DO STATIC-99R AND PCL-R SCORES PREDICT RISK OPINION EQUALLY ACROSS EVALUATORS IN SVP CIVIL COMMITMENT EVALUATIONS?

A Dissertation

Presented to

The Faculty of the Department of Psychology and Philosophy

Sam Houston State University

In Partial Fulfillment

of the Requirements for the Degree of

Doctor of Philosophy

---

by

Samantha J. Kurus

August, 2022

DO STATIC-99R AND PCL-R SCORES PREDICT RISK OPINION EQUALLY

ACROSS EVALUATORS IN SVP CIVIL COMMITMENT EVALUATIONS?

by

Samantha J. Kurus

APPROVED:

Jorge G. Varela, PhD
Committee Chair

Marcus T. Boccaccini, PhD
Committee Co-Chair

James W. Crosby, PhD
Committee Member

Darrell B. Turner, PhD
Committee Member

Chien-pin Li, PhD
Dean, College of Humanities and Social Sciences

ABSTRACT

Kurus, Samantha J., *Do Static-99R and PCL-R scores predict risk opinion equally across evaluators in SVP civil commitment evaluations?* Doctor of Philosophy (Clinical Psychology), August, 2022, Sam Houston State University, Huntsville, Texas.

Jurisdictional legislation attempts to provide the public protection from sexual offenders. Sexually Violent Predator (SVP) statutes allow for civil commitment of sexual offenders considered high risk for perpetrating additional sexual crimes upon completion of their punitive sentence. SVP evaluations integrate multiple considerations and although structured risk assessment measures decrease the incidence of evaluator differences, research suggests evaluator differences continue to impact clinical opinions. When multiple sources of information contribute to the evaluation, practitioners may place varying levels of importance on any given piece of data. This raises the question: What data do evaluators rely on most strongly to form their ultimate opinions? The current study examined whether scores from two measures used in the course of SVP evaluations, the Static-99R and the Psychopathy Checklist-Revised (PCL-R), predicted ultimate risk opinions and determination of behavioral abnormality equally across evaluators. Data was extracted from the evaluation records of 393 incarcerated men convicted of sexual offences who were evaluated by one of three evaluators for civil commitment in the state of Texas. It was expected that Static-99R scores would predict risk opinion equally across evaluators. With regard to PCL-R factors and facets, it was expected that Factor 2 scores, Facet 4 in particular, would predict risk opinion equally across evaluators, while Factor 1 scores might differentially predict risk. Results indicated that PCL-R Facet 4 scores did predict risk opinion equally across evaluators, as expected. All other relationships between measure score and risk opinion differed across evaluators. With regard to behavioral abnormality, it was expected that Static-99R scores would be differentially predictive across evaluators, but that PCL-R scores would be equally predictive. Results showed that PCL-R Facet 4 scores did equally predict behavioral abnormality opinion across evaluators. Static-99R scores predicted opinion about behavioral abnormality as well, which was not expected. All other relationships between measure scores and behavioral abnormality differed by evaluator, also contrary to hypothesis. Potential contributing factors to these varied results are discussed.

KEY WORDS: Forensic assessment; Risk assessment; Sexually violent predator; Evaluator differences; Static-99R; Psychopathy Checklist Revised; Behavioral abnormality; Civil commitment

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

Figure **Page**

CHAPTER I

Introduction

Volumes of policies and statutes have been created and revised with the intention of protecting the public from violent others. Many such policies address the heightened concern that those who have engaged in violent behaviors will do so again. Sex offenders are often considered to be a particularly critical threat to public safety, and concerns about the potential for an individual who has been convicted of a sexual offense to commit another sexual offense upon release abound. Indeed, surveys of public beliefs reveal that the majority of individuals believe sex offenders to be at extremely high risk for sexual re-offense (Levenson, Brannon, Fortney, & Baker, 2007). For example, respondents in Levenson & colleagues' (2007) study asserted that 74% of sex offenders will commit another sexual offense. In contrast, empirical data suggest that recidivism rates within the sex offender population are generally lower than in other offending populations (Hanson, Morton, & Harris, 2003), with reports ranging from 3% to 13% (Calleja, 2015; Vess & Skelton, 2010; Hanson & Morton-Bourgon, 2009). However, this concern raises an important question: how do we know which offenders will reoffend?

One way of addressing the public's concern and of managing particularly high-risk sex offenders is through Sexually Violent Predator (SVP) statutes. SVP statutes were established to help ensure public safety via civil commitment of those offenders thought to pose a continued risk of sexual violence after their period of incarceration. While the history of these statutes has endured some controversy with regard to challenges to due process and double jeopardy rights, the landmark case of *Kansas v Hendricks* (1997) established that, as long as these laws are enacted with the intent of incapacitating individuals who are dangerous, and not with the intent of further punishing these individuals, they are, in fact, constitutional. As such, risk assessment is an integral part of the SVP civil commitment process to ensure only particularly high-risk individuals who truly pose a danger to society are civilly committed. SVP determination is made in accordance with optional state statutes, which 20 of the 50 states have currently adopted (see DeMatteo et al., 2015 for review of state statutes).

Violence Risk Assessment

Evaluations of risk for future violence incorporate information from a variety of sources. For example, evaluators often review historical records, including legal, psychiatric, and institutional records, to gather information about past behavior and functioning. Additionally, recent functioning with regard to interpersonal relationships, affective, cognitive, and behavioral functioning, as well as treatment compliance may be reviewed. Actuarial measures are often completed and clinical interviews conducted to aid in assessing risk.

Research suggests that the most accurate estimates of future violence are provided by the results of standardized risk assessment measures (Guy, 2008; Hanson & Morton-Bourgon, 2009; Doren, 2002). Risk assessment instruments have been empirically shown to provide more reliable and accurate prediction of general and violent re-offense (Grove, Zald, Lebow, Snitz & Nelson, 2000; Ægisdóttir et al., 2006) and of sexual recidivism than clinical judgment (Hanson & Morton-Bourgon, 2004). As such, the use of standardized risk assessment instruments is becoming increasingly routine in the process of evaluating the risk that an individual will commit future offenses (Monahan & Skeem, 2016). Forensic risk assessment instruments are used in a wide variety of contexts.

Clinicians utilize these instruments in making decisions about pre-trial matters, sentencing decisions, parole release decisions, and insanity acquittee placement and release decisions (Viljeon, Jonnson, Cochrane, Vargen, & Vincent, 2019), just to name a few.

Sexual Violence Risk Assessment

Specialized evaluations allow for more nuanced assessment of risk of particular types of violence. For example, sexual violence risk assessment is a crucial part of the SVP process. Statutes vary by state, though laws generally stipulate that for civil commitment as an SVP an individual has been convicted of multiple crimes that were sexual in nature, has been determined to have a mental illness, is predisposed to commit sexually violent offenses, and, as a result, poses a continued risk to the community (Miller, Amenta, & Conroy, 2005). The evaluation process for such a determination is likewise governed by states, and generally includes evaluation of the inmates' mental health and risk for sexual dangerousness as a result of his or her mental health. As with other evaluations of risk, clinical interviews are typically conducted, and standardized risk assessments are administered.

While specific assessment instruments utilized in any given evaluation likely vary by evaluator, the requirements of the jurisdiction in which an evaluator practices also influence procedure. In Texas, the cases of incarcerated offenders with a history of multiple qualifying sexual offenses (i.e., primarily contact sexual offenses) are reviewed by a multidisciplinary team that includes various criminal justice, law enforcement, and mental health professionals. Offenders considered to be at high risk are referred to an external contractor for an evaluation to determine if s/he manifests a mental illness.

Following the decision in Kansas v. Hendricks (1997), distinctions between dangerousness of an individual generally and dangerousness associated with mental illness were drawn. For example, the decision in Kansas v. Crane (2002) asserted that, for civil commitment to be warranted, a mental abnormality must be relevant enough to the behaviors in question and severe enough to "distinguish the dangerous sexual offender…from the dangerous but typical recidivist convicted in an ordinary criminal case." As such, assessment of mental illness has become essential to the SVP evaluation process. In Texas, the presence of mental illness is referred to as the behavioral abnormality (BA) requirement (Tex. Health & Safety Code Ann. § 841.023, 2011). Moreover, the Texas statute specifically requires that assessment of psychopathy be included in the evaluation of behavioral abnormality, presumably as an examination of personality traits or disorder.

Evaluators are also tasked with providing a summary opinion as to the individuals' overall level of risk for committing future offenses of a sexual nature. Those offenders found to be at high risk may be recommended for civil commitment, typically for an indeterminate amount of time, following the completion of their sentence.

Standardized Sexual Violence Risk Assessment

Numerous assessment instruments have been created to aid in the specific measurement of risk factors for future sexual violence, such as the Static-99R (Hanson & Thornton, 2000; Helmus, Thornton, Hanson, & Babchishin, 2012), Sex Offender Risk Appraisal Guide (SORAG; Quinsey, Harris, Rice, & Cormier, 1998), the Sexual Violence Risk-20 (SVR-20; Boer, Hart, Kropp, & Webster, 1997), Screening Scale for Pedophilic Interests (SSPI; Seto & Lalumiere, 2001), Violence Risk Scale: Sex Offender Version

(VRS:SO; Wong, Olver, Nicholaichuk, & Gordon, 2004), and the Structured Assessment of Violence Risk in Youth (SAVRY; Borum, Bartel, & Forth, 2006).

Static-99R. The Static-99R is specifically designed for measurement of risk of sexual re-offense. It is comprised of 10 items concerning static variables that relate to risk for future sexual offense, including several offense history details, offender age, and cohabitation history. Items are scored for the presence or absence of the factor, with higher total scores indicating higher risk for recidivism. It is the most widely used assessment for predicting sexual recidivism and has been determined time and again to exhibit strong predictive validity (Helmus et al., 2012; Helmus & Thornton, 2015; Boccaccini et al., 2017). For example, meta-analytic research has revealed strong effects with respect to sexual recidivism (d = .74, Hanson & Morton-Bourgon, 2009; AUC = 0.705, Helmus, Hanson, Thornton, Babchishin, & Harris, 2012). Additional recent reviews of the predictive accuracy of the Static-99R have shown somewhat lower yet still moderate results. Boccaccini and colleagues (2017) revealed effect sizes of d = .55 - .56 (AUC = .623 - .650) for the Static-99-R scores' prediction of sexual recidivism in a large Texas sample.

Psychopathy Checklist-Revised (PCL-R). Hare's Psychopathy Checklist-Revised (PCL-R; 2003) is an assessment instrument designed to measure the presence of psychopathic traits. It is comprised of 20 items, yielding four facet scores and two overarching factor scores. Factor 1 includes Facets 1 (Interpersonal) and 2 (Affective), while Factor 2 includes Facets 3 (Impulsive Lifestyle) and 4 (Antisocial Behavior). A total score is also calculated. Higher scores are associated with higher levels of psychopathic traits.

PCL-R Factor 1 (facets 1 and 2) items are considered to reflect personality traits. For example, items included under Factor 1 examine grandiose sense of self-worth, glibness or superficial charm, pathological lying, callousness or lack of empathy, and manipulativeness. These items might be considered to require somewhat more subjective rating than those of Factor 2. PCL-R Factor 2 (facets 3 and 4) is considered to be a reflection of behavioral and lifestyle factors, addressing factors such as parasitic lifestyle, need for stimulation, impulsivity, poor behavioral control, early behavioral problems, and versatility of criminal history.

The PCL-R is one of the most commonly used instruments in assessment of offender risk (Boccaccini, Turner, & Murrie, 2008), as it has been associated with risk for general, violent, and sexual recidivism (Hemphill, Templeman, Wong, & Hare, 1998; Salekin, Rogers & Sewell, 1996) as well as associated with other antisocial behaviors (Leistico, Salekin, DeCoster, & Rogers, 2008). While the PCL-R is not designed to be a sexual violence risk instrument, recent studies suggest that high psychopathy scores, when specifically combined with sexual deviance, have been shown to be strong predictors of sexual recidivism (Hawes et al., 2013). A meta-analysis analyzing 20 studies reported an overall effect size of $d = 0.40$ for PCL-R Total scores' prediction of sexual recidivism (Hawes, Boccaccini, & Murrie, (2013). Hence, it is no surprise this instrument is often used as a measure of psychopathy in SVP evaluations. With regard to the specific factors and facets assessed using the PCL-R, Factor 2 ($d = 0.44$, $p < .001$) and Facet 4 ($d = 0.40$, $p < .001$), specifically, have been shown to be strong predictors of sexual recidivism among sexual offenders (Hawes, Boccaccini, & Murrie, 2013; Walters, Knight, Grann, & Dahle, 2008).

Behavioral Abnormality

A variety of methods are available to examine for mental illness and behavioral functioning. Measurement of personality traits can help inform clinicians about the presence or absence of mental illness/behavioral abnormality, lending another reason for evaluators to include the PCL-R in SVP evaluations. For example, in unpublished analyses conducted previously using current data, AUC for PCL-R total scores and manifestation of behavioral abnormality was 0.73 ($p \leq .001$, 95% CI: 0.68 – 0.77), suggesting that individuals with higher PCL-R scores were more likely to be determined to manifest behavioral abnormality than those assigned lower scores (Kurus, et al., 2019). Cohen’s d effect size for PCL-R scores between those with and without behavioral abnormality is 0.85 (Kurus et al., 2019).

Further, the same unpublished analyses revealed associations between scores on the Static-99R and determination of the presence or absence of behavioral abnormality. Specifically, AUC for Static-99R total scores and manifestation of a behavioral abnormality was 0.725 ($p \leq .001$), indicating that individuals with higher Static-99R scores were more likely to be determined to manifest a BA than were those assigned lower scores (Kurus, et al., 2019). Cohen’s d effect size for Static-99R scores between those with and without a BA was 0.87 (Kurus, et al, 2019). However, additional literature addressing the relationship between risk assessment scores and the determination of behavioral abnormality is sparce.

Inter-Rater Agreement

While the use of standardized instruments does tend to improve prediction of risk, it is also important to recognize the role of human participation in scoring these instruments. The utility of a measure in real-world practice relies, in part, on rater agreement across evaluators. Although we hope that the manualized instructions and, ideally, proper training clinicians receive with regard to administering and scoring risk assessment instruments will result in evaluators' identical scoring, interpretation, and application of items, evaluator differences nonetheless may influence the scores assigned by evaluators and the resulting implications. It is, thus, important to examine inter-rater reliability of the measures used to make determinations of risk.

Rater agreement analyses allow researchers to examine the amount of variability in scores that is due to actual differences in traits possessed by the evaluee, thereby also allowing for the examination of the amount of variability due to error, including rater differences. Intraclass correlation coefficient (ICC) is a ratio of variance attributable to the individual evaluated divided by the person variance plus error (McGraw & Wong, 1996; Shalverson & Webb, 1991), i.e., a ratio of true possession of traits to measurement error. Absolute agreement ICC examines the specific values of scores assigned (as opposed to relative high versus low scores across evaluators).

Examinations of rater agreement in forensic assessments in practice have revealed varying levels of inter-rater reliability. In recent years, concern about discrepancies in ultimate opinions have suggested that evaluator disagreement is found in 15%-30% of competency to stand trial cases, 25%-35% of sanity cases (Guarnera & Murrie, 2017), and 45% of conditional release assessment cases (Acklin, Fuger, & Gowensmith, 2015).

With regard to sexual violence risk assessment specifically, the Static-99R has repeatedly been found to show good inter-rater reliability in research studies. Intraclass correlation coefficients (ICCs) of .85 to .90 are reported by the test developers (Barbaree et al., 2001), and additional reviews have reported agreement values of .90 (Hanson & Morton-Bourgon, 2009; Helmus, 2008). Similarly, the PCL-R has also shown consistently good inter-rater reliability in controlled studies. The instrument manual reports ICCs of .86 to .94 (Hare, 2003). Additional research studies corroborate good reliability, with ICCs of .84 (Levenson, 2004a).

Despite very high inter-rater reliability of the Static-99R and PCL-R reported in lab studies, over the last decade or so many researchers have examined the field reliability of measures and found that many measures perform less reliably in real-world clinical practice than in highly controlled research studies. For example, literature examining the application of Static-99R scores in clinical practice (as opposed to in research coding) indicates moderate rater agreement (ICC = .78; Miller, et al., 2012; ICCs = .79 & .88; Boccaccini, Murrie, Mercado, Quesada, Hawes, Rice, & Jeglic, 2012).

Likewise, although lab trial results indicate excellent rater-agreement, research regarding the PCL-R scores assigned by those administering the measure in clinical practice suggests a broader range of reliability, including ICC values above .90 (Kroner & Mills, 2001; Porter et al., 2003; Looman, Morphett, & Abracen, 2012; Rettenberger, Matthes, Boer, & Eher, 2010), as well as ICC values “lower than .80” (Boccaccini, Turner, Murrie, & Rufino, 2012; Lloyd, Clark, & Forth, 2010), ICC values of .60 (Rutherford, Cacciola, Alterman, McKay & Cook, 1999; Miller et al., 2012), and of .39 (single rater absolute ICC; Murrie et al., 2008). Boccaccini, Murrie, Rufino, and Gardner

(2014) found that about 32% of the variance in the PCL-R Total scores in their sample was attributable to differences among evaluators.

With regard to reliability of the PCL-R factor and facet scores, agreement tends to be good. For example, Porter and colleagues (2003) reported high ICC values: Factor 1 ICC = .81; Factor 2 ICC = .95. Ismail and Looman (2018) revealed single measure ICC for the PCL-R Factor 1 was .78 and for Factor 2 it was .90.

