 4.87, p < .05$ , and explained an additional 4% of variance in the dependent variable. Table 6 contains a summary of the model.

It should be noted that CATS scores had the weakest beta and partial correlation with the dependent variable ( $\beta = -.01, t = -.08, ns, pr = .01$ ) in the regression. Consequently, a final hierarchical regression was computed comparing the predictive utility of CATS scores beyond years of education. Years of education was chosen because it was consistently shown to be the strongest predictor of the time before participants were approved for transfer. This variable was entered into the regression first, and the CATS was entered in the second block. Interestingly, results indicated that CATS scores did not account for any variance beyond the years of education  $F_{\Delta}(2, 90) = .00, ns$ . Thus, while the CATS had the greatest predictive utility among the risk assessment instruments, the measure did not explain any variance in the dependent variable beyond the participants' years of education.

Table 6

*Summary of Hierarchical Regression Predicting the Duration of Maximum Security Treatment*

*With Select Demographic Variables, CATS, and PCL-R*

Variable	<i>B</i>	<i>SE B</i>	$\beta$
Step 1			
Years of Education	-7.57	1.84	-.39**
Age at Index Offense	-.73	.38	-.19
Step 2			
CATS	-.22	2.70	-.01
PCL-R	-1.55	.70	-.26

*Note.*  $R^2 = .21$  for Step 1;  $\Delta R^2 < .01$  for Step 2;  $\Delta R^2 = .04$  for Step 3; \* $p < .05$ .

## CHAPTER 4

### **Discussion**

Through employing methodology similar to Bolton (2006), the present investigation added to the scientific literature on three leading risk assessment instruments, the VRAG, PCL-R, and CATS. While the VRAG and PCL-R are among the most widely researched measures employed in forensic mental health settings, relatively fewer published studies have examined the CATS. The CATS merits greater consideration because it offers the potential of being a more concise and less costly index of psychopathy compared to the PCL-R.

Previous research has supported employing the CATS as an alternative to the PCL-R when calculating VRAG scores. However, the published investigations have been conducted with predominantly Caucasian samples and there is limited evidence at present supporting the measure's cross-cultural validity. Moreover, there is a paucity of research examining the validity of the CATS when calculated apart from the VRAG as a stand-alone index of psychopathy. The current study was designed to address these concerns by assessing the cross-cultural generalizability and predictive utility of the VRAG, CATS, and PCL-R in a mixed race sample.

The study employed a sample of 94 male NGRI patients committed for treatment at Virginia's Eastern State Hospital. Fifty percent of participants were Caucasian and 50% were African American. The participants varied considerably in terms of the demographic variables and their scores on the three risk assessment measures. Overall, the participants scored lower on



the VRAG and PCL-R compared to previous investigations examining forensic psychiatric inpatients. The analysis of demographic information, scores on risk assessment instruments, and the decision-making related to the participants' perceived level of risk yielded a number of notable findings.

Results from correlational analyses found a number of significant relationships between demographic variables and scores on risk assessment instruments. Both age and age at index offense showed moderate to strong negative correlations with scores on risk assessment instruments, indicating that younger participants tend to score higher on all three measures. This outcome was not surprising as it has been well established that an individual's risk for violent behavior naturally decreases with age (Quinsey et al., 2006). In addition to the negative correlation for age, there were also moderate to strong negative correlations between levels of education and scores on risk assessment instruments. This relationship was strongest for CATS and PCL-R scores. Although the correlation was somewhat stronger than expected, the finding that participants with fewer years of formal education tend to score higher on risk assessment measures is also not surprising as previous research has also found a negative correlation between educational attainment and psychopathy (Hare, 2003). Moreover, fewer years of formal education would be a logical consequence of the various child and adolescent misconduct variables assessed by the CATS. The implications of years of education on review panel decision-making will be discussed in more detail below.

A number of hypotheses were presented regarding the relationships between the various risk assessment measures. It was hypothesized that VRAG ratings would not significantly differ if the CATS was employed in lieu of PCL-R to calculate scores. As expected, VRAG-P and VRAG-C scores were highly correlated. Moreover, while results for the overall sample indicated

that VRAG-C and VRAG-P raw scores differed somewhat, the probability estimates for violent recidivism generated from each measure did not. In terms of race related scoring differences, no significant differences were found for either VRAG raw scores or actuarial risk ratings when the sample was divided by race. Thus, this pattern of results supports the hypothesis that VRAG ratings would not differ for either the overall sample or within racial groups depending on whether the CATS or PCL-R was used as the index of psychopathy.

In addition to hypothesizing the equivalence of VRAG-C and VRAG-P ratings, it was also expected that scores on the CATS and PCL-R would be highly correlated when used as stand-alone measures. As predicted, results showed the scores on both indices of psychopathy to be highly correlated, which suggests both instruments are measuring a similar construct. Furthermore, the strength of the associations between these scales was consistent with previous investigations (Skilling et al., 2002). These results demonstrate the concurrent validity of the CATS with the PCL-R, and offer further support for using the taxon scale as more efficient and less expensive index of psychopathy.

The relationship between CATS scores and the two factor scales of the PCL-R was also consistent with the results from Skilling et al.'s (2002) investigation. Specifically, CATS scores were found to be more highly correlated with the Factor 2 items of the PCL-R than with the Factor 1 items. Thus, the relationship was stronger for the social deviance dimensions than for the interpersonal and affective features of psychopathy as assessed by the PCL-R. This is not surprising given that the CATS items focus on observable antisocial behaviors rather than personality traits or relational variables. Harris et al. (1994) did not emphasize these items when constructing the CATS because they did not find evidence of a taxon underlying the interpersonal and affective factor of the PCL-R.