However, agreement coefficients for the PCL-R factors and facets, just like the Total score, are not immune to variability. Many researchers have found greater rater agreement for Factor 2 than for Factor 1 (Ismail & Looman, 2018; Edens, Boccaccini, Johnson, & Johnson, 2010), and greater agreement for Facet 4 than for Facets 1, 2, or 3 (Edens, Boccaccini, Johnson & Johnson, 2010; Miller at al., 2012). These findings make sense when considering that Factor 2 items may require less subjective judgment to score. They also coincide with the earlier reported trends in predictive validity of the factors and facets. Conversely, other studies have reported substantial scoring variability on both Factor 1 and Factor 2 due to evaluator differences, as well as for Facets 3 and 4 (Boccaccini, Murrie, Rufino, & Gardner, 2014). Clearly, scores across raters in real-world practice vary in their consistency.

Evaluator Use of Measures

As aforementioned, evaluators use a host of different information to form their opinions of an offenders' risk, including clinical interviews, records reviews, and assessment data. Moreover, clinicians not only use the results of one assessment instrument differently, but they differ in their integration of the results of several risk measures per case. Neal & Grisso (2014) revealed that most evaluators report the use of multiple assessment measures during the course of evaluations of risk – an average of 4 measures per case, and as many as 15 different risk assessment instruments per their findings. But, do all evaluators use the information acquired from these instruments to the same degree to inform their responses to their forensic questions in real-world practice?

As stated by McCallum, Boccaccini, and Bryson (2017), the "overall picture of contemporary risk assessment practice is to gather a whole bunch of information from different sources, administer several instruments, and integrate all or parts of the information to come to a single conclusion about risk" (p. 1214). In light of these trends in evaluation and the variability in evaluators' use of these sources, there is cause to wonder what real-life implementation of assessment looks like. How are evaluators incorporating the information gathered from multiple sources of information, in particular, the scores from multiple risk assessment instruments, to come to an informed opinion in actual practice?

Early studies suggested that evaluators' risk opinions were scarcely influenced by the results of risk assessments administered when other information was also available. For example, with regard to security placement needs of insanity acquittees, Violence Risk Appraisal Guide (r = -.06; VRAG; Quinsey, Harris, Rice, & Cormier, 1998) scores were not significantly associated with evaluators' recommendations (Hilton & Simmons, 2001). A follow-up study suggested that both VRAG and PCL-R scores were significantly associated (r = .33 to .47) with recommendations (Hilton, Simpson, & Ham 2016), but these findings still indicate only a modest role of assessment instruments results in forming a risk opinion. A recent examination of the association between scores on the HCR-20, LS/CSI, and PCL-R and summary risk ratings revealed statistically significant, though variable, correlations (Guy, Kusaj, Packer & Douglas, 2015). Specifically, correlations for HCR-20 indices ranged from .55 to .73; correlations for the LS/CSI indices ranged from .27 to .67; and PCL-R score correlations ranged from .37 to .64. The variability reported in the associations between actuarial scores and ultimate opinions/recommendations/ratings suggests that, while evaluators of risk do take assessment instrument results into account, there is wide variability across evaluators with regard to the extent to which those results factor into ultimate findings.

Previous literature examining the integration practices of forensic evaluators have revealed just that: evaluators do not similarly take into account the information acquired during the course of an evaluation. Some evaluators base their opinions solely on the results of actuarial measures, some base their opinions partially on such results, and still others report they administer instruments as a requirement but do not use those results in making their opinions (Miller & Maloney, 2013). Evaluation of risk for sexual re-offense, specifically, is no different with regard to variable predictivity of actuarial measures on ultimate opinions. For example, Chevalier, Boccaccini, Murrie, and Varela (2015) found that 49% of evaluators reported the results of administered Static-99Rs had “some” influence on their decision of offender risk, while 42% reported the assessment instrument findings had “a lot” of influence. One evaluator in their study reported basing his/her opinion of risk solely on the results of the Static-99R. Clearly, evaluators differ in their use of the results of the forensic risk measures they administer.

Indeed, Kahn and colleagues (2022) recently also found significant differences in the weight assigned to relevant, and even irrelevant, case details involved in SVP cases. Further, their results showed that the recommendation offered with regard to civil commitment appropriateness was strongly influenced by which evaluator was assigned to any given case.

Levenson & Morin (2006) found that factors measured by actuarial instruments were all significantly predictive of civil commitment of those recommended for SVP evaluation. However, they also reported that diagnoses of pedophilia or paraphilia NOS, the presence of psychopathy, and several factors with no empirical association with sexual recidivism were predictive of civil commitment decisions – young victim and nonminority race. Recently, and perhaps most encouragingly, McCallum, Boccaccini, and Bryson (2017) examined sex offender evaluations conducted in Colorado and reported that scores from the three actuarial measures used (Static-99, VRAG, SORAG) were each significantly associated with summary risk ratings (r = .72, .63, and .81, respectively), suggesting that even though several instruments were incorporated, the information resulting from each were used to inform opinions. But, when both the results of actuarial data and non-instrument risk factor information were incorporated, only the non-instrument data were predictive of risk containment recommendations (McCallum, Boccaccini, & Bryson, 2017).

Considering the critical impact of the results of SVP evaluations on the lives of the offenders, as well as on the resources allocated for their care and treatment, the process by which these decisions are made should be examined. When evaluators use several points of information, namely, multiple assessment measures, how do they incorporate all of the data to come to a conclusion? Do evaluators take the results of risk measures into account similarly?

CHAPTER II

Current Study

Given what we know about the variability in rater scoring, as well as previously reported variability in the extent to which any given assessment measure is taken into account in real-world clinical practice, the purpose of the current study was to examine whether Static-99R and PCL-R scores, when both administered in the course of SVP evaluations, are equally predictive across evaluators of ultimate risk opinion. To date, very few studies have examined these trends in effect on ultimate opinion across evaluators. Further, only a couple studies appear to have looked at the agreement across evaluators of PCL-R factors and facets – most examine PCL-R total scores - and no studies have addressed whether evaluators differentially use these specific facet scores in forming their risk opinions.

Specifically, I intended to determine whether Static-99R scores predict risk opinion equally across evaluators and whether PCL-R total and facet scores predict risk equally across evaluators. The Static-99R has previously demonstrated good field reliability and evaluators seem to have great familiarity with the measure. The static nature of the items to be scored suggests the scoring process is one of an objective task, which should result in little error due to evaluator differences. High predictive validity of the measure suggests professionals would value the results of the measure. Therefore, it was expected that Static-99R scores would predict risk opinion equally across evaluators.

The wide range of variability in inter-rater agreement correlations reported in field studies concerning the PCL-R made it difficult to surmise whether PCL-R scores would perform equally across evaluators in the current study. On one hand, previous studies report that clinicians working for the same side of the aisle tend to show more agreement than those on opposing sides of the adversarial picture. Evaluators in the current study are meant to act as neutral parties – neither defense nor for the state. Further, all evaluators here operated under Texas statute, which requires that measures of psychopathy be incorporated into the process. As such, and in light of many reports of high rater agreement, it might have been the case that clinicians in the current study would take PCL-R scores into account similarly in opining risk level.

On the other hand, the volume of literature to date reflecting agreement values from .39 through .93 suggest that low evaluator agreement might be found. Additionally, there is an indirect connection between psychopathy ratings and risk of sexual violence. As such it was believed that evaluators might rely to varying degree on PCL-R Total scores in concluding about risk of committing additional offenses sexual in nature.

As a small trend appears to be emerging with regard to stronger inter-rater reliability for PCL-R Factor 2 than Factor 1, it was expected that Factor 2 scores (Facets 3 and 4) would predict risk opinion equally across evaluators. As a suggested reason for lower reliability in Factor 1 scores relates to the subjectivity of the items to be scored, it was believed it might be the case here that Factor 1 scores (Facets 1 and 2) differentially predict risk opinions.

Additionally, examination of the process of determining BA was of interest here. Most often, both risk and behavioral abnormality are assessed within the same evaluation period. It stands to reason, then, that any findings or clinical judgements as a result of the risk assessment instruments administered may influence the evaluator's BA opinion. But, are these findings influenced similarly across differing evaluators? Specifically, I intended to assess whether Static-99R and PCL-R total and facet scores informed an evaluator's opinion of the presence or absence of BA, and whether that influence was equally predictive across evaluators.

Little previous research is available to inform expectations of the relationships between risk assessment scores and BA findings. However, as commission of a sexual offense, in itself, does not necessitate diagnosis of a mental illness, a direct correlation between Static-99R scores (designed to measure risk for sexual offense) and BA is not intuitive. Neither do the individual items on the measure directly correspond to traditional criteria for a mental health diagnosis. However, preliminary analyses mentioned above again indicate a strong relationship between Static-99R scores and BA opinion. But, with regard to separate evaluators, it was believed that the relationship between Static-99R scores and BA finding might differ, given the indirect relationship between risk and mental health diagnosis.

Further, the above-noted prior analyses indicate predictive value of PCL-R Total scores in identifying BA opinion (Kurus, et al., 2019). Additionally, given that personality traits can certainly contribute to diagnosis of mental illness, and personality disorder is not precluded from a finding of behavioral abnormality in the current jurisdiction, it was reasonable to believe that measures of personality characteristics, in this case, PCL-R scores, might predict behavioral abnormality opinion. It was expected the facet scores would evidence a similar trend.

Research Question and Statistical Methods

For the current study, I examined how evaluators use the Static-99R and PCL-R when formulating risk level opinions of offenders being considered for commitment as sexually violent predators. To this end, risk opinion was regressed on to Static-99R and various PCL-R scores, albeit in six differing analyses; specifically, the following: risk opinion was regressed onto 1) Static-99R scores, 2) PCL-R Total scores, 3) PCL-R Facet 1 scores, 4) PCL-R Facet 2 scores, 5) PCL-R Facet 3 scores, and 6) PCL-R Facet 4 scores. To specifically assess for variations across evaluators, in each regression "evaluator" was entered as a potential moderator and interactions were assessed.

Additionally, I examined how Static-99R and PCL-R scores influence an evaluator's determination of the presence or absence of BA. As such, BA was regressed on to Static-99R and the various PCL-R scores, entering evaluator as a potential moderating variable, albeit in six differing analyses; specifically, the following: BA was regressed onto 1) Static-99R scores, 2) PCL-R Total scores, 3) PCL-R Facet 1 scores, 4) PCL-R Facet 2 scores, 5) PCL-R Facet 3 scores, and 6) PCL-R Facet 4 scores. Again, to specifically examine any differences across evaluators, interactions were assessed in all models.

Any observed influence of assessment scores on risk level or BA opinions were then assessed for particular trends in the relationships, via correlation analyses. Specifically, ways in which the relationship between the assessment score and the respective outcome variable varied by evaluator were examined.

CHAPTER III

Methods

Sample

Data included in the current study were extracted from the evaluation records of 393 incarcerated sexual offenders who were evaluated for civil commitment as SVPs by one of three evaluators in Texas. The evaluations included were conducted between 2012 and 2019 and represent all of the evaluations conducted by each of the three evaluators during that time. The evaluators who conducted the current evaluations were all male and all licensed doctoral level clinicians who have completed specialized training in forensic psychology. Of those evaluations reviewed, 87 were conducted by Evaluator 1, 158 by Evaluator 2, and 139 by Evaluator 3. The remaining 9 were excluded from analyses due to lack of identification with regard to evaluator. Evaluations were conducted following offender referral from the multidisciplinary team. All evaluations occurred in correctional facilities. In the jurisdiction in which the current evaluations were conducted, offenders must have been convicted of at least two qualifying offenses in order to be considered for civil commitment as an SVP, which tend to be contact sex offenses. As such, all offenders in the current sample were convicted of at least one contact sexual offense.

Measures

Results from several measures were included in the current evaluations. For the purpose of this study, two of the measures administered during the course of these SVP evaluations were examined.

Static-99R

The Static-99R is a 10-item measure specifically designed to examine an offender’s risk for sexual re-offense (as opposed to general or violent re-offense). Items pertain to static, historical factors, such as offense characteristics, previous offense history, and previous lifestyle characteristics. Total scores are then used to report level of risk posed by the evaluee. For example, test developers suggest that scores of 0 or 1 indicate Low risk; scores of 2 or 3 indicate Low-Moderate risk; scores of 4 or 5 indicate Moderate-High risk, and scores of 6 and above indicate High risk for recidivism. In evaluation records available for the current study evaluators provided item and total scores, but did not report a risk level category specific to Static-99R scores.

Psychopathy Checklist-Revised (PCL-R)

The PCL-R (Hare, 2003) consists of 20 items designed to measure the presence or absence of psychopathic traits. Items are scored by a clinician typically with use of information from review of records and clinical interview. As mentioned above, the items provide information specifically with respect to interpersonal and affective functioning, as well as engagement in impulsive lifestyle and antisocial behaviors. A total score provides an overall measure of psychopathy.

Risk Level Opinion

Evaluators’ reports included an overall risk level opinion based upon the totality of all evaluation procedures, including review of records, clinical interview, and results of assessment measures. The evaluators reported the following range of risk levels: Low, Low-to-Moderate, Moderate, Moderate-to-High, High, Very High.

Procedure

Ethical approval was obtained during initial data collection by the Institutional Review Board of Sam Houston State University. Evaluation records included here were reviewed and reports written by one of three independent evaluators licensed in the state of Texas. In each file, evaluators reported the item and total scores of several risk assessment instruments, including the Static-99R and the PCL-R. In conjunction with other data sources, the evaluators then provided an overall Risk Level opinion following the six-category risk scale: Low, Low-to-Moderate, Moderate, High-to-Moderate, High, Very High. An opinion regarding the presence or absence of qualifying BA was also provided.

As part of a larger study, evaluation reports were coded on a number of factors. For current purposes, individual item scores, factor scores, facet scores, overall risk opinion, and BA opinion were examined.

CHAPTER IV

Results

Demographics

Demographic information for the evaluees appears in Tables 1 and 2. The average age of offenders in the overall sample was 48.17 (*SD* = 12.08). The three evaluator subsamples did differ significantly by age, $F(2,376) = 6.35$, $p=.002$, $\eta_p^2 = .03$. Specifically, evaluees on Evaluator 2's caseload (*M* = 50.76, *SD* = 11.21) were slightly older than those of Evaluator 1 (*M* = 46.72, *SD* = 12.31) or 3 (*M* = 46.13, *SD* = 12.43). Overall, evaluees' race was identified as 53.7% White, 26% Black, 18.1% Latino, and 6% Other. Offender race/ethnicity did not differ significantly across evaluators, $X^2(6, N = 381) = 5.88$, $p = .44$, Cramer's $V = .09$.

Table 1

Evaluee Age by Evaluator

	M	*SD*
Total Sample	48.17	12.08
Evaluator 1	46.72	12.31
Evaluator 2	50.76	11.21
Evaluator 3	46.13	12.43

Table 2

Evaluee Race by Evaluator

	Race	%
Total Sample	White	53.7
	Black	26.2
	Latino	18.1
	Other	1.5
Evaluator 1	White	56.3
	Black	26.4
	Latino	17.2
	Other	0
Evaluator 2	White	50
	Black	26.6
	Latino	19.0
	Other	3.2
Evaluator 3	White	57.6
	Black	23.7
	Latino	17.3
	Other	.7

Overall Measure Scores

Static-99R scores ranged from -3 to 10, with an average of 2.67 (*SD* = 2.24). Evaluator 1's Static-99R scores ranged from -2 to 9 (*M* = 3.09, *SD* = 2.04), scores assigned by Evaluator 2 ranged from -3 to 10 (*M* = 2.26, *SD* = 2.27), and Evaluator 3's Static-99R scores ranged from -3 to 8 (*M* = 2.86, *SD* = 2.28). See Table 3. There were significant differences in Static-99R scores between evaluators, $F(2,376) = 4.64$, p=.01**,** $\eta_p^{\ 2} = .02$. Specifically, independent samples t-tests revealed that Evaluator 1's Static-

99R scores were statistically higher than Evaluator 2's Static-99R scores, *t*(239) = 2.81, *p* = .005, *d* = .38.

Table 3

Static-99R Scores

	M	*SD*
Total Sample (*N* = 382)	2.67	2.24
Evaluator 1 (*n* = 85)	3.09	2.04
Evaluator 2 (*n* = 156)	2.26	2.27
Evaluator 3 (*n* = 138)	2.86	2.28

Overall, PCL-R Total scores ranged from 1 to 36, with an average of 18.40 (*SD* = 7.01). Scores assigned by Evaluator 1 ranged from 4 to 33 (*M* = 19.81, *SD* = 8.05). Evaluator 2's PCL-R Total scores ranged from 7 to 36 (*M* = 20.52, *SD* = 5.53) and Evaluator 3's scores ranged from 1 to 31 (*M* = 16.18, *SD* = 7.33). See Table 4. There were significant differences between evaluators in PCL-R Total scores, *F*(2,356) = 15.61, $p < .001$, $\eta_p^2 = .08$. Independent samples t-tests showed that Evaluator 3 assigned significantly lower PCL-R Total scores than Evaluator 1, *t*(212) = -3.33, *p* = .001, *d* = -.48) and Evaluator 2, *t*(282) = -5.65, $p < .001$, *d* = -.67).