Lastly, when the sub facets of the PCL-R were compared to CATS scores, a similar pattern of results emerged. The CATS was most strongly correlated with the two facet scales of the PCL-R that when combined form the antisocial/lifestyle factor (i.e. Factor 2), with the strongest correlation being with the antisocial facet (Facet 4). Thus, of all the behavioral dimensions assessed by the PCL-R, it is those items that directly reflect criminal behavior (e.g. criminal versatility, revocation of conditional release) that have the strongest relationship with the CATS. The fact that the CATS shows the strongest correlation with these dimensions bodes well for its utility as a stand-alone risk assessment instrument as the Factor 2 items of the PCL-R have the strongest relationship with recidivism (Quinsey et al., 1998). The relationship between PCL-R and CATS will be discussed in greater detail below.

The final set of analyses directly compared the CATS and PCL-R for race-related scoring difference. Analyses revealed no significant difference in scores on either measure attributable to the race of the participant. Therefore, similar to the results for the VRAG-C and VRAG-P, there was no evidence of any race-related bias in either CATS or PCL-R scores that would potentially undermine the cross-cultural generalizability of the measures.

### **Regression Results**

The remaining hypotheses concerned the predictive utility and criterion related validity of the risk assessment measures. It was anticipated that higher scores on the VRAG, CATS, and PCL-R would predict longer periods of hospitalization in a maximum security setting for both Caucasian and African American patients. Of the three measures included in the regression, only the CATS was significantly related to the length of time participants spent in maximum security before being approved for transfer. These mixed results were contrary to expectations and suggest that review panel decisions regarding the level of security an individual requires may

emphasize different factors than leading risk assessment measures. A review of the criteria considered by the Forensic Review Panel prior to approving transfer offers some perspective as to why this might be the case.

According to Virginia Department of Behavioral Health and Developmental Services policy, in order for a patient to be approved for transfer s/he must not be under any special safety precautions (e.g. one-to-one staff supervision to manage aggression) or judged to be an escape risk, and be free from any recent violent behavior (M. Fahey, personal communication, July 17, 2009). However, these are the minimum criteria as the Forensic Review Panel also evaluates relevant information from an updated Analysis of Aggressive Behavior (AAB). As detailed above, the AAB is a clinical risk assessment conducted with each NGRI acquttee. This risk assessment seeks to integrate established correlates and base rates for aggressive behavior in order to assess, monitor, and manage risk for forensic psychiatric patients (*Guidelines for the Management*, 2003). More than offering static predications of dangerousness, the AAB risk factors are continually updated in accordance with the patient's recent behavior and response to treatment. Compared to the VRAG, PCL-R, and the CATS, the AAB takes a broader array of patient characteristics into account and places an emphasis on dynamic risk factors that change according to the patients' response to treatment. Taken together, the review panel's emphasis on recent behavior within the hospital and a clinical assessment process emphasizing dynamic risk assessment, may account for the limited ability of risk assessment measures to predict the length of time before participants were approved for transfer out of maximum security.

These differences are perhaps most clearly delineated by considering the results for the VRAG. More than any other measure analyzed, the VRAG focuses on static risk factors and seeks to predict the probability of violent recidivism over the longer term. More proximal and

dynamic risk factors such as recent aggression, special precautions, or other indicators of maladjustment within the hospital that might prevent the transfer out of a maximum security setting, are not assessed by VRAG items. This difference may account for the results indicating the VRAG was the weakest predictor of the length of time participants spent in maximum security. Moreover, while the finding that VRAG scores did not predict the time required for participants to be approved for transfer was unexpected, it was not unprecedented as other studies have not found a link between the VRAG and decisions regarding level of security (Hilton & Simmons, 2001). Perhaps results would have been more in line with expectations if VRAG scores were compared to an alternative criterion variable, such as release-related review panel decisions with repeat violent offenders. In these situations, review panels may place a greater emphasis on the probability of long term recidivism and more static risk factors as compared to decisions that address the degree of supervision required in the hospital setting.

Given that PCL-R scores have repeatedly been associated with decisions not to transfer patients out of maximum security treatment settings in previous investigations (Manguno-Mire et al., 2007; McKee, 2007; Morrissey et al., 2007), the finding that it did not predict the length of time before participants were approved for transfer warrants further discussion. Once again, these results may have been due to this particular review panel's emphasis on more short term and dynamic risk factors. However, this explanation would appear to be less applicable to the PCL-R. That is, research has shown that PCL-R scores have been associated with rule violations and management problems within institutional settings similar to those considered by the review panel when assessing whether a patient continues to require treatment in a maximum security setting (Edens et al., 1999; Reiss et al., 1999).

An alternative explanation for why PCL-R scores did not predict the length of time participants spent in maximum security is that the deceptive skills of psychopathic individuals enable them to persuade evaluators that they have made treatment gains and are ready for transfer. Evidence for this possibility comes from recent research suggesting that psychopathic individuals are much more likely to convince parole board members they are ready for release (Porter, Brinke, & Wilson, 2009). It is also possible that calculating PCL-R scores without interviewing patients (i.e. scoring the PCL-R by record review alone) may have led to an underestimation of PCL-R scores for some participants and negatively impacted the predictive utility of the instrument. Previous research supports this possibility, as investigations of file only PCL-R ratings have suggested this method may yield lower scores for those individuals who are rated highest in psychopathy when the structured interview is also employed (Grann, Langstrom, Tangstrom, & Stalenheim, 1998; Wong, 1988).