Table 4

PCL-R Total Scores

	M	*SD*
Total Sample (*N* = 366)	18.40	7.01
Evaluator 1 (*n* = 75)	19.81	8.05
Evaluator 2 (*n* = 145)	20.52	5.53
Evaluator 3 (*n* = 139)	16.18	7.33

Across the sample, PCL-R Facet 1 scores ranged from 0 to 8, with an average of 3.34 (*SD* = 2.31). Scores assigned by Evaluator 1 ranged from 0 to 8 (*M* = 3.62, *SD* = 2.70). Evaluator 2's PCL-R Facet 1 scores ranged from 0 to 8 (*M* = 3.67, *SD* = 1.63) and Evaluator 3's scores ranged from 0 to 8 (*M* = 2.78, *SD* = 2.58). See Table 5. There were significant differences between evaluators in PCL-R Facet 1 scores, $F(2,353) = 6.34$, $p = .002$, $\eta_p^2 = .03$. Specifically, Evaluator 3 assigned significantly lower PCL-R Facet 1 scores than Evaluator 1, $t(208) = -2.21$, $p = .03$, $d = -.32$) and Evaluator 2, $t(282) = -3.51$, $p = .001$, $d = -.42$).

Table 5

PCL-R Facet 1 Scores

	M	*SD*
Total Sample (*N* = 360)	3.34	2.31
Evaluator 1 (*n* = 72)	3.62	2.70
Evaluator 2 (*n* = 146)	3.67	1.63
Evaluator 3 (*n* = 138)	2.78	2.58

Overall, PCL-R Facet 2 scores ranged from 0 to 8, with an average of 4.98 (*SD* = 2.38). Scores assigned by Evaluator 1 ranged from 0 to 8 (*M* = 4.76, *SD* = 2.85). Evaluator 2's PCL-R Facet 2 scores ranged from 2 to 8 (*M* = 5.72, *SD* = 1.33) and Evaluator 3's scores ranged from 0 to 8 (*M* = 4.25, *SD* = 2.72). See Table 6. There were significant differences between evaluators in PCL-R Facet 2 scores, $F(2,355) = 15.17$, $p < .001$, $\eta_p^2 = .08$. Specifically, Evaluator 2 assigned significantly higher PCL-R Facet 2 scores than Evaluator 1, $t(218) = 3.44$, $p = .001$, $d = .49$, and Evaluator 3, $t(284) = 5.86$, $p < .001$, $d = .69$.

Table 6

PCL-R Facet 2 Scores

	M	*SD*
Total Sample (*N* = 362)	4.98	2.38
Evaluator 1 (*n* = 72)	4.76	2.85
Evaluator 2 (*n* = 148)	5.72	1.33
Evaluator 3 (*n* = 138)	4.25	2.72

PCL-R Facet 3 scores ranged from 0 to 10, with an average of 4.53 (*SD* = 2.45). Scores assigned by Evaluator 1 ranged from 0 to 10 (*M* = 5.07, *SD* = 2.60). Evaluator 2's PCL-R Facet 3 scores ranged from 1 to 9 (*M* = 4.99, *SD* = 1.85) and Evaluator 3's scores ranged from 0 to 10 (*M* = 3.81, *SD* = 2.73). See Table 7. There were significant differences between evaluators in PCL-R Facet 3 scores, $F(2,352) = 10.81$, $p < .001$, $\eta_p^2 = .06$. Specifically, Evaluator 3 assigned significantly lower PCL-R Facet 3 scores than Evaluator 1, $t(208) = -3.22$, $p = .001$, $d = -.47$, and Evaluator 2, $t(281) = -4.26$, $p < .001$, $d = -.51$).

Table 7

PCL-R Facet 3 Scores

	M	*SD*
Total Sample (*N* = 359)	4.53	2.45
Evaluator 1 (*n* = 72)	5.07	2.60
Evaluator 2 (*n* = 145)	4.99	1.85
Evaluator 3 (*n* = 138)	3.81	2.73

Across the sample, PCL-R Facet 4 scores ranged from 0 to 10, with an average of 4.09 (*SD* = 2.73). Scores assigned by Evaluator 1 ranged from 1 to 10 (*M* = 5.38, *SD* = 2.60). Evaluator 2's PCL-R Facet 4 scores ranged from 0 to 9 (*M* = 4.04, *SD* = 2.51) and

Evaluator 3's scores ranged from 0 to 9 (*M* = 3.47, *SD* = 2.78). See Table 8. There were significant differences between evaluators in PCL-R Facet 4 scores, $F(2,359) = 12.30$, $p < .001$, $\eta_p^2 = .06$. Specifically, Evaluator 1 assigned significantly higher PCL-R Facet 4 scores than Evaluator 2, $t(222) = 3.68$, $p < .001$, $d = .53$, and Evaluator 3, $t(207) = 4.79$, $p < .001$, $d = .70$.

Table 8

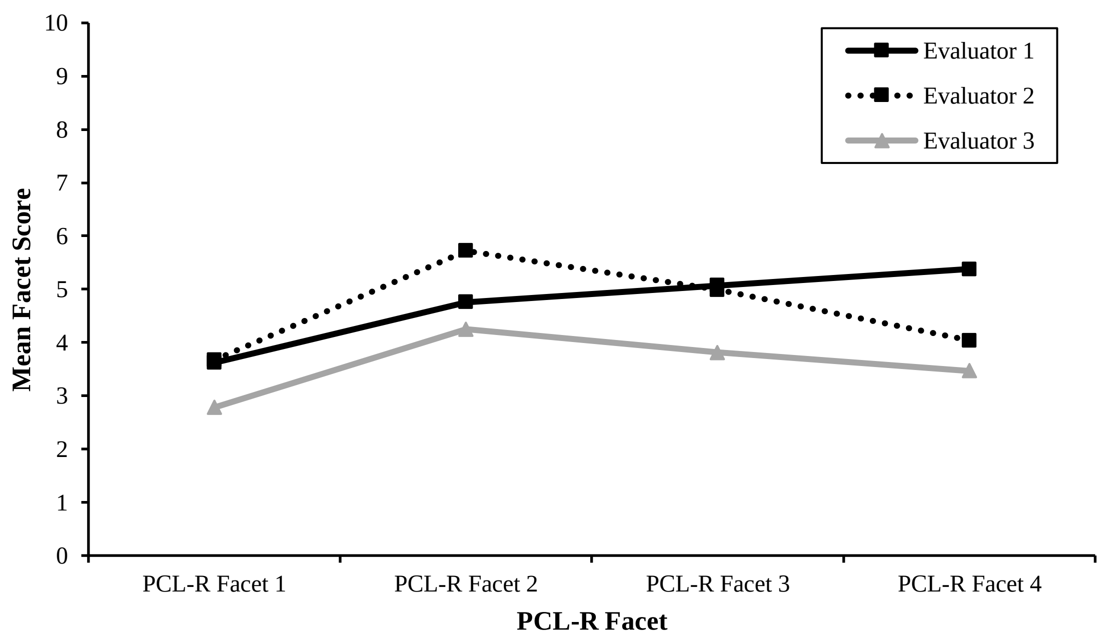
PCL-R Facet 4 Scores

	M	*SD*
Total Sample (*N* = 366)	4.09	2.73
Evaluator 1 (*n* = 71)	5.38	2.60
Evaluator 2 (*n* = 153)	4.04	2.51
Evaluator 3 (*n* = 138)	3.47	2.78

Trends in all four PCL-R facet scores across evaluators are depicted in Figure 1.

Figure 1

Trends in PCL-R Facet Scores by Evaluator



Overall Risk Opinion

Risk level frequencies are presented in Table 9. For the current sample, evaluators opined 21.1% (*n* = 83) of offenders to be at low risk for sexual re-offense; 29.8% (*n* = 117) to be at Low-to-Moderate risk; 10.9% (*n* = 43) at Moderate risk; 20.6% (*n* = 81) to be Moderate-to-High risk; 13.5% (*n* = 53) as High risk; and 0.5% (*n* = 2) to be at Very High risk. Fourteen (3.6%) reports did not include a risk level opinion.

Evaluator 1 opined 6.9% (*n* = 6) of cases to be at low risk; 25.3% (*n* = 22) to be Low-to-Moderate risk; 12.7% (*n* = 11) to be Moderate risk; 14.9% (*n* = 13) to be Moderate-to-High risk; and 32.2% (*n* = 28) to be at High risk. He did not opine any offenders to be at Very High risk. Evaluator 2 opined 26.6% (*n* = 42 of cases to be at low risk; 36.7% (*n* = 58) to be Low-to-Moderate risk; 3.2% (*n* = 5) to be Moderate risk; 24.7% (*n* = 39) to be Moderate-to-High risk; 6.3% (*n* = 10) to be at High risk; and 1.3% (*n* = 2) to be Very High risk. Evaluator 3 opined 23.7% (*n* = 33) of cases to be at low risk; 26.6% (*n* = 37) to be Low-to-Moderate risk; 18.7% (*n* = 26) to be Moderate risk; 18.7% (*n* = 26) to be Moderate-to-High risk; and 10.1% (*n* = 14) to be at High risk. He did not opine any offenders to be at Very High risk.

Table 9

Risk Level Opinion

	Risk Category	*N*	%
Total Sample	Low	83	21.1
	Low-to-Moderate	117	29.8
	Moderate	43	10.9
	Moderate-to-High	81	20.6
	High	54	13.5
	Very High	2	0.5
Evaluator 1	Low	6	6.9

(continued)

	Risk Category	*N*	%
	Low-to-Moderate	22	25.3
	Moderate	11	12.7
	Moderate-to-High	13	14.9
	High	28	32.2
	Very High	0	0.0
Evaluator 2	Low	42	26.6
	Low-to-Moderate	58	36.7
	Moderate	5	3.2
	Moderate-to-High	39	24.7
	High	10	6.3
	Very High	2	1.3
Evaluator 3	Low	33	23.7
	Low-to-Moderate	37	26.6
	Moderate	26	18.7
	Moderate-to-High	26	18.7
	High	14	10.1
	Very High	0	0.0

In analyses, risk level was coded as follows: Low = 1, Low-to-Moderate = 2, Moderate = 3, Moderate-to-High = 4, High = 5, and Very High = 6. The average risk level assigned regardless of evaluator was 2.76 (*SD* = 1.40; see Table 10). Average risk level assigned did differ between evaluators, $F(2,369) = 13.33$, $p < .001$, $\eta_p^2 = .07$. Specifically, independent samples t-tests revealed that evaluees on Evaluator 1's caseload ($M = 3.44$, $SD = 1.40$) were opined to be at higher levels of risk than those of Evaluator 2 ($M = 2.51$, $SD = 1.36$), $t(234) = 4.93$, $p < .001$, $d = .68$, or 3 ($M = 2.64$, $SD = 1.31$), $t(214) = 4.20$, $p < .001$, $d = .60$.

Table 10

Risk Level Opinion as Continuous Variable

	M	*SD*
Total Sample	2.76	1.40
Evaluator 1	3.44	1.40
Evaluator 2	2.51	1.36
Evaluator 3	2.64	1.31

Overall Behavioral Abnormality Opinion

Across the sample, evaluators opined 45.0% (*n* =177) of individuals to have a qualifying BA, while they opined 52.7% (*n* = 207) did not. Nine (2.3%) reports did not include a BA opinion. Evaluator 1 opined BA to be present in 48.3% (*n* = 42) of cases and absent in 46% (*n* = 40) of cases. Evaluator 2 opined present BA in 39.2% (*n* = 62) of cases and absent in 59.5% (*n* = 94) of cases, while Evaluator 3 assigned BA as present in 48.2% (*n* = 67) of cases and absent in 51.8% (*n* = 72) of cases. See Table 11. Behavioral abnormality opinion did not differ significantly across evaluators, $X^2(2, N = 377) = 3.57$, $p = .17$, Cramer's $V = .10$.

Table 11

Behavioral Abnormality Opinion

	Opinion	*n*	%
Total Sample	Yes	177	45.0
	No	207	52.7
Evaluator 1	Yes	42	48.3
	No	40	46.0
Evaluator 2	Yes	62	39.2
	No	94	59.5

(continued)

	Opinion	*n*	%
Evaluator 3	Yes	67	48.2
	No	72	51.8

Overall Association between Measure Scores and Risk Opinion

Across the entire sample, significant correlations were found between the various measure scores and ultimate opinion of risk level. All associations were found to be significant, though the strength of the relationship varied across the measures. The overall relationship between risk opinion and BA opinion was also significant. See Table 11.

Table 12

Correlation between Risk Opinion and Measure Scores/Behavioral Abnormality Across Sample

Measure	*M*	*SD*	*r*	*p*
Static-99R	2.67	2.24	.711	<.001***
PCL-R Total	18.73	7.10	.369	<.001***
PCL-R Facet 1	3.34	2.31	.128	.02*
PCL-R Facet 2	4.98	2.38	.193	<.001***
PCL-R Facet 3	4.53	2.45	.218	<.001***
PCL-R Facet 4	4.09	2.73	.384	<.001***
Behavioral Abnormality	-	-	.491	<.001***

Note. ****p* ≤ .001. ***p* ≤ .01. **p* ≤ .05.

Overall Association between Measure Scores and Behavioral Abnormality

Similarly, significant associations were found between each measure score and opinion regarding the presence or absence of BA. The strengths of these relationships tended to fall in the moderate range. See Table 13.

Table 13

Correlation between Measure Scores and Behavioral Abnormality Opinion Across Sample

Measure	*M*	*SD*	*r*	*p*
Static-99R	2.67	2.24	.372	<.001***
PCL-R Total	18.73	7.10	.391	<.001***
PCL-R Facet 1	3.34	2.31	.231	<.001***
PCL-R Facet 2	4.98	2.38	.301	<.001***
PCL-R Facet 3	4.53	2.45	.306	<.001***
PCL-R Facet 4	4.09	2.73	.222	<.001***

Note. ***$p \leq .001$. **$p \leq .01$. *$p \leq .05$.

Risk Level

To assess the potential moderating effect of evaluator on the relationship between actuarial measure scores and opinion of risk level, six sets of hierarchical multiple regression models were examined with risk level opinion as the dependent variable. The risk measure score (centered) was entered in the first model. The second model added two dummy coded variables to represent the three evaluators. The third model added two interaction terms to allow for the examination of a potential moderating effect. Regression statistics are reported in Table 14 and intercorrelations between the multiple regression variables are reported in Tables 15-20.

Static-99R Scores and Risk

With respect to the relationship between Static-99R scores and evaluators' ultimate opinion of risk level, the hierarchical regression revealed that at Step one, Static-99R scores contributed significantly to the regression model, $F(1,367) = 375.51$, $p < .001$, and accounted for 50.55% of the variation in risk level. Introducing Evaluator variables explained an additional 3.5% of the variation in risk level and this change in R^2 was significant, $F(2,365) = 13.78$, $p < .001$. Finally, adding the interaction terms between Static-99R score and Evaluator explained an additional 1.0% of the variation in risk level. This change in R^2 was again significant, $F(2,363) = 4.08$, $p=.02$.

In the final model, both of the interaction terms were statistically significant. The regression coefficients for these effects compare the effect for one evaluator (i.e., the evaluator coded 1 on the dummy variable) and the reference evaluator (the evaluator coded 0 on both dummy variables, which is Evaluator 1). Thus, the interaction term for evaluator 2 examines whether the association between Static-99R scores and risk opinions differs for Evaluator 2 and Evaluator 1. The positive sign of the regression coefficient ($\beta = .14$, $p = .02$) indicates that the association between Static-99R scores and risk was stronger for Evaluator 2 than Evaluator 1. Similarly, the interaction between Evaluator 3 and Static-99R score indicates that the association between Static-99R scores and risk was stronger for Evaluator 3 than Evaluator 1 ($\beta = .18$, $p = .005$).

In other words, the statistically significant interaction effects indicate that the association between Static-99R scores and risk opinions differed depending on which evaluator assigned the score.

Table 14

Linear Regression Results Examining the Potential Moderating Effect of Evaluator on the Association between Measure Scores and Evaluator Opinion of Risk Level

Measure	*b*	*SE*	95% CI of *b*	*t*	*p*	*sr*
Static-99R Total (Model R^2 = .51, p < .001) (Block R^2 change = .01, p = .02)						
Static-99R Total	.30	.05	[.20, .41]	5.68	<.001	.29***
Evaluator 2	-.63	.13	[-.89, -.37]	-4.73	<.001	-.24***
Evaluator 3	-.74	.14	[-1.01, -.48]	-5.48	<.001	-.28***
Static-99R x Evaluator 2	.14	.06	[.02, .31]	2.28	.02	.12*
Static-99R x Evaluator 3	.18	.06	[.06, .31]	2.83	.005	.15**

(continued)

Measure	*b*	*SE*	95% CI of *b*	*t*	*p*	*sr*
PCL-R Total (Model R^2 = .25, *p* = .008) (Block R^2 change = .02, *p* = .008)						
PCL-R Total	.12	.02	[.09, .16]	7.05	<.001	.35***
Evaluator 2	-1.02	.18	[-1.37, -.86]	-5.81	<.001	-.30***
Evaluator 3	-.57	.18	[-.92, -.22]	-3.20	.002	-.17**
PCL-R Total x Eval 2	-.06	.03	[-.11, -.01]	-2.39	.02	-.13*
PCL-R Total x Eval 3	-.07	.02	[-.11, -.02]	-3.01	.003	-.16**
PCL-R Facet 1 (Model R^2 = .11, *p* = .14) (Block R^2 change = .01, *p* = .14)						
PCL-R Facet 1	.17	.06	[.05, .29]	2.88	.004	.15**
Evaluator 2	-1.02	.19	[-1.40, -.64]	-5.27	<.001	-.27***
Evaluator 3	-.81	.20	[-1.20, -.43]	-4.14	<.001	-.22***
PCL-R Facet 1 x Eval 2	-.13	.09	[-.30, .05]	-1.41	.16	-.08
PCL-R Facet 1 x Eval 3	-.14	.07	[-.29, .00]	-1.95	.05	-.11*
PCL-R Facet 2 (Model R^2 = .17, *p* = .001) (Block R^2 change = .03, *p* = .001)						
PCL-R Facet 2	.30	.05	[.19, .40]	5.63	<.001	.29***
Evaluator 2	-1.14	.19	[-1.52, -.76]	-5.89	<.001	-.30***
Evaluator 3	-.89	.19	[-1.26, -.52]	-4.71	<.001	-.25***
PCL-R Facet 2 x Eval 2	-.28	.10	[-.47, -.09]	-2.91	.004	-.16**
PCL-R Facet 2 x Eval 3	-.22	.07	[-.35, -.09]	-3.34	.001	-.18***
PCL-R Facet 3 (Model R^2 = .17, *p* < .001) (Block R^2 change = .04, *p* < .001)						
PCL-R Facet 3	.33	.06	[.21, .45]	5.52	<.001	.29
Evaluator 2	-.99	.19	[-1.36, -.61]	-5.20	<.001	-.27***
Evaluator 3	-.71	.19	[-1.09, -.33]	-3.71	<.001	-.20***
PCL-R Facet 3 x Eval 2	-.21	.08	[-.37, -.04]	-2.49	.01	-.13**
PCL-R Facet 3 x Eval 3	-.30	.07	[-.44, -.16]	-4.20	<.001	-.22***
PCL-R Facet 4 (Model R^2 = .20, *p* = .29) (Block R^2 change = .01, *p* = .29)						
PCL-R Facet 4	.251	.06	[.14, .36]	4.38	<.001	.23***
Evaluator 2	-.73	.19	[-1.11, -.35]	-3.78	<.001	-.20***
Evaluator 3	-.49	.20	[-.87, -.10]	-2.46	.01	-.13**
PCL-R Facet 4 x Eval 2	-.06	.07	[-.20, .07]	-.91	.36	-.05
PCL-R Facet 4 x Eval 3	-.11	.07	[-.24, .03]	-1.56	.12	-.08

Note. *sr* = semipartial correlation. ***$p \leq .001$. **$p \leq .01$. *$p \leq .05$.