While PCL-R scores did not predict the length of time participants spent in maximum security, CATS scores did. This is an interesting finding given that both instruments purport to measure psychopathy. One possible explanation for the superior predictive power of the CATS is that the sample examined in the present investigation contained individuals of lower risk compared to those employed in previous research. As stated above, the average scores on risk assessment measures were somewhat lower than those of previous investigations employing a forensic psychiatric patient sample. Of the 94 individuals included in the investigation, only five participants exceeded the traditional cutoff for diagnosing psychopathy (i.e. a PCL-R total score of 30 or greater), while 28 individuals had CATS scores that fell within the highest range (Quinsey et al., 2006). Thus, while the sample contained few individuals that could be identified as psychopaths using the PCL-R, 30% of participants scored in the most heavily weighted range

on the taxon scale. Stated another way, these results suggest that a greater proportion of participants had clinically significant elevations on the CATS compared to the PCL-R. These results are consistent with previous research indicating that the base rate for psychopathy as defined by the CATS (Harris, 1994; Quinsey et al., 2006) is higher than estimated base rate for psychopathy as measured by the PCL-R (Hare, 1993).

One possible implication of the finding that more individuals fall within an elevated range on the taxon scale compared to the PCL-R is that the CATS may be a more sensitive instrument for risk assessment with moderate risk populations. That is, the CATS may have had more explanatory power with the sample employed in the present investigation because participants had lower scores on risk assessment instruments compared to other published studies of forensic psychiatric patients. Future research might be able to shed light on this possibility by evaluating whether the PCL-R is able to predict review panel decisions and recidivism with moderate risk populations or individuals who score below published cutoff scores for the instrument.

### **Predictive Power of the Demographic Variables**

In addition to being highly correlated with scores on risk assessment measures, years of education best predicted the length of time participants were confined in maximum security before being approved for transfer. This parallels the results of the Bolton (2006) investigation which also found this variable to be the best predictor of review panel decisions. Bolton suggested that review panel members may place an emphasis on education in their decision-making because it tends to be seen as associated with better premorbid and postmorbid functioning. Another potential explanation is that educational attainment is emphasized by review panels because of its modest but significant negative correlation with criminal activity

(Andrews & Bonta, 1998). In addition, cognitive impairment is one of the risk factors included in the AAB (*Guidelines for the management*, 2003). While not the same, years of education would certainly correlate with cognitive impairment. Finally, there may be various factors associated with low educational attainment that may mediate the relationship such as difficulty adjusting to institutional settings or poorer response to treatment.

Whereas race has predicted review panel decisions in previous research, this variable was not associated with the length of time before patients were approved for transfer out of maximum security in the present investigation. There are a number of possible explanations for this result. First, it appears that race was predominantly found to be associated with discharge related decisions in investigations that employed smaller sample sizes or older datasets (e.g. Bolton, 2006; Callaghan & Silver, 1998; Monson et al., 2001). The forensic review panel process, and risk assessment more generally, have undergone significant change over the past two decades. It is likely that increased scrutiny has coincided with this evolution and that these clinical-forensic processes have become more sensitive to any potential racial bias in their decision-making. However, it is also possible that race was not predictive because the review panel decisions examined in the present investigation did not involve the discharge of NGRI aquitees into the community. Thus, finances, insurance, or other such resources that may facilitate reintegration into the community and be unequally distributed between racial groups, would not have affected the decision-making examined in the current study.

The relationship between CATS scores and the time spent in maximum security also differed across racial groups as none of the risk assessment measures were significantly correlated with this variable among African Americans. The lack of a significant relationship for CATS scores and perceived risk among African Americans may be due in part to the decreased



sample size when the participants were split by race. Although not significant, the length of time before African Americans were approved for transfer tended to be associated with higher scores on the taxon scale. However, it also appears that there was variation in the strength of the correlation between a key age-related demographic variable and the length of maximum security hospitalization for African American and Caucasian participants that may help explain this result.

While the study found a negative correlation between age at index offense and time spent in maximum security for Caucasian participants, there was no significant relationship between these variables for African Americans. This result suggests that other age-related variables may be differentially applied between racial groups. Thus, it appears that the CATS variables, which reflect early childhood misconduct, may have been emphasized more heavily by review panel members when assessing risk for Caucasian participants. It may also be the case that participant variables that were not included in the present investigation, such as severity of the index offense, are also differentially emphasized depending on race. Perhaps the differential weighting of certain variables is a product of clinical, rather than actuarial risk assessment, and subjective decision-making by review panels. Future research should explore for this possibility by including a wider assortment of developmental, demographic, and treatment related variables with a larger racially mixed participant sample.

### **Implications**

The results of the present research offer further evidence that the CATS can be employed in lieu of the PCL-R when calculating the VRAG. Consistent with prior research, scores from the VRAG-C were highly correlated with scores from the VRAG-P, and the actuarial probability estimates for future violence derived from each measure did not significantly differ.

The present investigation also addressed whether the CATS can be used as an alternative index of psychopathy to the PCL-R. Results revealed the CATS had superior criterion validity than either the PCL-R or the VRAG as it was the only risk assessment instrument that was able to predict the length of time before participants were approved for transfer out of maximum security. Moreover, measures of association established further evidence for the concurrent validity of the CATS, in that its scores were highly correlated with those from the PCL-R. Finally, no evidence was found for race-related scoring differences on the CATS when used as either a stand-alone measure or to calculate VRAG scores. These results, when considered together, lend support for both the validity and cross-cultural generalizability of the CATS.

In addition to establishing further evidence for the validity of the CATS, results offer some perspective on which dimensions of psychopathic behavior are most accurately captured by the instrument. Given that CATS scores were most strongly correlated with Factor 2 and Facet 4 of the PCL-R, CATS scores appear to more accurately reflect adult criminal conduct than the personality traits widely associated with psychopathy. Thus, while the CATS and the PCL-R appear to be measuring a similar construct, the taxon scale may be more narrowly focused than Hare's measure.

These findings bear upon to the larger debate surrounding how to best conceptualize and assess the persistent antisocial behavior exhibited by psychopathic individuals. Those who adhere most closely to the construct of psychopathy as conceptualized by Cleckley and Hare, would likely suggest that the CATS underemphasizes the qualitatively distinct interpersonal and affective characteristics that lie at the core of the construct. For example, Hart and Hare (1989) argued that the failure to assess these trait dimensions of psychopathy, yields a diagnostic model with insufficient conceptual coverage and gives rise to an overly broad and heterogeneous

clinical population. However, those who stress clearly discernable behavioral descriptions such as those employed by the CATS have argued that the difference between personality based and behavioral models of psychopathy is more apparent than real (Skilling et al., 2002). This is to say, the personality characteristics assessed by the PCL-R are subsumed by behavioral descriptions of psychopathy such as those employed by the CATS (Skilling et al., 2002).

At present, however, it is unclear whether the more narrowly focused conceptualization of psychopathy delineated by the CATS is any less effective than the PCL-R for purposes of risk assessment. Perhaps any lingering questions surrounding whether the CATS is measuring the same construct as the PCL-R could be best examined by exploring whether Hare's instrument has any incremental contribution to predicting behavior compared to the taxon scale.

These issues aside, the results of the present investigation provide further evidence that the CATS can be employed as an alternative to the PCL-R when assessing psychopathy. Moreover, as described in detail above, the CATS is much more concise and easy to score than the PCL-R, and could be administered at a fraction of the cost. Thus, even if only employed as a screening instrument, the CATS could represent a valuable tool for managing risk in clinical-forensic and correctional settings where funding is often scarce.

Perhaps the present study's most significant results pertain to the analysis of variables associated with the time participants spent in maximum security before being approved for transfer. While findings suggested that CATS scores were able to account for a small but significant amount of the variance in time to transfer, the PCL-R and the VRAG scores did not. Instead, the number of months spent in maximum security before being approved for transfer was better accounted for by two demographic variables, age at index offense, and the variable found to be the best overall predictor, years of education. In fact, the analysis revealed the

CATS was unable to account for any variance in the time participants were confined in maximum security beyond years of education. These results are similar to previous investigations that have found variables with little or no association with risk, such as education, race, and attractiveness, to be among the best predictors of review panel decisions (Bolton, 2006; Hilton & Simmons, 2001; McKee et al., 2007).

These findings in many ways parallel the research and debate surrounding unstructured versus actuarial risk assessment. As detailed in the introduction, unstructured risk assessments have been criticized on the basis that they are subjective, overemphasize the importance of single variables, and are overly impressionistic (Hilton & Simmons, 2001; Quinsey et al., 1998). Given results indicating that decisions about when to transfer forensic patients out of maximum security appear to have little or no relationship with scores on risk assessment instruments, and may place a differential emphasis on demographic variables such as age at index offense and years of education, the forensic review panel decision-making process may be subject to these same criticisms. Moreover, given that these decisions are largely based on a clinical risk assessment process that purports to systematically integrate empirically supported risk factors, one wonders to what extent semi-structured and scientifically informed clinical assessments improve upon unaided clinical opinion.

It may be that semi-structured clinical risk assessments like the AAB, are limited in that they do not provide mechanisms for weighing the individual risk factors or utilize algorithmic decision rules to estimate overall risk as is the case with actuarial instruments. Though the various risk factors may be empirically supported, they may be emphasized in an idiosyncratic manner by the clinician completing the assessment. Moreover, given that they do not yield scores that can be employed to compare the relative risk of one patient to another, they may be

more open to subjective interpretation by forensic review panel members. In short, the assessment and decision-making process employed by review panels may not accurately reflect the patient's actual level of risk. However, there may be valid reasons why the length of time participants spent in maximum security did not coincide with PCL-R and VRAG scores. It is possible that security-related decisions within institutional settings should involve the consideration of a wider range of risk factors and place greater emphasis on dynamic variables in order to monitor treatment gains. Moreover, given that forensic review panels involve the intersection of psychological and legalistic decision-making, they may require a more flexible and idiographic approach to assessment. Nevertheless, the results of the present investigation suggest that the review panel process employed with forensic psychiatric patients needs to be better understood and that these administrative bodies would likely benefit from employing a more structured (i.e. actuarial/algorithmic) decision-making process.

### **Limitations to the Current Study**

There are a number of limitations that should be considered when interpreting the results of the present investigation. First, although larger than the sample employed by Bolton (2006), the sample employed in the present dissertation research was still relatively small. Given that a number of findings were trending in the expected direction, but did not achieve statistical significance, it is likely that a larger sample size would have yielded more significant findings. Due to this same limitation, only a relatively small number of predictor variables could be meaningfully employed in the regression analyses. As delineated above, the unexpected regression results may have been better understood had more demographic and treatment related predictor variables been employed.

In addition to being limited by the number of participants included in the analysis, the present investigation was also somewhat limited by other sample characteristics. One obvious drawback is that no female participants were included. Whereas female participants were deliberately excluded in order to simplify the analysis, it is unclear whether any of the results of the present investigation generalize to this population. Secondly, relatively few individuals scored in the psychopathic range on the PCL-R and the mean score on this risk assessment measure was somewhat lower than what has been reported for forensic psychiatric patients in previous investigations (Hare, 2003). Consequently, comparatively few psychopathic individuals were included in the analysis. This could explain why a limited amount of variance was accounted for by the CATS and PCL-R.

In addition to constraints related to the sample, the present investigation was also limited by certain aspects of the research design employed. As previously mentioned, PCL-R scores were calculated retrospectively by file review alone. Although this method of scoring the PCL-R is common for research purposes, there is reason to believe that this method may have underestimated scores for the most psychopathic participants.