Specifically, there was a moderate correlation between Static-99R scores and risk level for Evaluator 1, r = .42 (p<.001), but there was a strong correlation between Static-99R scores and risk level for both Evaluator 2, r = .75 (p<.001) and Evaluator 3, r = .84 (p<.001). In other words, the association between Static-99R scores and risk opinions was stronger for Evaluators 2 and 3 than for Evaluator 1. See Table 15.

Table 15

Correlation between Static-99R Total Scores and Evaluator Risk Opinion by Evaluator

Evaluator	Variable	*M*	*SD*	Static-99R Total	Risk Level	*p*
Evaluator 1						
	Static-99R Total	3.09	2.04			
	Risk Level	3.44	1.40	.422	1	<.001***
Evaluator 2						
	Static-99R Total	2.26	2.27			
	Risk Level	2.51	1.36	.749	1	<.001***
Evaluator 3						
	Static-99R Total	2.86	2.28			
	Risk Level	2.64	1.31	.837	1	<.001***

Note. ***$p \leq .001$. **$p \leq .01$. *$p \leq .05$.

PCL-R Total Scores and Risk

With respect to the relationship between PCL-R Total scores and evaluators' opinion of risk level, the hierarchical regression revealed that at Step one, PCL-R Total scores contributed significantly to the regression model, $F(1,354) = 55.69$, $p < .001$, and accounted for 13.61% of the variation in risk level. Introducing Evaluator variables explained an additional 9% of the variation in risk level and this change in R^2 was significant, $F(2,352) = 20.47$, $p < .001$. Finally, adding the interaction terms between PCL-R Total score and evaluator explained an additional 2.1% of the variation in risk level. This change in R^2 was again significant, $F(2,350) = 4.96$, p=.008 (see Table 14).

Once again, both of the interaction terms were statistically significant. In this instance, the regression coefficients for the interaction terms were negative, indicating the association between PCL-R Total scores and risk was smaller for Evaluator 2 than for Evaluator 1 (β = -.06, p = .02). Similarly, the association was smaller for Evaluator 3 than for Evaluator 1 (β = -.07, p = .003). These statistically significant interaction effects again indicate the association between PCL-R scores and risk opinion differed depending on which evaluator assigned the score.

Specifically, there was a small correlation between PCL-R Total scores and risk level for Evaluator 2, $r(142) = .26$, p=.002 (see Table 16). There was a small-to-moderate correlation between the variables for Evaluator 3, $r(135) = .31$, $p < .001$., and a strong correlation between PCL-R Total scores and risk level for Evaluator 1, $r(70) = .73$, $p < .001$. In other words, the association between PCL-R Total score and risk opinion is a strong indicator of Evaluator's 1 assignment of risk level, whereas PCL-R Total score is at most a moderate indicator of assignment of risk for Evaluators 2 and 3. See Table 16.

Table 16

Correlation between PCL-R Total Scores and Evaluator Risk Opinion by Evaluator

Evaluator	Variable	*M*	*SD*	PCL-R Total	Risk Level	*p*
Evaluator 1						
	PCL-R Total	19.81	8.05			
	Risk Level	3.44	1.40	.732	1	<.001***
Evaluator 2						
	PCL-R Total	20.52	5.53			
	Risk Level	2.51	1.36	.262	1	.002***
Evaluator 3						
	PCL-R Total	16.18	7.33			
	Risk Level	2.64	1.31	.309	1	<.001***

Note. ***$p \leq .001$. **$p \leq .01$. *$p \leq .05$.

PCL-R Facet 1 Scores and Risk

With respect to the relationship between PCL-R Facet 1 scores and evaluators' opinion of risk level, the hierarchical regression revealed that at Step one, PCL-R Facet 1 scores contributed significantly to the regression model, $F(1,348) = 5.81$, $p = .02$, and accounted for 1.64% of the variation in risk level. Introducing evaluator variables explained an additional 8.3% of the variation in risk level and this change in R^2 was significant, $F(2,346) = 15.86$, $p < .001$. Finally, adding the interaction between PCL-R Facet 1 score and evaluator explained an additional 1.0% of the variation in risk level. This change in R^2 was not significant, $F(2,344) = 2.00$, $p = .14$ (see Table 14).

In this case, only one interaction term, representing the interaction of the variables for Evaluator 3, was statistically significant. The regression coefficient for the interaction term was negative, indicating the association between PCL-R Facet 1 scores and risk was smaller for Evaluator 3 than for Evaluator 1 ($\beta = -.14$, $p = .05$). There was no significant difference in the association between PCL-R Facet 1 scores and risk between Evaluator 1 and Evaluator 2 ($\beta = -.13$, $p = .16$). The variation in statistically significant interaction effects indicates that the association between PCL-R Facet 1 scores and risk opinions differed between some evaluators but not others.

Specifically, the relationship between PCL-R Facet 1 scores and risk was significant only for Evaluator 1, $r = .40$ (p=.001) (see Table 17). The relationship between PCL-R Facet 1 score and risk level was not significant for Evaluator 2, $r = .06$ ($p = .52$) or Evaluator 3, $r = .05$ ($p = .56$). In other words, PCL-R Facet 1 score is a moderate indicator of Evaluator 1's assignment of risk level, whereas PCL-R Facet 1 score does not significantly relate to opinion of risk for Evaluators 2 or 3. See Table 17.

Table 17

Correlation between PCL-R Facet 1 Scores and Evaluator Risk Opinion by Evaluator

Evaluator	Variable	*M*	*SD*	PCL-R Facet 1	Risk Level	*p*
Evaluator 1						
	PCL-R Facet 1	3.62	2.70			
	Risk Level	3.44	1.40	.398	1	.001***
Evaluator 2						
	PCL-R Facet 1	3.67	1.63			
	Risk Level	2.51	1.36	.055	1	.52
Evaluator 3						
	PCL-R Facet 1	2.78	2.58			
	Risk Level	2.64	1.31	.051	1	.56

Note. ***$p \leq .001$. **$p \leq .01$. *$p \leq .05$.

PCL-R Facet 2 Scores and Risk

With respect to the relationship between PCL-R Facet 2 scores and evaluators' opinion of risk level the hierarchical regression revealed that at Step one, PCL-R Facet 2 scores contributed significantly to the regression model, $F(1,350) = 13.58$, $p < .001$, and accounted for 3.72% of the variation in risk level. Introducing Evaluator variables explained an additional 10% of the variation in risk level and this change in R^2 was significant, $F(2,348) = 20.10$, $p <.001$. Finally, adding the interaction between PCL-R Facet 2 score and Evaluator explained an additional 3.3% of the variation in risk level. This change in R^2 was again significant, $F(2,346) = 6.85$, $p=.001$ (see Table 14).

Here, both of the interaction terms were statistically significant. The regression coefficients for the interaction terms were again negative, indicating that the association between PCL-R Facet 2 scores and risk was smaller for Evaluator 2 than for Evaluator 1 ($\beta = -.28$, $p = .004$). Similarly, the association was smaller for Evaluator 3 than Evaluator 1 ($\beta = -.22$, $p = .001$). These results indicate once again that the association between

PCL-R Facet 2 scores and risk opinions differed depending on which evaluator assigned the score.

Specifically, the relationship was significant only for Evaluator 1, *r* = .69 (*p* < .001) (see Table 18), for whom a strong positive correlation was found. The relationship between PCL-R Facet 2 score and risk level was positive, albeit non-significant, for both Evaluator 2, *r* = .02 (*p*=.81) and Evaluator 3, *r* = .16 (*p* = .07). In other words, the association between PCL-R Facet 2 score and risk opinion was strong for Evaluator 1's assignment of risk level, yet PCL-R Facet 2 score is not a significant indicator of risk for Evaluators 2 or 3. See Table 18.

Table 18

Correlation between PCL-R Facet 2 Scores and Evaluator Risk Opinion by Evaluator

Evaluator	Variable	*M*	*SD*	PCL-R Facet 2	Risk Level	*p*
Evaluator 1						
	PCL-R Facet 2	4.76	2.85			
	Risk Level	3.44	1.40	.688	1	<.001***
Evaluator 2						
	PCL-R Facet 2	5.72	1.33			
	Risk Level	2.51	1.36	.021	1	.81
Evaluator 3						
	PCL-R Facet 2	4.25	2.72			
	Risk Level	2.64	1.31	.155	1	.07

Note. ****p* ≤ .001. ***p* ≤ .01. **p* ≤ .05.

PCL-R Facet 3 Scores and Risk

With respect to the relationship between PCL-R Facet 3 scores and evaluators' opinion of risk level the hierarchical regression revealed that at Step one, Facet 3 scores contributed significantly to the regression model, $F(1,347) = 17.28$, $p < .001$, and accounted for 4.75% of the variation in risk level. Introducing evaluator variables explained an additional 8.4% of the variation in risk level and this change in R^2 was significant, $F(2,345) = 16.61$, $p <.001$. Finally, adding the interaction between PCL-R Facet 3 score and Evaluator explained an additional 4.2% of the variation in risk level. This change in R^2 was again significant, $F(2,343) = 8.80$, $p < .001$ (see Table 14).

Both of the interaction terms were statistically significant. In this instance, the regression coefficients for the interaction terms were both negative, indicating once again that the association between PCL-R Facet 3 scores and risk was smaller for Evaluator 2 than for Evaluator 1 ($\beta = -.21$, $p = .01$). Similarly, the association was smaller for Evaluator 3 than for Evaluator 1 ($\beta = -.30$, $p < .001$). Statistical significance found in the interaction effects show that the association between PCL-R Facet 3 scores and risk opinions differed depending on which evaluator assigned the score.

Specifically there was a moderate-to-strong correlation between PCL-R Facet 3 and risk level for Evaluator 1, $r = .58$ ($p < .001$) (see Table 19), but the correlation between the variables for Evaluator 2 was small, $r(142) = .17$, $p = .04$, and there was no significant correlation between PCL-R Facet3 scores and risk level for Evaluator 3, $r = .06$ ($p = .50$). In other words, Evaluator 1's risk level opinion is more strongly related to PCL-R Facet 3 than that of Evaluators 2 or 3. Risk level opinion is moderately-to-strongly associated with PCL-R Facet 3 score for Evaluator 1, weakly associated for Evaluator 2, and not significantly associated Evaluator 3. See Table 19.

Table 19

Correlation between PCL-R Facet 3 Total Scores and Evaluator Risk Opinion by Evaluator

Evaluator	Variable	*M*	*SD*	PCL-R Facet 3	Risk Level	*p*
Evaluator 1						
	PCL-R Facet 3	5.07	2.60			
	Risk Level	3.44	1.40	.583	1	<.001***
Evaluator 2						
	PCL-R Facet 3	4.99	1.85			
	Risk Level	2.51	1.36	.171	1	.04*
Evaluator 3						
	PCL-R Facet 3	3.81	2.73			
	Risk Level	2.64	1.31	.058	1	.503

Note. ***$p \leq .001$. **$p \leq .01$. *$p \leq .05$.

PCL-R Facet 4 Scores and Risk

With respect to the relationship between PCL-R Facet 4 scores and evaluators' opinion of risk level the hierarchical regression revealed that at Step one, PCL-R Facet 4 scores contributed significantly to the regression model, $F(1,356) = 61.52$, $p < .001$, and accounted for 14.75% of the variation in risk level. Introducing Evaluator variables explained an additional 4.5% of the variation in risk level and this change in R^2 was significant, $F(2,354) = 9.77$, $p <.001$. Finally, adding the interaction between PCL-R Facet 4 score and Evaluator explained an additional 0.6% of the variation in risk level, though this change in R^2 was not significant, $F(2,352) = 1.23$, $p = .29$ (see Table 14).

This time, neither interaction term was statistically significant. The association between PCL-R Facet 4 scores and risk did not differ between Evaluator 2 and Evaluator 1 ($\beta = -.06$, $p = .36$) nor between Evaluator 3 and Evaluator 1 ($\beta = -.11$, $p = .12$). Thus, the lack of significant interaction effects indicates that the association between PCL-R Facet 4 scores and risk opinions was similar across evaluators.

Specifically, there were moderate correlations between PCL-R Facet 4 scores and risk for Evaluator 1, *r*(68) = .45, *p*<.001, and for Evaluator 2, *r*(150) = .35, *p*<.001. There was a small-to-moderate correlation between PCL-R Facet 4 and risk for Evaluator 3, *r*(134) = .29, *p*=.001. PCL-R Facet 4 scores similarly predicted risk opinion across evaluators. See Table 20.

Table 20

Correlation between PCL-R Facet 4 Scores and Evaluator Risk Opinion by Evaluator

Evaluator	Variable	*M*	*SD*	PCL-R Facet 4	Risk Level	*p*
Evaluator 1						
	PCL-R Facet 4	5.38	2.60			
	Risk Level	3.44	1.40	.450	1	<.001***
Evaluator 2						
	PCL-R Facet 4	4.04	2.51			
	Risk Level	2.51	1.36	.346	1	<.001***
Evaluator 3						
	PCL-R Facet 4	3.47	2.78			
	Risk Level	2.64	1.31	.294	1	.001***

Note. ***$p \leq .001$. **$p \leq .01$. *$p \leq .05$.

Behavioral Abnormality

Logistic regression analyses were conducted to investigate whether there are relationships between the various measure scores and an evaluator's opinion of the presence or absence of BA, per evaluator. Six sets of stepwise logistical regression models were examined with BA opinion as the dependent variable. The risk measure score (centered) was entered in the first model. As above, the second model added two dummy coded variables to represent the three evaluators. The third model added two interaction terms to allow for the examination of a potential moderating effect. Regression statistics are reported in Table 21 and intercorrelations between the multiple regression variables are reported in Tables 22-27.

Static-99R Scores and Behavioral Abnormality

With respect to the relationship between Static-99R scores and evaluators' opinion regarding the presence or absence of BA, the hierarchical regression revealed that at Step one, Static-99R scores contributed significantly to the regression model. The unstandardized Beta weight for the Constant, *B* = (-.09), *SE* = .24, *Wald* = .13, *p*=.72, was significantly different from the unstandardized Beta weight for Static-99R Total score, *B* = (.33), *SE* =.12, *Wald* = 6.99, *p*=.008. The estimated odds ratio favored an increase of nearly 38% [*Exp (B)* = 1.38, 95% CI (1.09, 1.76)] in the chances of BA opined as present for every one unit increase in Static-99R Total score.

Introducing evaluator variables did not contribute to the regression model. Considering Evaluator 2 as opposed to Evaluator 1 did not uniquely contribute to the overall model [*B* = (-.26), *SE* = .30, *Wald* = .75, *p*=.39, *Exp (B)* = .77, 95% CI (.43, 1.39)], nor did Evaluator 3 as opposed to Evaluator 1, [*B* = (-.03), *SE* = .30, *Wald* = .01, *p*=.91, *Exp (B)* = .97, 95% CI (.54, 1.74)].

Similarly, the addition of the interaction terms also did not contribute significantly to the model. The unstandardized Beta weight for the Constant, *B* = (-.09), *SE* = .24, *Wald* = .13, *p*=.72, did not differ significantly from the unstandardized Beta weight for the interaction between Static-99R Total score and Evaluator 2 as opposed to Evaluator 1, *B* = (.16), *SE* =.16, *Wald* = .99, *p*=.32, nor from the unstandardized Beta weight for the interaction between Static-99R Total score and Evaluator 3 as opposed to Evaluator 1, *B* = (-.01), *SE* =.15, *Wald* = .00, *p*=.96. The odds ratio indicated a nonsignificant increase of 17% [*Exp(B)* = 1.17, 95% CI (.86, 1.60)] in the chances of BA opined as present as Static-99R Total scores increased for Evaluator 2 as opposed to Evaluator 1. For

Evaluator 3 as opposed to Evaluator 1, the odds ratio indicated a nonsignificant decrease of 1% [*Exp(B)* = .99, 95% CI (.74, 1.33) in the chances of BA opined as present as Static-99R Total score increased. See Table 21.

In other words, the absence of statistically significant interaction effects indicates that the association between Static-99R scores and BA did not differ based on which evaluator made the opinion.

Table 21

Logistical Regression Results Examining the Potential Moderating Effect of Evaluator on the Association between Measure Scores and Evaluator Opinion of Presence of Behavioral Abnormality

Measure	*b*	*SE*	95% CI of *b*	*Wald*	*p*	*Exp(B)*
Behavioral Abnormality (Model R^2 = .15, p < .001[1]) (Block R^2 = .141, p < .001)						
Constant	-.09	.24		.13	.72	.92
Static-99R Total	.33	.12	[1.09, 1.76]	6.99	.008	1.38
Evaluator 2	-.26	.30	[.43, 1.39]	.75	.39	.77
Evaluator 3	-.03	.30	[.54, 1.74]	.01	.91	.97
Static-99R Total x Eval 2	.16	.16	[.86, 1.60]	.99	.32	1.17
Static-99R Total x Eval 3	-.01	.15	[.74, 1.33]	.00	.96	.99
Behavioral Abnormality (Model R^2 = .26, p < .001) (Block R^2 = .188, p < .001)						
Constant	.28	.37		.67	.42	1.32
PCL-R Total	.30	.06	[1.19, 1.53]	22.44	.00	1.35
Evaluator 2	-.66	.38	[.24, 1.09]	3.04	.08	.52
Evaluator 3	.25	.41	[.57, 2.88]	.36	.55	1.28
PCL-R Total x Eval 2	-.30	.07	[.65, .86]	17.62	.00	.75
PCL-R Total x Eva 3	-.08	.07	[.80, 1.07]	012	.29	.93

(continued)

[1] Cox & Snell

Measure	*b*	*SE*	95% CI of *b*	*Wald*	*p*	*Exp(B)*
Behavioral Abnormality (Model R^2 = .10, p < .001) (Block R^2 = .069, p < .001)						
Constant	.19	.27		.48	.49	1.21
PCL-R Facet 1	.43	.12	[1.22, 1.95]	13.17	.00	1.54
Evaluator 2	-.53	.32	[.32, 1.11]	2.66	.10	.59
Evaluator 3	-.11	.33	[.47, 1.70]	.12	.73	.89
PCL-R Facet 1 x Eval 2	-.52	.16	[.43, .84]	10.86	.001	.59
PCL-R Facet 1 x Eval 3	-.16	.14	[.65, 1.13]	1.22	.27	.86
Behavioral Abnormality (Model R^2 = .16, p < .001) (Block R^2 = .122, p < .001)						
Constant	.48	.35		1.83	.18	1.61
PCL-R Facet 2	.78	.17	[1.57, 3.01]	22.07	.00	2.18
Evaluator 2	-.89	.40	[.19, .91]	4.87	.03	.41
Evaluator 3	-.40	.40	[.31, 1.46]	1.00	.32	.67
PCL-R Facet 2 x Eval 2	-.71	.21	[.33, .74]	11.52	.001	.49
PCL-R Facet 2 x Eval 3	-.53	.18	[.41, .83]	8.82	.003	.59
Behavioral Abnormality (Model R^2 = .15, p < .001) (Block R^2 = .117, p < .001)						
Constant	.04	.29		.02	.89	1.04
PCL-R Facet 3	.63	.16	[1.37, 2.57]	15.18	.00	1.88
Evaluator 2	-.49	.34	[.31, 1.20]	2.03	.16	.61
Evaluator 3	.10	.35	[.56, 2.21]	.09	.77	1.11
PCL-R Facet 3 x Eval 2	-.56	.19	[.40, .82]	9.06	.003	.57
PCL-R Facet 3 x Eval 3	-.29	.18	[.53, 1.06]	2.65	.10	.75
Behavioral Abnormality (Model R^2 = .08, p < .001) (Block R^2 = .062, p < .001)						
Constant	.00	.27		.00	1.00	1.00
PCL-R Facet 4	.26	.10	[1.07, 1.59]	6.72	.01	1.30
Evaluator 2	-.45	.32	[.35, 1.19]	1.98	.16	.64
Evaluator 3	.05	.33	[.56, 1.99]	.03	.87	1.05
PCL-R Facet 4 x Eval 2	-.21	.12	[.64, 1.03]	3.03	.08	.81
PCL-R Facet 4 x Eval 3	-.03	.12	[.77, 1.23]	.06	.81	.97

Specifically, there were small-to-moderate correlations between Static-99R scores and BA for Evaluator 1, $r(81) = .30$, $p < .001$, and for Evaluator 3, $r(137) = .32$, $p < .001$. There was a moderate correlation between Static-99R scores and BA for Evaluator 2, $r(155) = .43$, $p < .001$. See Table 22.

Table 22

Correlation between Static-99R Total Scores and Behavioral Abnormality Opinion by Evaluator

Evaluator	Variable	*M*	*SD*	Static-99R	BA Opinion	*p*
Evaluator 1						
	Static-99R Total	3.09	2.04			
	BA Opinion	.51	.50	.302	1	.006**
Evaluator 2						
	Static-99R Total	2.26	2.27			
	BA Opinion	.40	.49	.429	1	<.001***
Evaluator 3						
	Static-99R Total	2.86	2.28			
	BA Opinion	.48	.50	.324	1	<.001***

Note. ***$p \leq .001$. **$p \leq .01$. *$p \leq .05$.

In other words, Static-99R scores were a moderate indicator of BA for Evaluator 2 and a weak-to-moderate predictor of BA for Evaluators 1 and 3.

PCL-R Total Scores and Behavioral Abnormality

With respect to PCL-R Total score, PCL-R Total score was found to contribute to the model. The unstandardized Beta weight for the Constant, *B* = (.28), *SE* = .34, *Wald* = .67, *p*=.41 differed significantly from the unstandardized Beta weight for PCL-R Total score, *B* = (.30), *SE* =.06, *Wald* = 22.44, *p*<.001. The estimated odds ratio favored an increase in opining BA as present of nearly 35% [*Exp (B)* = 1.35, 95% CI (1.19, 1.53)] for every one unit increase in PCL-R Total score.

Introducing evaluator variables did not significantly contribute to the model. Considering Evaluator 2 as opposed to Evaluator 1 did not uniquely contribute to the overall model [*B* = (-.66), *SE* = .38, *Wald* =3.04, *p*=.08, *Exp (B)* = .53, 95% CI (.24, 1.09)], nor did Evaluator 3 as opposed to Evaluator 1, [*B* = (.25), *SE* = .41, *Wald* = .36, *p*=.55, *Exp (B)* = 1.28, 95% CI (.57, 2.88)].

In the final model, the one of the interaction terms was statistically significant. In this instance, the regression coefficients for the interaction terms were negative, indicating that the association between PCL-R Total scores and BA was smaller for Evaluator 2 than for Evaluator 1, *B* = (-.30), *SE* =.07, *Wald* = 17.62, $p<.001$. The association between PCL-R Total scores and BA did not differ significantly between Evaluator 3 and Evaluator 1, *B* = (-.08), *SE* =.07, *Wald* = 1.12, $p=.29$. The odds ratio indicated a 25% decrease [*Exp(B)* = .75, 95% CI (.65, .86)] in the chances that behavioral abnormality is opined as present as PCL-R Total scores increased for Evaluator 2 as opposed to Evaluator 1. For Evaluator 3 as opposed to Evaluator 1, the odds ratio indicated a nonsignificant decrease of 7% [*Exp(B)* = .93, 95% CI (.80, 1.07) in the chances of BA opined as present as PCL-R Total score increased. The varied statistical significance in the interaction effects indicates that the association between PCL-R Total scores and BA differs between some evaluators but not others. See Table 21.

Specifically, there was a strong correlation between PCL-R Total scores and BA for Evaluator 1, $r(70) = .73$, $p < .001$, and a moderate-to-strong correlation for Evaluator 3, $r(138) = .58$, $p < .001$. However, the relationship between PCL-R Total scores and BA for Evaluator 2, $r(142) = .01$, $p = .90$, was not significant. In other words, PCL-R Total score is a strong indicator of Evaluator 1's determination of BA and a moderate-to-strong indicator of Evaluator 3's opinion of BA, whereas PCL-R Total score is not a significant indicator of BA for Evaluator 2. See Table 23.

Table 23

Correlation between PCL-R Total Scores and Behavioral Abnormality Opinion by Evaluator

Evaluator	Variable	*M*	*SD*	PCL-R Total	BA Opinion	*p*
Evaluator 1						
	PCL-R Total	19.81	8.05			
	BA Opinion	.51	.50	.731	1	<.001***
Evaluator 2						
	PCL-R Total	20.52	5.53			
	BA Opinion	.40	.49	.011	1	.90
Evaluator 3						
	PCL-R Total	16.18	7.33			
	BA Opinion	.48	.50	.577	1	<.001***

Note. ****p* ≤ .001. ***p* ≤ .01. **p* ≤ .05.

PCL-R Facet 1 Scores and Behavioral Abnormality

With respect to PCL-R Facet 1 score, PCL-R Facet 1 score was found to contribute to the model. The unstandardized Beta weight for the Constant, *B* = (.19), *SE* = .27, *Wald* = .48, *p*=.49 was significantly different than the unstandardized Beta weight for PCL-R Facet 1 Total score, *B* = (.43), *SE* =.12, *Wald* = 13.17, *p*<.001. The estimated odds ratio demonstrated a 54% increase [*Exp (B)* = 1.54, 95% CI (1.22, 1.95)] in the presence of BA for every one unit increase in PCL-R Facet 1 score.

Introducing evaluator variables again did not contribute significantly to the model. Considering Evaluator 2 as opposed to Evaluator 1 did not uniquely contribute to the overall model [*B* = (-.53), *SE* = .32, *Wald* = 2.66, *p*=.10, *Exp (B)* = .59, 95% CI (.32, 1.11)], nor did Evaluator 3 as opposed to Evaluator 1, [*B* = (-.11), *SE* = .33, *Wald* = .11, *p*=.73, *Exp (B)* = .89, 95% CI (.47, 1.70)].

In the final model, the addition of one of the interaction terms was statistically significant. Once again, the regression coefficients for the interaction terms were negative, indicating that the association between PCL-R Facet 1 scores and BA was smaller for Evaluator 2 than for Evaluator 1, *B* = (-.52), *SE* =.16, *Wald* = 10.86, $p<.001$. The association between PCL-R Facet 1 scores and BA was not significantly different for Evaluator 3 than for Evaluator 1, *B* = (-.16), *SE* =.14, *Wald* = 1.22, $p=.27$. The odds ratio indicated a 41% decrease [*Exp(B)* = .59, 95% CI (.43, .81)] in the chances that BA was opined as present as PCL-R Facet 1 scores increased for Evaluator 2 as opposed to Evaluator 1. Once again, the varied statistical significance in the interaction effects indicates that the association between PCL-R Facet 1 scores and BA differs across some evaluators. See Table 21.

Specifically, there was a moderate positive correlation between PCL-R Facet 1 scores and BA for both Evaluator 1, $r(67) = .46$, $p<.001$, and Evaluator 3, $r(137) = .33$, $p<.001$. In contrast, for Evaluator 2, the relationship was negative, though not significant, $r(143) = -.07$, $p=.39$. In other words, PCL-R Facet 1 scores are a moderate indicator of Evaluator 1's opinion of BA as well as Evaluator 3's opinion, though PCL-R Facet 1 scores are not a significant indicator of BA for Evaluator 2. See Table 24.

Table 24

Correlation between PCL-R Facet 1 Scores and Behavioral Abnormality Opinion by Evaluator

Evaluator	Variable	*M*	*SD*	PCL-R Facet 1	BA Opinion	*p*
Evaluator 1						
	PCL-R Facet 1	3.62	2.70			
	BA Opinion	.51	.50	.463	1	<.001***
Evaluator 2						
	PCL-R Facet 1	3.67	1.63			
	BA Opinion	.40	332	-.072	1	.39
Evaluator 3						
	PCL-R Facet 1	2.78	2.58			
	BA Opinion	.48	.50	.332	1	<.001***

Note. ***$p \leq .001$. **$p \leq .01$. *$p \leq .05$.

PCL-R Facet 2 Scores and Behavioral Abnormality

PCL-R Facet 2 score was found to contribute to the model. The unstandardized Beta weight for the Constant, *B* = (.48), *SE* = .35, *Wald* = 1.83, *p*=.18, was different from the unstandardized Beta weight for PCL-R Facet 2 score, *B* = (.78), *SE* =.17, *Wald* = 22.07, *p*<.001. The estimated odds ratio showed an increase of nearly 118% [*Exp (B)* = 2.18, 95% CI (1.57, 3.01)] in the opinion of the presence of BA for every one unit increase in PCL-R Facet 2 score.

One evaluator variable contributed to the model, while the other did not significantly contribute. The evaluator 2 variable (Evaluator 2 as opposed to Evaluator 1) contributed to the overall model, [*B* = (-.88), *SE* = .40, *Wald* = 4.87, *p*=.03] with the odds ratio indicating a decrease of 59% [*Exp (B)* = .41, 95% CI (.19, .91)] in the chances of BA being opined as present when considering the effect of Evaluator 2 as opposed to Evaluator 1. However, the influence of Evaluator 3 as opposed to Evaluator 1 did not uniquely contribute to the model, [*B* = (-.40), *SE* = .40, *Wald* = 1.00, *p*=.32, *Exp (B)* = .67, 95% CI (.31, 1.46)].

In the final model, the interaction terms were statistically significant. Both regression coefficients were negative, again indicating that the association between PCL-R Facet 2 scores and BA was smaller for both Evaluator 2 than for Evaluator 1, *B* = (-.71), *SE* =.21, *Wald* = 11.52, *p*=.001, and for Evaluator 3 as opposed to Evaluator 1, *B* = (-.53), *SE* =.18, *Wald* = 8.82, *p*=.003. The odds ratio indicated a decrease of 51% [*Exp(B)* = .49, 95% CI (.33, .74)] in the occurrence of BA opined as present as PCL-R Facet 2 scores increased for Evaluator 2 as opposed to Evaluator 1. For Evaluator 3 as opposed to Evaluator 1, the odds ratio indicated a decrease of 41% [*Exp(B)* = .59, 95% CI (.41, .83) in the opinion of the presence of BA as PCL-R Facet 2 score increased. The statistically significant interaction effects indicate that the association between PCL-R Facet 2 scores and BA is different depending on evaluator. See Table 21.

Specifically, there was a strong correlation between PCL-R Facet 2 scores and BA for Evaluator 1, $r(67) = .71$, $p < .001$. There was a small-to-moderate correlation between PCL-R Facet 2 scores and BA opinion for Evaluator 3, $r(137) = .31$, $p<.001$. There was not a significant relationship between the variables for Evaluator 2, $r(145) = .05$, $p=.59$. In other words, PCL-R Facet 2 scores are a strong indication of BA opinion for Evaluator 1 and a small-to-moderate indicator of BA for Evaluator 3. However, PCL-R Facet 2 scores are not an indication of opined BA for Evaluator 2. See Table 25.

Table 25

Correlation between PCL-R Facet 2 Scores and Behavioral Abnormality Opinion by Evaluator

Evaluator	Variable	*M*	*SD*	PCL-R Facet 2	BA Opinion	*p*
Evaluator 1						
	PCL-R Facet 2	4.76	2.85			
	BA Opinion	.51	.50	.712	1	<.001***
Evaluator 2						
	PCL-R Facet 2	5.72	1.33			
	BA Opinion	.40	.49	.045	1	.59
Evaluator 3						
	PCL-R Facet 2	4.25	2.72			
	BA Opinion	.48	.50	.312	1	<.001***

Note. ***$p \leq .001$. **$p \leq .01$. *$p \leq .05$.

PCL-R Facet 3 Scores and Behavioral Abnormality

With respect to PCL-R Facet 3 score, PCL-R Facet 3 score was found to contribute to the prediction of BA opinion. The unstandardized Beta weight for the Constant, *B* = (.04), *SE* = .29, *Wald* = .02, *p*=.89 differed significantly from the unstandardized Beta weight for PCL-R Facet 3 score, *B* = (.63), *SE* =.16, *Wald* = 15.18, *p*<.001. The estimated odds ratio favored an increase of nearly 88% [*Exp (B)* = 1.88, 95% CI (1.37, 2.57)] in the presence of BA for every one unit increase in Facet 3 score.

Evaluator variables did not contribute to the model. Dummy variable Evaluator 2 did not uniquely contribute to the overall model [*B* = (-.49), *SE* = .34, *Wald* = 2.03, *p*=.16, *Exp (B)* = .61, 95% CI (.31, 1.20)], nor did dummy variable for Evaluator 3, [*B* = (.10), *SE* = .35, *Wald* = .09, *p*=.77, *Exp (B)* = 1.11, 95% CI (.56, 2.21)].