There were also methodological limitations associated with the various predictor and criterion variables included in the regression analyses. Specifically, while members of the forensic review panel in the Bolton (2006) investigation were not provided with VRAG and CATS scores, they were expected to review PCL-R scores as part of their decision-making process. None of these measures were routinely available to members of the review panel in the present investigation. Had the CATS, VRAG, and PCL-R scores for each participant been provided to forensic review panel members prior to making transfer decisions, the dispositions may have been more reflective of the participants' scores on these instruments.

In terms of the criterion variable, it appears that review panel decisions about whether to transfer participants out of maximum security may be based in large part on their recent behavior and variables associated with short-term risk. This likely also complicated comparing the predictive utility of the CATS, VRAG, and PCL-R, as these measures primarily predict the potential for risk over the long-term.

### **Future Directions**

The results of the present investigation suggest that there are potentially more parsimonious and less costly alternatives to assessing psychopathy with the PCL-R. Future investigations could continue to add to the growing literature on the CATS by examining if the taxon scale is able to predict other risk-related variables such as the revocation of NGRI privileges. This could be accomplished though employing a retrospective research design similar to the present investigation that examines the participants' progression through the later hospital privileges, noting the frequency with which their privileges were suspended or revoked. Future research could also examine whether the PCL-R offers any incremental validity over the CATS by comparing the recidivism rates among discharged patients who score in the high range on each instrument. For example, researchers could follow the progress of discharged individuals who score greater than thirty on the PCL-R and above five on the CATS. Finally, given that female patients were not included in either the present investigation or the research of Bolton (2006), examining whether the CATS predicts risk and review panel decisions in this population would also represent a valuable contribution to the scientific literature on the taxon scale.

The results of the present investigation also raise some questions about the use of the PCL-R. For example, in some jurisdictions (e.g. Louisiana) the PCL-R is administered to every

patient as part of the risk assessment process for person acquitted as NGRI. The results of the present investigation suggest that a large proportion of forensic psychiatric patients have relatively low scores on this instrument. The use of the PCL-R may not be supported in all cases, particularly given the time and expense of the procedure. Consequently, both the risk assessment literature and the institutions charged managing forensic psychiatric patients would benefit from research examining how PCL-R scores are used in these settings. Specifically, examining whether the PCL-R has any predictive utility among individuals who score in the sub-clinical range and whether less costly measures such as the CATS could first be used as a screening measure to identify individuals who might benefit from a more detailed psychopathy assessment.

With respect to research into the decision-making process employed by forensic review panels, future research could examine if demographic variables related to premorbid functioning other than years of education will predict panel dispositions. Variables such as intelligence test scores and employment history could be examined in terms of the variance in decision-making they account for. Moreover, it would also be valuable to investigate whether there are any race related differences in how these variables are employed. Finally, it would be helpful to conduct longitudinal research examining whether the decisions of forensic review panels are able to predict risk-related variables such as aggression within the hospital, privilege revocations, or post-discharge recidivism at a greater than chance rate. Addressing these areas of research could lay the groundwork for more structured and empirically informed decision-making process.



## REFERENCES

- Acheson, S. K. (2005). Review of the Hare Psychopathy Checklist-Revised (2<sup>nd</sup> ed.). In R.A. Spies & B.S. Plake (Eds.), *The sixteenth mental measurements yearbook* (pp. 429–431). Lincoln, NE: Buros Institute of Mental Measurements.
- Andrews, D. A., & Bonta, J. (1998). *The psychology of criminal conduct* (2nd ed.). Cincinnati, OH: Anderson.
- Bartosh, D. L., Garby, T., Lewis, D., & Gray, S. (2003). Differences in the predictive validity of actuarial risk assessments in relation to sex offender type. *International Journal of Offender Therapy and Comparative Criminology*, *47*, 422-438.
- Belmore, M. F., & Quinsey, V. L. (1994). Correlates of psychopathy in a noninstitutionalized sample. *Journal of Interpersonal Violence*, *9*, 339-349.
- Bolton, J. C. (2006). Assessing risk for violent recidivism: A comparison of the Psychopathy Checklist-Revised and the Child and Adolescent Taxon Scale when calculating Violence Risk Appraisal Guide scores. (Doctoral dissertation, Indiana State University, 2006). *Dissertation Abstracts International*, *67*(5-B), 2823.
- Callahan, L. A., & Silver, E. (1998). Factors associated with the conditional release of persons acquitted by reason of insanity: A decision tree approach. *Law and Human Behavior*, *22*, 147-163.
- Cleckley, H. (1972). *The mask of sanity*. Saint Louis, MO: Mosby.
- Cohen, J. (1992). A power primer. *Psychological Bulletin*, *112*(1), 155-159.

Commonwealth of Virginia, Department of Behavioral Health and Developmental Services.

(2003). *Guidelines for the management of individuals found not guilty by reason of insanity*. Retrieved from <http://www.dbhds.virginia.gov/OFO-NGRI.htm>

Cooke, D. J., Hart, S. D., & Michie, C. (2004). Cross-national differences in the assessment of psychopathy: Do they reflect variation in rater's perception of symptoms? *Psychological Assessment, 16*, 335-339.

Doninger, N. A., & Kosson, D. S. (2001). Interpersonal construct systems among psychopaths. *Personality and Individual Differences, 30*, 1263-1281.

Doyle, M., Dolan, M., & McGovern, J. (2002). The validity of the North American risk assessment tools in predicting violent behavior in England. *Legal and Criminological Psychology, 7*, 141-154.

Edens, J. F., & Campbell, J. S. (2007). Identifying youths at risk for institutional misconduct: A meta-analytic investigation of the psychopathy checklist measures. *Psychological Services, 4*, 13-27.