Once again, one of the interaction terms was statistically significant. Again, the regression coefficients for the interaction terms were negative. The association between PCL-R Facet 3 scores and BA was smaller for Evaluator 2 than it was for Evaluator 1, *B* = (-.56), *SE* =.19, *Wald* = 9.06, *p*=.003. There was not a significant difference in the association between PCL-R Facet 3 scores and BA for Evaluator 3 as opposed to Evaluator 1, *B* = (-.29), *SE* =.18, *Wald* = 2.65, *p*=.10. The odds ratio indicated a 43% decrease [*Exp(B)* = .57, 95% CI (.40, .82)] in the opinion of BA as present as PCL-R Facet 3 scores increased for Evaluator 2 as opposed to Evaluator 1. For Evaluator 3 as opposed to Evaluator 1, the odds ratio indicated a nonsignificant decrease of 25% [*Exp(B)* = .75, 95% CI (.53, 1.06) in the BA opined as present as PCL-R Facet 3 score increased. The varied statistical significance in the interaction effects shows that the association between PCL-R Facet 3 scores and behavioral abnormality opinion varies by evaluator. See Table 21.

Specifically, there was a moderate-to-strong correlation between PCL-R Facet 3 scores and BA for Evaluator 1, $r(67) = .59$, $p < .001$. There was a moderate correlation between PCL-R Facet 3 scores and BA for Evaluator 3, $r(137) = .40$, $p < .001$. The relationship between the variables was not significant for Evaluator 2, $r(142) = .06$, $p = .46$. Put another way, PCL-R Facet 3 score is a moderate-to-strong predictor of BA opinion for Evaluator 1 and a moderate indicator of BA for Evaluator 3. However, PCL-R Facet 3 is not an indicator of BA for Evaluator 2. See Table 26.

Table 26

Correlation between PCL-R Facet 3 Scores and Behavioral Abnormality Opinion by Evaluator

Evaluator	Variable	*M*	*SD*	PCL-R Facet 3	BA Opinion	*p*
Evaluator 1						
	PCL-R Facet 3	5.07	2.60			
	BA Opinion	.51	.50	.588	1	<.001***
Evaluator 2						
	PCL-R Facet 3	4.99	1.85			
	BA Opinion	.40	.49	.062	1	.46
Evaluator 3						
	PCL-R Facet 3	3.81	2.73			
	BA Opinion	.48	.50	.404	1	<.001***

Note. ***$p \leq .001$. **$p \leq .01$. *$p \leq .05$.

PCL-R Facet 4 Scores and Behavioral Abnormality

Finally, PCL-R Facet 4 score was found to contribute to the prediction of evaluators' opinion of the presence or absence of BA. The unstandardized Beta weight for the Constant, *B* = (.00), *SE* = .27, *Wald* = .00, *p*=1.00, differed significantly from the unstandardized Beta weight for PCL-R Facet 4 score, *B* = (.26), *SE* =.10, *Wald* = 6.72, *p*=.01. Per the estimated odds ratio, there was a 30% increase [*Exp (B)* = 1.30, 95% CI (1.07, 1.59)] in the opinion that BA was present for every one unit increase in PCL-R Facet 4 score.

Neither evaluator variable contributed to the model. Considering Evaluator 2 as opposed to Evaluator 1 did not uniquely contribute to the overall model [*B* = (-.45), *SE* = .32, *Wald* = 1.98, *p*=.16, *Exp (B)* = .64, 95% CI (.35, 1.19)], nor did Evaluator 3 as opposed to Evaluator 1, [*B* = (.05), *SE* = .33, *Wald* = .03, *p*=.87, *Exp (B)* = 1.05, 95% CI (.56, 1.99)].

Similarly, introducing interaction terms was not statistically significant. The association between PCL-R Facet 4 score and BA was not different for Evaluator 2 as opposed to Evaluator 1, *B* = (-.22), *SE* =.12, *Wald* = 3.03, *p*=.08, nor for Evaluator 3 as opposed to Evaluator 1, *B* = (-.03), *SE* =.12, *Wald* = .06, *p*=.81. The odds ratio indicated a nonsignificant decrease of 19% [*Exp(B)* = .81, 95% CI (.64, 1.03)] with regard to the BA opined as present as PCL-R Facet 4 scores increased for Evaluator 2 as opposed to Evaluator 1. For Evaluator 3 as opposed to Evaluator 1, the odds ratio indicated a nonsignificant decrease of 3% [*Exp(B)* = .97, 95% CI (.77, 1.23) in the chances of BA opined as present as PCL-R Facet 4 score increased. The lack of statistical significance in the interaction demonstrates that the association between PCL-R Facet 4 scores and BA is similar regardless of which evaluator offered the opinion. See Table 21.

A moderate correlation was found between PCL-R Facet 4 scores and BA for Evaluator 1, *r*(68) = .36, *p* = .002. There was a small-to-moderate correlation between PCL-R Facet 4 scores and BA for Evaluator 3, *r*(137) = .31, *p* < .001. There was no significant relationship between the variables for Evaluator 2, *r*(150) = .06, *p*=.44. Said differently, PCL-R Facet 4 score is a moderate indicator of Evaluator 1's BA opinion. PCL-R Facet 4 is a small-to-moderate indicator of BA for Evaluator 3. PCL-R Facet 4 is not an indicator of BA for Evaluator 2. See Table 27.

Table 27

Correlation between PCL-R Facet 4 Scores and Behavioral Abnormality Opinion by Evaluator

Evaluator	Variable	*M*	*SD*	PCL-R Facet 4	BA Opinion	*p*
Evaluator 1						
	PCL-R Facet 4	5.38	2.60			
	BA Opinion	.51	.50	.359	1	.002**
Evaluator 2						
	PCL-R Facet 4	4.04	2.51			
	BA Opinion	.40	.49	.063	1	.442
Evaluator 3						
	PCL-R Facet 4	3.47	2.78			
Evaluator	Variable	*M*	*SD*	PCL-R Facet 4	BA Opinion	*p*
	BA Opinion	.48	.50	.307	1	<.001***

Note. ***p ≤ .001. **p ≤ .01. *p ≤ .05.

CHAPTER V

Discussion

The current research demonstrated that evaluator effects influence ultimate opinions in SVP civil commitment evaluations. Specifically, the strength and, at times, direction, of the relations between Static-99R, PCL-R Total, and PCL-R Facets 1, 2, 3, and 4 scores, respectively, and opined risk level were dependent upon the evaluator conducting the evaluation. Similar moderating effects of evaluator were found for the relationships between measure scores and the opinion about presence or absence of BA.

Risk

Given that the Static-99R is comprised of objectively scored items (i.e., little variation in scoring accuracy is expected) and that high predictive validity makes this a valuable resource in decision-making when it comes to risk for future sexual violence, it was expected that evaluators would take Static-99R scores into account similarly in determining overall level of risk for future sexual violence. Alternatively, it was hypothesized that PCL-R Total scores would demonstrate varied predictive utility across evaluators with regard to risk, due to the indirect relationship between psychopathy and sexual recidivism. PCL-R Facets 1 and 2 were also expected to show differing prediction due to the subjective nature of items. On the other hand, it was supposed that PCL-R Facets 3 and 4 might predict risk opinion equally across evaluators, due to the more objective nature of the items. Also, Facet 4, in particular, was shown in previous research to have a stronger relationship with sexual recidivism (Hawes, Boccaccini, & Murrie, 2013; Walters, Knight, Grann & Dahle, 2008), adding support for this supposition.

Instead, Static-99R was found to variably predict risk level by evaluator. The relationship with PCL-R Facet 3 scores was also found to vary, also contrary to prediction. While PCL-R Total, Facet 1, and Facet 2 scores were also shown to have differing relationships based on evaluator, these associations were as predicted. PCL-R Facet 4 scores predicted equally across evaluators, as hypothesized.

Overall, Evaluator 1's risk opinions appeared to be more strongly associated with PCL-R scores than Static-99R scores. Evaluator 1's risk opinion was more strongly correlated than that of either Evaluator 2 or Evaluator 3 to nearly all PCL-R scores (Total, Facet 1, Facet 2, and Facet 3). In fact, for some facets, Evaluator 1's opinion was the only opinion correlated with the PCL scores, and only Evaluator 1's opinions were ever strongly correlated with PCL scores.

To the contrary, Evaluator 2 took Static-99R scores into account strongly when considering overall risk level. Evaluator 2's risk opinion was not particularly associated with the various PCL-R scores, save for the PCL-R Facet 4 score (moderate) and, to a small degree, the Facet 3 and Total scores.

Evaluator 3 also took Static-99R scores into large account in opining overall risk level. However, Evaluator 3's risk opinion was the least associated with PCL-R scores, with only a small-to-moderate association with PCL-R Facet 4 lending to a weak association with PCL-R Total.

Behavioral Abnormality

The Static-99R is a risk measure comprising objectively scored items pertaining to historical/static risk factors. As such, a direct connection between the Static-99R and BA consideration did not seem intuitive, and it was believed that any relationship that did exist might differ across evaluators. Alternatively, given that measurements of psychopathy represent examinations of maladaptive personality traits, and maladaptive personality traits lend to diagnoses of mental illness, it was believed that PCL-R scores might equally predict evaluators' opinions about the presence or absence of BA.

Surprisingly, all evaluators' opinions of BA were moderately correlated with Static-99R scores, rendering that hypothesis unsupported. Of note, the finding here corresponds with the overall relationship found in previous analyses of the current data assessing the relationship between Static-99R and risk level opinion (without moderating variables).

On the other hand, PCL-R scores did not perform as expected. Rather, there were significant differences in the relationships between measure scores and BA opinion across evaluators. Overall, Evaluator 1's opinion of the presence or absence of BA was more strongly correlated than that of either Evaluator 2 or Evaluator 3 to almost all PCL-R scores (Total, Facet 2, Facet 3, and Facet 4). Once again, only Evaluator 1's opinion regarding BA was ever strongly correlated with PCL-R scores (Total and Facet 2). Further, Evaluator 1's BA opinions appear to be more associated with the varied PCL-R scores than with Static-99R scores. Evaluator 2 took Static-99R scores into account and did so slightly more than Evaluators 1 or 3 when considering presence of BA. Evaluator 2's BA opinion was not significantly associated with any of the various PCL-R scores. Evaluator 3 also took Static-99R scores into account in opining BA, but the association between Evaluator 3's BA opinion and PCL-R scores varied. In general, Evaluator 3 took PCL-R scores into account to a moderate degree.

Factors Influencing Evaluators' Opinions

Ultimately, evaluators differed in the weight they gave to the measures of risk and personality both in considering risk level and presence of BA. In other words, in evaluating an individual's appropriateness for commitment as a sexually violent predator, evaluators do not take into account the same factors when determining risk level, or even when determining whether an individual has a qualifying BA.

Much is at stake in the decision of whether an offender meets criteria for SVP civil commitment. While reports indicate there is variability in how strongly or even whether expert testimony influences a jury's decision-making in SVP civil commitment trials (Krauss & Scurich, 2014; Boccaccini, et al., 2013; McCabe, Krauss, & Lieberman, 2010), a most recent report (Elwood, 2019) stated that courts agreed with expert opinion 67% of the time. While the exact nature of the agreement is unknown (i.e., causal or not), it is important that evaluators provide the trier of fact with the most accurate depiction of true, research-based risk estimates. As such, it is important we, as a field, continually assess and understand the factors which impact evaluators' opinions. However, few studies have examined how the scores from routinely used assessment measures relate to the ultimate opinions put forth to the court in sexually violent predator civil commitment proceedings. The purpose of the current study was to gain a better understanding of the factors which influence evaluators' opinions in the course of these evaluations, and whether they do so uniformly.

Present results demonstrate that the standardized measures of sexual violence risk and psychopathic personality traits are taken into account when considering whether an individual meets fundamental criteria for civil commitment under SVP statute in Texas.

However, evaluators are not taking them into account in similar ways or to similar degrees. This is in line with previous research asserting that evaluators utilize instruments differently in the course of evaluations of risk for sexual re-offense (Chevalier, Boccaccini, Murrie, and Varela, 2015; Miller & Maloney, 2013). But why might this be?

With regard to risk level opinion, it makes sense that Static-99R scores would be predictive of risk level finding. However, I was surprised to find variance in the strength of the relationship across evaluators. When examining the findings, it appears that difference may be due not to a lack of value on Static-99R utility but, rather, a difference in the value placed on other measures in the mix. Specifically, there was an overall reliance on all PCL-R scores for one evaluator, and slightly differing reliance by the other two evaluators.

When considering why the PCL-R would significantly predict a risk level opinion in the first place, a reason for this relationship could be within the language of the statute itself. The current sample was required to administer a measure of psychopathy as per state statute. Such a requirement may communicate to evaluators within this jurisdiction the expectation that psychopathy would, or perhaps should, be directly taken into account when considering an individual’s risk for future sexual violence. In terms of the variation in PCL-R score application in the current sample, perhaps variation in the interpretation of the statutory language was involved.

Another factor possibly contributing to differential utility across evaluators of the PCL-R scores was noted earlier. That is, previous literature has shown that, while psychopathy does function as a predictor of sexual recidivism, psychopathy alone may not be the strongest predictor of sexual recidivism (Hawes, Boccaccini, & Murrie, (2013).

As such, it makes sense that individual evaluators may place differing emphasis on psychopathy given the varied reports of its predictive value.

Moreover, it may be that some evaluators administer the PCL-R without any intention of utilizing it as a measure of risk. Boccaccini, Chevalier, Murrie, and Varela (2017) found that SVP evaluators in their study endorsed multiple reasons for administering the PCL-R, including as a measure of risk for future sexual offending, as a measure of risk for future violent offending, and as a measure of mental illness/abnormality. Thus, it is possible evaluators here, while required to assess for psychopathy, may define the function of such a measure differently.

Of course, PCL-R Facet 4 scores predicted risk level opinion equally across evaluators, despite the fact that the relationship between all other PCL-R facet scores and risk opinion was stronger for Evaluator 1 than for Evaluators 2 or 3. This finding is in line with previous research (Hawes, Boccaccini, & Murrie, 2013; Walters, Knight, Grann, & Dahle, 2008). Additionally, it makes intuitive sense when one considers that PCL-R Facet 4 assesses for history of antisocial behavior. It is not surprising, therefore, that the resulting score would be influential when considering risk for a future antisocial behavior (i.e., sexual offending).

With regard to BA opinion, due to previous analyses of the current data, it was not surprising that Static-99R score would so significantly relate to the BA opinion. However, the influence of evaluator on that relationship was not previously assessed and, given the lack of nexus between sexual offense and mental illness as well as a lack of direct relevance of Static-99R items to mental health diagnoses, it was believed different people might conceptualize this relationship differently. In fact, all evaluators seemed to place moderate weight on Static-99R scores in determining the presence or absence of qualifying behavioral abnormality.

On the other hand, while PCL-R scores were expected to behave similarly across evaluators, significant differences were found. Given that a measure of psychopathic traits is, by definition, a measure of personality, it is surprising that personality is not equivalently considered in determining BA. However, as noted above, evaluators may conceptualize the type of information obtained from the PCL-R differently (Boccaccini, Chevalier, Murrie, & Varela, 2017).

Further, despite the fact that all states with SVP statutes deem personality disorder diagnoses as appropriately meeting the standards of mental illness encapsulated in the statutes, there has been some debate regarding the qualification of personality disorder diagnoses in SVP civil commitment criteria. Specifically, antisocial personality disorder, and by extension psychopathy, are contentious uses of mental illness/abnormalities in these proceedings (Cauley, 2007; Hamilton, 2002; Johnson & Elbogen, 2013; Prentky et al., 2008; Zonana et al., 1999). As such, it is entirely possible that evaluators herein represent differing schools of thought on this subject, leading to differing weights placed on data pertaining to personality factors.

Additionally, it is possible the particular diagnostic makeup of offenders differed across evaluators, as data are representative of independent samples for each evaluator. Prior research suggests PCL-R scores tend to be significantly different across diagnostic profiles, and particularly with regard to pedophilic diagnoses (Vess, Murphy, & Arkowitz, 2004; Elwood, Doren, & Thornton, 2008; Jackson & Richards, 2007; Becker, Stinson, Tromp & Messer, 2003; Levenson 2004b). Thus, significant differences in the diagnosis of pedophilic as opposed to non-pedophilic disorders across the evaluators' samples may have contributed to individual differences in PCL-R scores. What's more, PCL-R scores are more strongly associated with sexual recidivism for offenders whose offenses involved adults, and less so for those whose offenses involved minors (Witt & Conroy, 2009). Thus, differences in diagnostic makeup may have influenced the subjective value of the PCL-R in considering risk.

Moreover, discrepancies in the relationship between risk measures and BA opinion may have nothing to do with the measures at all. It has been found that SVP commitment evaluators often do not demonstrate strong inter-rater agreement regarding assigned diagnoses, both with paraphilic disorders but especially with nonparaphilic disorders. It may well be the case here that the BA finding was more strongly related to differences in diagnosis than Static-99R and PCL-R scores.

The variable predictive effects across evaluators of the measures used in the current study add to the literature suggesting professional opinions are not developed via similar factors. While it is outside the scope of this writing to assert which factors, specifically, should determine the qualification for SVP civil commitment, it is concerning that empirically-based guidance has been posited (Rettenberger & Craig, 2020; Ahrendt & O'Donohue, 2019; Kelley & Thornton, 2015; Craig & Beech, 2010; Witt & Conroy, 2009) but does not appear to drive practice equally, and little consistency in what does drive decision-making appears to have been demonstrated. It should be acknowledged that no single method of determining risk for sexual re-offense has been labeled as "best." Indeed, those well-versed in the reality of sexual recidivism risk research acknowledge that some professional judgement is inherent in developing opinions that address varied legal questions: "Scoring an actuarial risk tool is not a risk assessment," (Hanson, 2009). It is also true, however, that whether an individual is confined for an indeterminate amount of time and whether significant resources are spent on confining such an individual should not be determined by the luck of the draw. It is our duty as a profession to ensure a degree of consistency across evaluators informed by the best science available.

Limitations and Future Directions

There were several limitations to this study. First, this sample consisted of evaluations conducted by only three evaluators. An examination incorporating a larger sample of SVP commitment evaluators would allow for a more comprehensive examination of factors influencing professional decision-making.

Further, current data includes only those evaluations completed within a southeastern region of Texas. While this geographic restriction may have limited, to some extent, the differences in populations evaluated across evaluators in the study, the results here may not generalize to populations within all jurisdictions in which SVP civil commitment evaluations are conducted. This may be especially true with regard to the requirement for examination of psychopathy within the evaluation of qualifying mental illness. Future research should assess for similar trends in a more widespread group of clinicians practicing across different jurisdictions.

Additionally, the current analyses did not take into account the specific training or years of experience of each evaluator. The amount of time a clinician has performed his/her duties can impact, for example, diagnostic accuracy (Spengler et al., 2009) and future research should include a look into whether clinicians of differing training background and/or years of SVP evaluation experience influence these outcome variables.

It should also be noted that many variables, including several additional measures, were reviewed by evaluators in the course of their evaluations, and only a few were examined for the purpose of the current study. As such, a number of other variables not specifically controlled for could have impacted these findings. For example, when considering the influence of PCL-R scores on risk for future sexual violence, sexual deviance was not assessed in current analyses. It is possible there were inherent differences among the evaluators' samples with regard to sexual deviance which resulted in differing consideration of PCL-R scores.

Furthermore, while many variables were reviewed by each of the evaluators, the specific information available to any given evaluator in the course of his evaluation differed. For instance, some case files may have contained more or less complete psychiatric treatment records, incarceration mental health contact records, institutional misconduct reports, law enforcement investigation records, or psychiatric assessment reports, to name just a few possible sources of information. It is unknown how this variance in the totality of available information may impact clinical decision-making.

Conclusion

It is important that we make life-altering and resource-intensive decisions based on the most objective and factually accurate information available to us. Indeed, our forensic psychology ethics tell us to "strive for accuracy, honesty, and truthfulness in the science…and practice of forensic psychology" (American Psychological Association, 2013). Unfortunately, no matter how objective the measures and process, there is still human error to be entered into the equation. As such, it is imperative we, as psychologists, honestly examine the factors influencing our opinions, including those injected as a result of individual differences among clinicians, and aim to decrease irrelevant factors as much as possible. To that end, opinions offered to the court in the course of SVP civil commitment proceedings should continue to be examined for the utmost fidelity to available best practices.

REFERENCES

Acklin, M. W., Fuger, K., & Gowensmith, W. N. (2015). Examiner agreement and judicial consensus in forensic mental health evaluations. *Journal of Forensic Psychology Practice*, *15*, 318-343. https://dx.doi.org/10.1080/15228932.2015.1051447

Ahrendt, A. J., & O'Donohue, W. T. (2019). Sexually violent predator evaluations: Problems and proposals. In W. T. O'Donohue & D. S. Bromberg (Eds.), *Sexually violent predators: A clinical science handbook.* (pp. 199–216). Springer Nature Switzerland AG. https://doi-org.ezproxy.shsu.edu/10.1007/978-3-030-04696-5_12

American Psychological Association. (2013). Specialty guidelines for forensic psychology. *American Psychologist*, *68*(1), 7–19. https://doi.org/10.1037/a0029889

Barbaree, H. E., Seto, M. C., Langton, C. M., & Peacock, E. J. (2001). Evaluating the predictive accuracy of six risk assessment instruments for adult sex offenders. *Criminal Justice and Behavior*, *28*, 490-521.

Becker, J. V., Stinson, J., Tromp, S., & Messer, G. (2003). Characteristics of individuals petitioned for civil commitment. *International Journal of Offender Therapy and Comparative Criminology*, *47*, 185-195.

Boccaccini, M. T., Chevalier, C. S., Murrie, D. C., & Varela, J. G. (2017). Psychopathy Checklist—Revised use and reporting practices in sexually violent predator evaluations. *Sexual Abuse: Journal of Research and Treatment*, *29*(6), 592–614. https://doi-org.ezproxy.shsu.edu/10.1177/1079063215612443

Boccaccini, M. T., Murrie, D. C., Mercado, C., Quesada, S., Hawes, S., Rice, A. K., Jeglic, E. L. (2012). Implications of Static-99 field reliability findings for score use and reporting. *Criminal Justice and Behavior, 39*, 1, 42-58.

Boccaccini, M. T., Murrie, D. C., Rufino, K. A., & Gardner, B. O. (2014). Evaluator differences in Psychopathy Checklist-Revised factor and facet scores. *Law and Human Behavior, 38*, 4, 337-345.

Boccaccini, M. T., Rice, A. K., Helmus, L. M., Murrie, D. C., & Harris, P. B. (2017). Field validity of Static-99/R scores in a statewide sample of 34,687 convicted sexual offenders. *Psychological Assessment*, *29*, 611-623. https://doi:10.1037/pas0000377

Boccaccini, M. T., Turner, D. B., & Murrie, D. C. (2008). Do some evaluators report consistently higher or lower PCL-R scores than others? Findings from a statewide sample of sexually violent predator evaluations. *Psychology, Public Policy, and Law*, *14*(4), 262–283. https://doi-org.ezproxy.shsu.edu/10.1037/a0014523

Boccaccini, M., Turner, D., Murrie, D., Henderson, C., & Chevalier, C. (2013). Do scores from risk measures matter? *Psychology, Public Policy, and Law, 19*, 259–269. https://doi:10.1037/a0031354

Boccaccini, M. T., Turner, D. B., Murrie, D. C., & Rufino, K. A. (2012). Do PCL-R scores from state or defense experts best predict future misconduct among civilly committed sex offenders? *Law and Human Behavior*, *36*, 159-169.

Boer, D. P., Hart, S. D., Kropp, P. R., & Webster, C. D. (1997). *The Sexual Violence Risk-20: Professional guidelines for assessing risk of sexual violence.* Institute Against Family Violence.

Borum, R., Bartel, P., & Forth, A. (2006). *Structured Assessment of Violence Risk in Youth (SAVRY)*. Psychological Assessment Resources, Inc.

Calleja, N. G. (2015). Juvenile sex and nonsex offenders: A comparison of recidivism and risk. *Journal of Addictions & Offender Counseling*, *36*(1), 2–12. https://doi:10.1002/j.2161-1874.2015.00031.x

Cauley, D. R. (2007). The diagnostic issue of antisocial personality disorder in civil commitment proceedings: A response to DeClue. *Journal of Psychiatry & Law*, *35*(4), 475–497.

Chevalier, C., Boccaccini, M. T., Murrie, D. C., & Varela, J. G. (2015). Static-99R reporting practices in sexually violent predator cases: Does norm selection reflect adversarial allegiance? *Law and Human Behavior, 39*, 209-218.

Craig, L. A., & Beech, A. R. (2010). Towards a guide to best practice in conducting actuarial risk assessments with sex offenders. *Aggression and Violent Behavior*, *15*(4), 278–293. https://doi-org.ezproxy.shsu.edu/10.1016/j.avb.2010.01.007

DeMatteo, D., Murphy, M., Galloway, M., & Krauss, D. (2015). A national survey of United States Sexually Violent Person legislation: Policy, procedures, and practice. *International Journal of Forensic Mental Health, 14*, 245-266.

Doren, D. M. (2002). *Evaluating sex offenders: A manual for civil commitments and beyond*. Sage Publishers.

Edens, J. F., Boccaccini, M. T., Johnson, D., & Johnson J. (2010). Inter-rater reliability of the PCL-R total and factor scores among psychopathic sex offenders: Are personality features more prone to disagreement than behavioral features? *Behavioral Sciences & The Law*, *28*, 106-119.

Elwood, R. W. (2019). Agreement between courts and SVP evaluators in the State of Wisconsin. *Criminal Justice and Behavior, 46*(6), 853–865. https://doi-org.ezproxy.shsu.edu/10.1177/0093854819839746

Elwood, R. W., Doren, D. M., & Thornton, D. (2008). Diagnostic and risk profiles of men detained under Wisconsin's sexually violent person law. *International Journal of Offender Therapy and Comparative Criminology, 54*, 187-196.

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.

Guarnera, L. A., & Murrie, D. C. (2017). Field reliability of competency and sanity opinions: A systematic review and meta-analysis. *Psychological Assessment*, *29*(6), 795–818. https://doi-org.ezproxy.shsu.edu/10.1037/pas0000388

Guy, L. S. (2008). *Performance indicators of the structured professional judgment approach for assessing risk for violence to others: A meta-analytic survey* (Unpublished doctoral dissertation). Simon Fraser University.

Guy, L. S., Kusaj, C., Packer, I. K., & Douglas, K. S. (2015). Influence of the HCR-20, LS-CMI, and PCL-R on decision about parole suitability among lifers. *Law and Human Behavior, 39*, 232-243.

Hamilton, G. S. (2002). The blurry line between “mad” and “bad”: Is “lack-of-control” a workable standard for sexually violent predators? *University of Richmond Law Review, 36*(2), 480–508. http://scholarship.richmond.edu/lawreview/vol36/iss2/9

Hanson, R. K. (2009). The psychological assessment of risk for crime and violence. *Canadian Psychology, 50*(3), 172–182. https://doi.org/10.1037/a0015726

Hanson, R. K., & Morton-Bourgon, K. E. (2004). *Predictors of sexual recidivism: An updated meta-analysis.* Public Works and Government Services.

Hanson, R. K., & Morton-Bourgon, K. E. (2009). The accuracy of recidivism risk assessments for sexual offenders: A meta-analysis of 118 prediction studies. *Psychological Assessment, 12,* 1-21.

Hanson, R. K., Morton, K. E., & Harris, A. J. R. (2003). Sexual offender recidivism risk: What we know and what we need to know. In R. A. Prentky, E. S. Janus, & M. C. Seto (Eds.), *Understanding and managing sexually coercive behavior* (Vol. 989, pp. 154–166). Annals of the New York Academy of Sciences.

Hanson, R. K., & Thornton, D. (2000). Improving risk assessments for sex offenders: A comparison of three actuarial scales. *Law and Human Behavior*, *24*(1), 119–136. https://doi-org.ezproxy.shsu.edu/10.1023/A:1005482921333

Hare, R. D. (2003). *PCL-R technical manual* (2nd Ed.). Multi-Health Systems.

Hawes, S. W., Boccaccini, M. T., & Murrie, D. C. (2013). Psychopathy and the combination of psychopathy and sexual deviance as predictors of sexual recidivism: Meta-analytic findings using the Psychopathy Checklist—

Revised. *Psychological Assessment*, *25*(1), 233–243. https://doi-org.ezproxy.shsu.edu/10.1037/a0030391

Helmus, R. K., (2008). Static-99 replications: Descriptive information. *Static-99 website*. http://www.static99.org/pdfdocs/static-99replicationsdescriptives.pdf

Helmus, L., Hanson, R. K., Thornton, D., Babchishin, K. M., & Harris, A. R. (2012). Absolute recidivism rates predicted by Static-99R and Static-2002R sex offender risk assessment tools vary across samples: A meta-analysis. *Criminal Justice and Behavior*, *39*, 1148-1171. https://doi.org/10.1177/0093854812443648

Helmus, L., & Thornton, D. (2015). Stability and predictive and incremental accuracy of the individual items of Static-99R and Static-2002R in predicting sexual recidivism: A meta-analysis. *Criminal Justice and Behavior*, *42*, 917-937. https://doi:10.1177/0093854814568891

Helmus, L., Thornton, D., Hanson, R. K., & Babchishin, K. M. (2012). Improving the predictive accuracy of Static-99 and Static-2002 with older sex offenders: Revised age weights. *Sexual Abuse: Journal of Research and Treatment*, *24*(1), 64–101. https://doi-org.ezproxy.shsu.edu/10.1177/1079063211409951

Hemphill, J. F., Templeman, R., Wong, S., & Hare, R. D. (1998). Psychopathy and crime: Recidivism and criminal careers. In D. Cooke, A. Forth, & R. D. Hare (eds.), *Psychopathy: Theory, research, and implications for society* (pp. 375-398). Kluwer Academic.

Hilton, N. Z., & Simmons, J. L. (2001). The influence of actuarial risk assessment in clinical judgements and tribunal decisions about mentally disordered offenders in maximum security. *Law and Human Behavior, 25*, 393-408.

Hilton, N. Z., Simpson, A. I., & Ham, E. (2016). The increasing influence of risk assessment on forensic patient review board decisions. *Psychological Services, 13*, 223-231.

Ismail, G., & Looman, J. (2018). Field inter-rater reliability of the Psychopathy Checklist-Revised. *International Journal of Offender Therapy and Comparative Criminology, 62*, 2, 468-481.

Jackson, R. L., & Richards, H. J. (2007). Diagnostic and risk profiles among civilly committed sexual offenders in Washington state. *International Journal of Offender Therapy and Comparative Criminology, 51*, 313-324.

Johnson, S. C., & Elbogen, E. B. (2013). Personality disorders at the interface of psychiatry and the law: Legal use and clinical classification. *Dialogues in Clinical Neuroscience*, *15*(2), 203–211. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3811091/

Kahn, R. E., Kelley, S. M., Mundt, J. C., Ambroziak, G., Barahal, R. M., & Thornton, D. (2022, February 7). Risk-tinted spectacles: What influences evaluator decision making in sexually violent persons examinations. *Psychology, Public Policy, and Law*. Advance online publication. https://dx.doi.org/10.1037/law0000333

Kansas v. Crane. 534 U.S. 407 (2002).

Kansas v. Hendricks. 521 U.S. 346 (1997).

Kelley, S. M., & Thornton, D. (2015). Assessing risk of sex offenders with major mental illness: Integrating research into best practices. *Journal of Aggression, Conflict and Peace Research*, *7*(4), 258–274. https://doi-org.ezproxy.shsu.edu/10.1108/JACPR-02-2015-0162

Kroner, D. G., & Mills, J. F. (2001). The accuracy of five risk appraisal instruments in predicting institutional misconduct and new convictions. *Criminal Justice and Behavior, 28,* 471-489.

Kurus, S.J., Holdren, S.M., Rubenstein, L., Varela, J. G., Harris, P. B., Strauss, J. P., Franklin, D. W., Turner, D. B., Hamilton, P. M. (2019). *Correlates of behavioral abnormality among sexual offenders evaluated for civil commitment as sexually violent predators*. Manuscript in preparation.

Krauss, D., & Scurich, N. (2014). The impact of case factors on jurors' decisions in a sexual violent predator hearing. *Psychology, Public Policy, and Law*, *20*(2), 135–145. https://doi-org.ezproxy.shsu.edu/10.1037/law0000007

Leistico, A.-M. R., Salekin, R. T., DeCoster, J., & Rogers, R. (2008). A large-scale meta-analysis relating the Hare measures of psychopathy to antisocial conduct. *Law and Human Behavior, 32*(1), 28–45. https://doi-org.ezproxy.shsu.edu/10.1007/s10979-007-9096-6

Levenson, J. S. (2004a). Reliability of sexually violent predator civil commitment criteria in Florida. *Law and Human Behavior, 8*, 357-368.

Levenson, J. S. (2004b). Sexual predator civil commitment: A comparison of selected and released offenders. *International Journal of Offender Therapy and Comparative Criminology, 48*, 638-648.

Levenson, J. S., Brannon, Y. N., Fortney, T., & Baker, J. (2007). Public perceptions about sex offenders and community protection policies. *Analyses of Social Issues and Public Policy (ASAP), 7*(1), 137–161.

Levenson, J. S., & Morin, J. W. (2006). Factors predicting selection of sexual violent predators for civil commitment. *International Journal of Offender Therapy and Comparative Criminology*, *50*, 609-629.

Levenson, J. S., & Morin, J. W. (2006). Factors predicting selection of sexually violent predators for civil commitment. *International Journal of Offender Therapy and Comparative Criminology*, *50*(6), 609–629. https://doi-org.ezproxy.shsu.edu/10.1177/0306624X06287644

Looman, J., Morphett, N., & Abracen, J. (2012). Does consideration of psychopathy and sexual deviance add to the predictive validity of the Static-99R? *International Journal of Offender Therapy and Comparative Criminology*. Advance online publication. https://doi:10.1177/ 0306624x12444839

Lloyd, C. D., Clark, H. J., & Forth, A. E. (2010). Psychopathy, expert testimony, and indeterminate sentences: Exploring the relationship between Psychopathy Checklist-Revised testimony and trial outcome in Canada. *Legal and Criminological Psychology*, *15*, 323-339.

McCabe, J. G., Krauss, D. A., & Lieberman, J. D. (2010). Reality check: A comparison of college students and community samples of mock jurors in a simulated sexual violent predator civil commitment. *Behavioral Sciences & the Law*, *28*, 730–750. https://doi.org/10.1002/bsl.902

McCallum, K. E., Boccaccini, M. T., & Bryson, C. N. (2017). The influence of risk assessment instrument scores on evaluators' risk opinions and sexual offender containment recommendations. *Criminal Justice and Behavior, 44*, 9, 1213-1235.

McGraw, K. O., & Wong, S. P. (1996). Forming inferences about some intraclass correlation coefficients. *Psychological Methods, 1,* 30-46.

Miller, H. A., Amenta, A. E., & Conroy, M. A. (2005). Sexually violent predator evaluations: Empirical evidence, strategies for professionals, and research directions. *Law and Human Behavior*, *29*(1), 29–54. https://doi-org.ezproxy.shsu.edu/10.1007/s10979-005-1398-y

Miller, C. S., Kimonis, E. R., Otto, R. K., Kline, S. M., & Wasserman, A. L. (2012). Reliability of risk assessment measures used in sexually violent predator proceedings. *Psychological Assessment*, *24*, 4, 944-953.