Edens, J. F., Poythress, N. G., & Lilienfeld, S. O. (1999). Identifying inmates at risk for disciplinary infractions: A comparison of two measures of psychopathy. *Behavioral Sciences and the Law, 17*, 435-443.

Edens, J. F., Marcus, D. K., Lilienfeld, S. O., & Poythress, N. G. (2006). Psychopathic, not psychopath: Taxometric evidence for the dimensional structure of psychopathy. *Journal of Abnormal Psychology, 115*, 131-144.

Folsom, J., & Atkinson, J. L. (2007). The generalizability of the LSI-R and the CAT to the prediction of recidivism in female offenders. *Criminal Justice and Behavior, 34*, 1044-1056.

- Glover, A. J. J., Nicholson, D. E., Hemmati, T., Bernfeld, G. A., & Quinsey, V. L. (2002). A comparison of predictors of general and violent recidivism among high-risk federal offenders. *Criminal Justice and Behavior, 29*, 235-249.
- Grann, M., Langstrom, N., Tengstrom A., & Stalenheim, G.E. (1998). Reliability of file-based retrospective ratings of psychopathy with the PCL-R. *Journal of Personality Assessment, 70*, 416-426
- Grove, W. M., & Meehl, P. E. (1996). Comparative efficiency of informal (subjective, impressionistic) and formal (mechanical, algorithmic) prediction procedures: The clinical-statistical controversy. *Psychology, Public Policy, and Law, 2*, 293-323.
- Grove, W. M., Zald, D. H., Lebow, B. S., Snitz, B. E., & Nelson, C. (2000). Clinical versus mechanical prediction: A meta-analysis. *Psychological Assessment, 12*, 19-30.
- Guay, J., Ruscio, J., Knight, R. A., & Hare, R. D. (2007). A taxometric analysis of the latent structure of psychopathy: Evidence for dimensionality. *Journal of Abnormal Psychology, 116*, 701-716.
- Hare, R. D. (1985). Comparison of procedures for the assessment of psychopathy. *Journal of Consulting and Clinical Psychology, 53*, 7-16.
- Hare, R. D. (1993). *Without conscience*. New York, NY: Guilford.
- Hare, R. D. (2002). Psychopathy and risk for recidivism and violence. In N. Gray, J. Laing, & L. Noaks (Eds.), *Criminal justice, mental health, and the politics of risk* (pp. 27-47). London, England: Cavendish Publishing.
- Hare, R. D. (2003). *Hare Psychopathy Checklist-Revised (PCL-R) technical manual*. (2<sup>nd</sup> ed.). New York, NY: Mental-Health Systems.

- Hare, R. D. (2007). Psychological instruments in the assessment of psychopathy. In A. Felthouse & S. Henning, (Eds.). *The international handbook of psychopathic disorders and the law* (pp. 41 - 68). West Sussex, England: John Wiley & Sons, Ltd.
- Hare, R. D., Harpur, T. J., Hakstian, A. R., Forth, A. E., Hart, S. D., & Newman, J. P. (1990). The revised psychopathy checklist: Descriptive statistics, reliability, and factor structure. *Psychological Assessment: A Journal of Consulting and Clinical Psychology*, 2, 338-341.
- Hare, R. D., Hart, S. D., & Harpur, T. J. (1991). Psychopathy and the DSM-IV criteria for antisocial personality disorder. *Journal of Abnormal Psychology*, 100, 391-398.
- Hare, R. D., & Neumann, C. S. (2009). Psychopathy and its measurement. In P. Corr & G. Matthews (Eds.), *Cambridge Handbook of Personality* (pp. 660-686). Cambridge, MA: Cambridge University Press.
- Harris, G. T., Rice, M. E., & Cormier, C. A. (2002). Prospective replication of the Violence Risk Appraisal Guide in predicting violent recidivism among forensic patients. *Law and Human Behavior*, 26, 377-394.
- Harris, G. T., Rice, M. E., & Quinsey, V. L. (1993). Violent recidivism of mentally disordered offenders: The development of a statistical prediction instrument. *Criminal Justice and Behavior*, 20, 315-335.
- Harris, G. T., Rice, M. E., & Quinsey, V. L. (1994). Psychopathy as a taxon: Evidence that psychopaths are a discrete class. *Journal of Consulting and Clinical Psychology*, 62, 387-397.
- Hart, S. D., Forth, A. E., & Hare, R. D. (1990). Performance of criminal psychopaths on selected neuropsychological tests. *Journal of Abnormal Psychology*, 99, 374-379.

- Hart, S. D., & Hare, R. D. (1989). Discriminant validity of the psychopathy checklist in a forensic psychiatric population. *Psychological Assessment, 1*, 211-218
- Hemphill, J. R., Hare, R. D., & Wong, S. (1998). Psychopathy and recidivism: A review. *Legal and Criminological Psychology, 3*, 139-170.
- Hicks, B. M., Markon, K. E., Patrick, C. J., Krueger, R. F., & Newman, J. P. (2004). Identifying psychopathy subtypes based on personality structure. *Psychological Assessment, 16*, 276-288.
- Hill, C., Neumann, C. S., & Rogers, R. (2004). Confirmatory factor analysis of the Psychopathy Checklist: Screening Version (PCL:SV) in offenders with Axis I disorders. *Psychological Assessment, 16*, 90-95.
- Hilton, N. Z., & Simmons, J. L. (2001). The influence of actuarial risk assessment in clinical judgments and tribunal decisions about mentally disordered offenders in maximum security. *Law and Human Behavior, 25*, 393-408.
- Kosson, D., Smith, S. S., & Newman, J. P. (1990). Evaluating the construct validity of psychopathy in black and white inmates: Three preliminary studies. *Journal of Abnormal Psychology, 99*, 250-259.
- Lalumiere, M. L., Chalmers, L. J., Quinsey, V. L., & Seto, M. C. (1996). A test of the mate deprivation hypothesis of sexual coercion. *Ethology and Sociobiology, 17*, 299-318.
- Lalumiere, M. L., & Quinsey, V. L. (1996). Sexual deviance, antisociality, mating effort, and the use of sexually coercive behaviors. *Personality and Individual Differences, 21*, 33-48.
- Lynn, R. (2002). Racial and ethnic differences in psychopathic personality. *Personality and Individual Differences, 32*, 273-316.