Miller, J., & Maloney, C. (2013). Practitioner compliance with risk/needs assessment tools: A theoretical and empirical assessment. *Criminal Justice and Behavior, 40*, 716-736.

Monahan, J., & Skeem, J. (2016). Risk assessment in criminal sentencing. *Annual Review of Clinical Psychology, 12*. Advance online publication. https://dx.doi.org/10.1146/annurev-clinpsy-021815-092945

Murrie, D. C., Boccaccini, M. T., Johnson, J. T., & Janke, C. (2008). Does interrater (dis)agreement on psychopathy on Psychopathy Checklist scores in sexually violent predator trials suggest partisan allegiance in forensic evaluations? *Law and Human Behavior, 32* 351-362.

Neal, T. M. S., & Grisso, T. (2014). Assessment practices and expert judgment methods in forensic psychology and psychiatry: An international snapshot. *Criminal Justice and Behavior*, *41*, 1406-1421.

Porter, S., Woodworth, M., Earle, J., Drugge, J., & Boer, D. P. (2003). Characteristics of violent behavior exhibited during sexual homicides by psychopathic and non-psychopathic murders. *Law and Human Behavior, 27*, 459-470.

Prentky, R. A., Coward, A. I., & Gabriel, A. M. (2008). Commentary: Muddy diagnostic waters in the SVP courtroom. *The Journal of the American Academy of Psychiatry and the Law, 36*(4), 455–458. https://pubmed.ncbi.nlm.nih.gov/19092061/

Quinsey, V. L., Harris, G. T., Rice, M. E., & Cormier, C. (1998). *Violent offenders: Appraising and managing risk.* American Psychological Association.

Rettenberger, M., & Craig, L. A. (2020). Sexual Violence Risk Assessment. In J. S. Wormith, L. A. Craig, & T. E. Hogue (Eds.), *The Wiley handbook of what works in violence risk management: Theory, research, and practice.* (pp. 183–201). Wiley Blackwell. https://doi-org.ezproxy.shsu.edu/10.1002/9781119315933.ch9

Rettenberger, M., Matthes, A., Boer, D. P., & Eher, R. (2010). Prospective actuarial risk assessment: A comparison of five risk assessment instruments in different sexual offender subtypes. *International Journal of Offender Therapy and Comparative Criminology*, *54*, 169 –186. https://doi: 10.1177/0306624X08328755

Rutherford, M., Cacciola, J. S., Alterman, A. I., McKay, J. R., & Cook, T. G. (1999). The 2-year test retest reliability of the Psychopathy Checklist-Revised in methadone patients. *Assessment, 6*, 285-291.

Salekin, R. T., Rogers, R., & Sewell, K. W. (1996). A review and meta-analysis of the Psychopathy Checklist and Psychopathy Checklist—Revised: Predictive validity of dangerousness. *Clinical Psychology: Science and Practice, 3*(3), 203–215. https://doi-org.ezproxy.shsu.edu/10.1111/j.1468-2850.1996.tb00071.x

Seto, M. C., & Lalumiere, M. L. (2001). A brief screening scale to identify pedophilic interests among child molesters. *Sexual Abuse*, *13,* 15-25.

Shalverson, R. J., & Webb, N. M. (1991). *Generalizability theory: A primer*. Sage Publishers.

Spengler, P.M., White, M. J., Ægisdóttir, S., Maugherman, A. S., Anderson, L. A., Cook, R. S., et al. (2009). The meta-analysis of clinical judgment project: Effects of experience on judgment accuracy. *The Counseling Psychologist, 37*, 350–399. http://dx.doi.org/10. 1177/0011000006295149.

Tex. Health & Safety Code Ann. § 841.023, 2011.

Vess, J., Murphy, C., & Arkowitz, S. (2004). Clinical and demographic differences between sexually violent predators and other commitment types in a state forensic hospital. *Journal of Forensic Psychiatry and Psychology, 15*, 669-681.

Vess, J., & Skelton, A. (2010). Sexual and violent recidivism by offender type and actuarial risk: Reoffending rates for rapists, child molesters and mixed-victim offenders. *Psychology, Crime & Law*, *16*(7), 541–554. https://doi:10.1080/ 10683160802612908

Viljoen, J. L., Jonnson, M. R., Cochrane, D. M., Vargen, L. M., & Vincent, G. M. (2019). Impact of risk assessment instruments on rates of pretrial detention, postconviction placements, and release: A systematic review and meta-analysis. *Law and Human Behavior*, *43*(5), 397-420.

Walters, G. D., Knight, R. A., Grann, M., & Dahle, K. P. (2008). Incremental validity of the Psychopathy Checklist facet scores: Predicting release outcome in six samples. *Journal of Abnormal Psychology, 117*, 396 – 405. https://doi:10.1037/0021-843X.117.2.396

Witt, P. H., & Conroy, M. A. (2009). *Evaluation of sexually violent predators*. Oxford University Press.

Wong, S., Olver, M. E., Nicholaichuk, T. P., & Gordon, A. (2004). *The Risk Scale: Sex offender version.* Regional Psychiatric Center and University of Saskatchewan.

Zonana, H., Abel, G., Bradford, J., Hoge, S. K., Metzner, J., Becker, J., Bonnie, R., Fitch, L., & Hughes, L. (1999). Dangerous sex offenders: A task force report of the American Psychiatric Association. *American Psychiatric Association*.

Ægisdóttir, S., White, M. J., Spengler, P. M., Maugherman, A. S., Anderson, L. A., Cook, R. S., … Rush, J. D. (2006). The meta-analysis of clinical judgment project: Fifty-six years of accumulated research on clinical versus statistical prediction. *The Counseling Psychologist, 34*, 341–382. https://doi-org.ezproxy.shsu.edu/10.1177/0011000005285875

VITA

Samantha J. Kurus, M.S.

EDUCATION

Candidate (Expected 2022)
Doctor of Philosophy, Clinical Psychology with Forensic Emphasis
Sam Houston State University – Huntsville, TX
Dissertation: Do Static-99R and PCL-R scores predict risk opinion equally across evaluators in SVP civil commitment evaluations?
Co-Chairs: Jorge G. Varela, Ph.D. and Marcus T. Boccaccini, Ph.D.

2014
Master of Science, Psychology (Clinical)
California State University, Fullerton – Fullerton, CA
Thesis: Public attitudes toward sex offenders, sex offender treatment, and sex offender rehabilitation: A comparison of attitudes between the United States and the United Kingdom
Chair: Nancy Panza, Ph.D.

2008
Bachelor of Arts, Psychology
University of Wisconsin, Madison
Minor: Criminal Justice

SUPERVISED CLINICAL POSITIONS

08/2021 To 08/2022
Pre-Doctoral Intern, APA-Accredited Internship
Mendota Mental Health Institute
Populations: Ethnically diverse, male adults committed for competency to stand trial evaluation, treatment to competence, or acquittal of not guilty by reason of insanity; ethnically diverse justice-involved youth
Training Directors: David Lee, Ph.D.; Karyn Gust-Brey, Ph.D.

06/2020 To 08/2021
Neuropsychology Student Clinician
Center for Optimal Brain Health
Population: Ethnically diverse, male and female adults (majority older adult) in the community.
Supervisors: Andres Tapia, Ph.D. (primary); Anthony Ward, Ph.D.
.

03/2018 To 08/2021
Practicum Student – Individual Evaluator
Psychological Services Center at Sam Houston State University
Population: Ethnically diverse, justice-involved youth.
Supervisor: Darryl Johnson, Ph.D.

12/2017 To 08/2021
Assistant Forensic Evaluator
Psychological Services Center at Sam Houston State University
Population: Ethnically diverse, male and female, adults and adolescents involved in the justice system in several rural counties
Supervisors: Mary Alice Conroy, Ph.D., ABPP, Wendy Elliot, Ph.D., & Darryl Johnson, Ph.D.

08/2017 **To** **08/2021**	***Practicum Student – Individual Therapist & Evaluator*** **Psychological Services Center at Sam Houston State University** *Population:* Diverse, low-income, multi-ethnic community population of children, adolescents, and adults with a range of diagnoses. *Supervisors:* Jaime Anderson, Ph.D., Craig Henderson, Ph.D., Darryl Johnson, Ph.D., Chelsea Ratcliff, Ph.D., Jorge Varela, Ph.D.
06/2019 **To** **09/2020**	***Student Clinician*** **Rusk State Hospital** *Population:* Ethnically diverse, male and female, adults committed due to competence restoration, and not guilty by reason of insanity (NGRI). *Supervisor:* Sarah Rogers, Ph.D. (primary); Patricia Plasay, Ph.D.
05/2018 **To** **06/2019**	***Clinic Coordinator*** **Psychological Services Center at Sam Houston State University** *Population:* Diverse, low-income, multi-ethnic community population of children, adolescents, and adults with a range of diagnoses. *Supervisors:* Mary Alice Conroy, Ph.D., ABPP, Clinic Director (primary); Wendy Elliot, Ph.D., Darryl Johnson, Ph.D., Jaime Anderson, Ph.D., & Chelsea Ratcliff, Ph.D.
09/2012 **To** **09/2013**	***Marriage and Family Therapist Trainee*** **Straight Talk Counseling** *Population:* Diverse, low-income, multi-ethnic community population of children, adolescents, and adults with a range of diagnoses. *Supervisors:* Marilyn Davis, Ph.D., Arthur Holden, Ph.D.
05/2008 **To** **08/2008**	***Undergraduate Student Clinician*** **Forensic Units at Mendota Mental Health Institute** *Population:* Ethnically diverse, male adults hospitalized for voluntary and involuntary commitment, competence restoration, and not guilty by reason of insanity (NGRI). *Supervisor:* Michael Vitacco, Ph.D.

TEACHING AND SUPERVISORY EXPERIENCE

05/2020 **To** **08/2020**	***Peer Supervisor*** **Beginning Doctoral Practicum** **Sam Houston State University** *Supervisor:* Mary Alice Conroy, Ph.D., ABPP
05/2020 **To** **08/2020**	***Graduate Teaching Assistant*** **Beginning Doctoral Practicum** **Sam Houston State University** *Supervisor:* Mary Alice Conroy, Ph.D., ABPP
08/2018 **To** **06/2020**	***Peer Supervisor*** **Capstone Practicum** **Sam Houston State University** *Supervisors:* Craig Henderson, Ph.D., Temilola Salami, Ph.D, & David Nelson, Ph.D.

08/2019 *Graduate Teaching Assistant*
To **Assessment of Intelligence and Achievement**
05/2020 **Sam Houston State University**
<u>Supervisor:</u> Craig Henderson, Ph.D.

09/2016 *Instructor of Record*
To **Sam Houston State University**
09/2017 <u>Supervisor:</u> Christopher Wilson, Ph.D.

01/2016 *Adjunct Faculty*
To **Department of Psychology; Department of Education**
08/2016 **Lonestar College**
<u>Supervisors:</u> Debra Parish; Gloria Maristany, M.Ed.

08/2014 *Part-Time Lecturer*
To **Department of Psychology**
08/2015 **California State University, Fullerton**
<u>Supervisor:</u> Jack Mearns, Ph.D.

PERIODICAL AND PEER-REVIWED PUBLICATIONS

Ricardo, M. M. & **Kurus**, S. J. (2020). Kahler v. Kansas: A new challenge to the insanity defense. *Texas Psychologist, Winter 2020, 79* (3).

Vitacco, M.J., Erickson, S.K., **Kurus**, S.J. & Apple, B. (2012). The role of the Violence Risk Appraisal Guide and Historical, Clinical, Risk-20 in U.S. courts: A case law survey. *Psychology, Public Policy, and Law, 18*, 361-391.

Vitacco, M.J., Erickson, S.K., **Kurus**, S.J., Apple, B.N., Lamberti, J.S., & Gasser, D. (2011). Evaluating conditional release in female insanity acquittees: A risk management perspective. *Psychological Services, 8,* 332-342.

Rogers, R., Vitacco, M.J., & **Kurus**, S.J. (2010). Assessment of malingering with repeat forensic evaluations: Patient variability and possible misclassification on the SIRS and other feigning measures. *Journal of the American Academy of Psychiatry and the Law, 38,* 109-114.

Vitacco, M.J., Caldwell, M., Ryba, N., Malesky, A., & **Kurus**, S. (2009). Assessing risk in adolescent sex offenders: Recommendations for clinical practice. *Behavioral Sciences & the Law,* 27, 929-940.

CONFERENCE PRESENTATIONS

Kurus, S. J., Holdren, S. M., Rubenstein, L., Ricardo, M. M., Varela, J. G., Boccaccini, M. T., Turner, D. B., Hamilton, P. M. (2022). *Static or dynamic? Evaluator differences in trends in Static-99R Scoring, sex offender treatment completion, and risk level determination in sexually violent predator civil commitment evaluations*. Poster presented at the Annual American Psychology-Law Society Conference, Denver, CO.

Kurus, S.J., Holdren, S.M., Rubenstein, L., Varela, J. G., Boccaccini, M. T., Harris, P. B., Turner, D. B., Hamilton, P. M. (2020). *Evaluator differences in trends in PCL-R scoring and risk*

level determination in sexually violent predator civil commitment evaluations. Poster presented at the Annual American Psychology-Law Society Conference, New Orleans, LA.

Rubenstein, L., Varela, J. G., Holdren, S. M., **Kurus**, S. J., Harris, P. B., Turner, D. B. (2020). *Predictors of Behavioral Abnormality findings in Sexually Violent Predator Evaluations*. Poster presented at the Annual American Psychology-Law Society Conference, New Orleans, LA.

Ricardo, M. M., Henderson, C. E., Anderson-White, E., Christensen, M. R., Krembuszewski, B. & **Kurus**, S. J. (2019). Assumptions of Defendant Identity at the Intersection of Crime and Substance Use. Poster presented at the annual convention of the American Psychological Association, Chicago, IL.

Kurus, S.J., Holdren, S.M., Rubenstein, L., Varela, J. G., Harris, P. B., Strauss, J. P., Franklin, D. W., Turner, D. B., Hamilton, P. M. (2019). *Correlates of behavioral abnormality among sexual offenders evaluated for civil commitment as sexually violent predators*. Poster presented at the Annual American Psychology-Law Society Conference, Portland, OR.

Kurus, S.J., & Luzzi, M. (2015). Fostering student success: Effective strategies to support probation and disqualified students in regaining academic success. Presented at the 2015 Annual Academic Advisors Professional Development Conference, California State University,Fullerton.

RESEARCH EXPERIENCE

01/20 **To** **08/2022**	***Co-Principal Investigator*** **Does Presence and Timing of OTT Impact Course of Restoration in Those Ordered to Treatment to Competency?** **Mendota Mental Health Institute** *Supervisors:* David M. Lee, Ph.D. & Karyn L. Gust-Brey, Ph.D.
01/2019 **To** **08/2022**	***Principal Investigator*** **Do Static-99R and PCL-R Scores Predict Risk Opinion Equally Across Evaluators in SVP civil commitment evaluations?** **Sam Houston State University** *Supervisors:* Jorge Varela, Ph.D. & Marcus Boccaccini, Ph.D.
10/2016 **To** **08/2021**	***Graduate Research Assistant*** **Department of Psychology and Philosophy** **Sam Houston State University** *Supervisors:* Marc Boccaccini, Ph.D.; Jorge Varela, Ph.D.
09/2018 **To** **07/2019**	***Graduate Research Assistant*** **Department of Psychology and Philosophy** **Sam Houston State University & Law Psychiatry and Public Policy at the University of Virginia** *Supervisors:* Marc Boccaccini, Ph.D., Brett Gardner, Ph.D., Daniel Murrie, Ph.D., Dr. Angela Torres, Ph.D.
01/2012 **To** **08/2014**	***Principal Investigator*** **Public Attitudes Toward Sex Offenders, Sex Offender Treatment, and Sex Offender Rehabilitation: A Comparison of Attitudes Between the United States and United Kingdom (Master's Thesis)** **California State University, Fullerton** *Supervisor:* Nancy Panza, Ph.D.

05/2010 *Research Assistant*
To **Risk Assessment in Court Cases**
07/2011 **Georgia Regents University**
Supervisor: Michael Vitacco, Ph.D.

01/2009 *Undergraduate Research Assistant*
To **Stability of the Structured Interview of Reported Symptoms**
01/2010 **Mendota Mental Health Institute**
Supervisor: Michael Vitacco, Ph.D.

08/2008 *Undergraduate Research Assistant*
To **Sex Offenders and Insanity**
01/2009 **Mendota Mental Health Institute**
Supervisor: Michael Vitacco, Ph.D.

11/2007 *Undergraduate Research Assistant*
To **Juvenile Assessment Data Compilation**
05/2008 **Mendota Juvenile Treatment Center**
Supervisor: David McCormick, Ph.D.

PROFESSIONAL DEVELOPMENT SERVICE

09/2020 *Volunteer Event Assistant*
American Academy of Forensic Psychology
Web-based Continuing Education Workshops

03/2018 *Student Volunteer*
American Psychology-Law Society Annual Conference
Memphis, TN

08/2017 *Secretary*
To **Graduate Student Psychology Organization**
08/2018 **Sam Houston State University**

12/2016 *Student Manuscript Reviewer*
To **Law and Human Behavior**
02/2017 **Faculty Supervisor: Marc Boccaccini, Ph.D.**

PROFESSIONAL MEMBERSHIPS

2016 – Present American Psychology-Law Society (Division 41)
2011 – Present American Psychological Association
2021 – 2022 Society for Clinical Neuropsychology (Division 40)
2018 – 2020 Therapeutic Assessment Institute
2016 – 2019 Association for Psychological Science