- Manguno-Mire, G. M., Thompson, J. W., Bertman-Pate, L. J., Burnett, D. R., & Thompson, H. W. (2007). Are release recommendations for NGRI acquittees informed by relevant data? *Behavioral Sciences and the Law*, *25*, 43-55.
- McDermott, B. E., & Thompson, J. W. (2006). The review panel process: An algorithm for the conditional release of insanity acquittees. *International Journal of Law and Psychiatry*, *29*, 101-111.
- McKee, S. A., Harris, G. T., & Rice, M. E. (2007). Improving forensic tribunal decisions: The role of the clinician. *Behavioral Sciences and the Law*, *25*, 485-506.
- Melton, G. B., Petrila, J., Poythress, N. G., Slobogin, C. (1997). *Psychological evaluations for the courts: A handbook for mental health professionals and lawyers*. (2<sup>nd</sup> ed.). New York, NY: Guilford Press.
- Meehl, P. E., & Golden, R. (1982). Taxometric methods. In P. Kendall & J. Butcher (Eds.), *Handbook of research methods in clinical psychology* (pp. 127–181).
- Morrissey, C., Mooney, P., Hogue, T. E., Lindsay, W. R., & Taylor, J. L. (2007). Predictive Validity of the PCL-R for offenders with intellectual disability in a high security hospital. *Journal of Intellectual and Developmental Disability*, *32*, 125-133.
- Porter, S., Brinke, L. T., Wilson, K. (2009). Crime profiles and conditional release performance of psychopathic and non-psychopathic sexual offenders. *Legal and Criminological Psychology*, *14*, 109-118.
- Quinsey, V. L., Book, A., & Lalumiere, M. L. (2001). A factor analysis of traits related to individual differences in antisocial behavior. *Criminal Justice and Behavior*, *28*, 522-536.

- Quinsey, V. L., Book, A., & Skilling, T. A. (2004). A follow-up of deinstitutionalized men with intellectual disabilities and histories of antisocial behavior. *Journal of Applied Research in Intellectual Disabilities, 17*, 243-253.
- Quinsey, V. L., Harris, G. T., Rice, M. E., & Cormier, C. A. (1998). *Violent offenders: Appraising and managing risk*. Washington, DC: American Psychological Association.
- Quinsey, V. L., Harris, G. T., Rice, M. E., & Cormier, C. A. (2006). *Violent offenders: Appraising and managing risk*. (2<sup>nd</sup> ed.). Washington, DC: American Psychological Association.
- Reiss, D., Grubin, D., & Meux, C. (1999). Institutional performance of male “psychopaths” in a high-security hospital. *The Journal of Forensic Psychiatry, 2*, 290-299
- Snowden, R. J., Gray, N. S., Taylor, J., & Fitzgerald, S. Assessing risk of future violence among forensic psychiatric inpatients with the classification of violence risk (COVR). *Psychiatric Services, 60*, 1522-1526.
- Schroeder, M. L., & Schroeder, K. G., & Hare, R. D. (1983). Generalizability of a checklist for assessment of psychopathy. *Journal of Consulting and Clinical Psychology, 4*, 511-516.
- Serin, R. C. (1992). The clinical application of the Psychopathy Checklist-Revised (PCL-R) in a prison population. *Journal of Clinical Psychology, 48*, 637-642.
- Seto, M. C., Khattar, N. A., Lalumiere, M. L., & Quinsey, V. L. (1997). Deception and sexual strategy in psychopathy. *Personality and Individual Differences, 22*, 301-307.
- Skeem, J. L., Edens, J. F., Sanford, G. M., & Colwell, L. H. (2003). Psychopathic personality and racial/ethnic differences reconsidered: A reply to Lynn (2002). *Personality and Individual Differences, 35*, 1439-1462.

- Skilling, T. A., Harris, G. T., Rice, M. E., & Quinsey, V. L. (2002). Identifying persistently antisocial offenders using the Hare Psychopathy Checklist and DSM antisocial personality disorder criteria. *Psychological Assessment, 14*, 27-38.
- Skilling, T. A., Quinsey, V. L., & Craig, W. M. (2001). Evidence of a taxon underlying serious antisocial behavior in boys. *Criminal Justice and Behavior, 28*, 450-470.
- Tengstrom, A., Grann, M., Langstrom, N., & Kullgren, G. (2000). Psychopathy (PCL-R) as a predictor of violent recidivism among criminal offenders with schizophrenia. *Law and Human Behavior, 24*(1), 45-58.
- Thornquist, M. H., & Zuckerman, M. (1995). Psychopathy, passive-avoidance learning and basic dimensions of personality. *Personality and Individual Differences, 19*, 525-534.
- Toldson, I. A. (2002). The relationship between race and psychopathy: An evaluation of selected psychometric properties of the Psychopathy Checklist Revised (PCL-R) for African American Men. (Doctoral dissertation, Temple University, 2002). *Dissertation Abstracts International, 63*, 553.
- Vasey, M. W., Kotov, R., Frick, P.J., & Loney, B. R. (2005). The latent structure of psychopathy in youth: A taxometric investigation. *Journal of Abnormal Child Psychology, 33*, 411-429.
- Viding, E. (2004). Annotation: Understanding the development of psychopathy. *Journal of Child Psychology and Psychiatry, 45*, 1329-1337.
- Vitacco, M. J., Neumann, C. S., & Jackson, R. L. (2005). Testing a four-factor model of psychopathy and its association with ethnicity, gender, intelligence, and violence. *Journal of Consulting and Clinical Psychology, 73*, 466-476.
- Wrightsmann, L. S. (2001). *Forensic psychology*. Belmont, CA: Wadsworth.



**APPENDIX A: DEMOGRAPHIC DATA FORM**

Participant ID# \_\_\_\_\_

Age \_\_\_\_\_

Age at Index Offense \_\_\_\_\_

Race:      Caucasian                      African American

Year of Education \_\_\_\_\_

Length of Hospitalization \_\_\_\_\_

Number of Previous Hospitalizations \_\_\_\_\_

Length of time before being approved for transfer to a civil psychiatric setting \_\_\_\_\_

Date data was recorded \_\_\_\_\_

**APPENDIX B: VIOLENCE RISK APPRAISAL GUIDE**

**(VRAG) – QUINSEV, HARRIS, RICE, & CORMIER, 1998**

Lived with both parents to age 16 (except parental death)

Yes

No

Elementary School Maladjustment (choose one):

No problems

Slight (minor discipline or attendance) or moderate problems

Severe problems (frequent disruptive behavior and/or attendance or behavior resulting in expulsion or serious suspension)

History of alcohol problems (check all that apply):

Alcohol abuse in biological parent

Teenage alcohol problem

Adult alcohol problem

Alcohol involved in a prior offense

Alcohol involved in the index offense

Marital History (choose one):

Ever married (or lived common law in same home for at least 6 months).

Never married.

Enter number of arrests for each prior (to index) offense type:

Robbery (bank, store)

Robbery (purse snatching)

Arson and fire setting (church, house, barn)

Arson and fire setting (garbage can)

Threatening with a weapon

Threatening (uttering threats)

Theft (grand larceny amount) including car theft and possession of stolen property

Enter number of arrests for each prior offense type that applies:

Mischief to property of grand larceny amount

Break and enter and commit indictable offense (burglary)

Theft or possession of stolen goods under grand larceny amount

Mischief to property under grand larceny amount

Break and enter (includes B and E with intent)

- Fraud (extortion, embezzlement)
- Fraud (forged check, impersonation)
- Possession of a prohibited or restricted weapon
- Procuring for or living on proceeds of prostitution
- Trafficking in narcotics
- Dangerous or impaired driving (DWI)
- Obstructing police officer (including resisting arrest)
- Causing a disturbance
- Wearing a disguise with intent to commit an offense
- Indecent exposure

Failure on prior conditional release (includes parole or probation violation or revocation, failure to comply, bail violation, and any new arrest while on conditional release)?

- Yes
- No

Age at index offense:

Most serious victim injury for index offense (Skip if scoring SORAG)

- Death
- Hospitalized
- Treated and released
- None, slight, or no victim

Any female victim for index offense (Skip is scoring SORAG)?

- Yes     No

Meets DSM-III criteria for any personality disorder?

- Yes     No

Meets DSM-III criteria for schizophrenia?

- Yes     No

Psychopathy Score

**APPENDIX C: PSYCHOPATHY CHECKLIST – REVISED**

**(PCL-R 2<sup>nd</sup> ed.) – HARE (2003)**

Factor 1: Interpersonal/Affective Items

Facet 1: Interpersonal

Glibness/superficial charm  
 Grandiose sense of self worth  
 Pathological lying  
 Conning/manipulative

Facet 2: Affective

Lack of remorse or guilt  
 Shallow affect  
 Callous/lack of empathy  
 Failure to accept responsibility for own actions

Factor 2: Social Deviance Items

Facet 3: Lifestyle

Need for stimulation/proneness to boredom  
 Parasitic lifestyle  
 Lack of realistic long-term goals  
 Impulsivity  
 Irresponsibility

Facet 4: Antisocial

Poor behavior controls  
 Early behavior problems  
 Juvenile delinquency  
 Revocation of conditional release  
 Criminal versatility

Items that do not load on Factor or Facets:

Promiscuous sexual behavior  
 Many short-term marital relationships

**APPENDIX D: CHILD AND ADOLESCENCE TAXON SCALE**

**(CATS) – HARRIS, RICE, & QUINSEY (1994)**

Severe Elementary School (1-8) Maladjustment (frequent disruptive behavior and/or attendance or behavior resulting in expulsion or serious suspensions).

Teenage alcohol problem

Childhood Aggression Rating (check one)

No evidence of childhood aggression

Occasional moderate childhood aggression

Occasional or frequent extreme aggression

Ever suspended or expelled from school

Arrested under the age of 16

Parental alcoholism

Lived with both parents to age 16 (except for death of parents)

**CHECK ALL THAT APPLY BEFORE AGE 15:**

Truancy (at least 5 days per year for at least 2 years, not including last year of school).

Expulsion or suspension from school for misbehavior.

Juvenile delinquency (arrest or referral) for behavior.

Running away from parental or surrogate home at least twice.

Persistent lying.

Repeated sexual intercourse in a casual relationship.

Repeated drunkenness or substance abuse.

Thefts.

Vandalism.

School grades markedly below expectations in relation to estimated or known IQ (may have resulted in repeating a year).

Chronic violations of rules at home and/or at school (other than truancy).

Initiation of fights.

Ever suspended or expelled from school

Arrested under the age of 16

Parental Alcoholism

Lived with Both Parents to age 16