Are File Review-Based SAVRY Ratings of Violence Risk Reliable?

A Dissertation

Submitted to the Faculty

of

Drexel University

by

Jeffrey D. Burl

in partial fulfillment of the

requirements for the degree

of

Doctor of Philosophy

May 2012

Dedications

To Melissa, my patient, loving wife. You have worked around the demands of my academic pursuits and have sacrificed many of your own goals and desires in the process. Through it all, you have not hesitated or voiced a complaint, but have stood by me and given nothing but encouragement. On the occasions when I have doubted my abilities, you have believed.

And to my parents, Jeff and Maureen. You have been constant and unwavering in your support since I began my graduate education (and, of course, throughout all my earlier schooling as well). Thank you for teaching me to strive for excellence and to always follow my path.

Acknowledgments

David DeMatteo is an incredible advisor, teacher, and friend. His guidance and mentorship over the past five years have not only helped me grow from a green first year student into a competent and skilled clinician, but have also taught me how to define "success" as a psychologist. Thank you for all your help on this project, including the great volumes of time you spent talking with me in its early development, fielding my countless questions throughout, and providing constructive feedback during the writing.

I am also grateful for my dissertation committee. Kirk Heilbrun was instrumental to this study's success. He helped me conceptualize the project and generously opened up his clinic's archives to allow me to examine the de-identified juvenile files. Naomi Goldstein, Jamie Loving, and Aileen Fink each lent their clinical expertise to this study and were encouraging from proposal to defense.

Casey LaDuke, Heidi Strohmaier, Megan Strobel, Meghann Hudson, and Sarah Filone were the best research assistants I could have asked for. They gave their time, energy, and effort without compensation or reward. I cannot thank each of you enough for all that you gave to this project.

Table of Contents

List of Tables vii

Abstract viii

Chapter 1: Introduction and Literature Review 1

1.1 Background 1

1.2 Risk Assessment 2

1.2.1 Unstructured clinical judgment. 3

1.2.2 Subsequent developments in risk assessment. 4

1.3 Juvenile Risk Assessment 6

1.3.1 Risk factors for juvenile violence. 6

1.3.2 Juvenile risk assessment measures. 8

1.4 The SAVRY 11

1.4.1 Standard SAVRY coding procedure. 13

1.4.2 Psychometric data on the SAVRY. 15

1.4.3 Use of multiple sources in coding the SAVRY. 15

1.5 The Current Study 16

1.6 Hypotheses 18

Chapter 2: Methods 20

2.1 Setting 20

2.2 Participants 21

2.3 Procedure 23

2.3.1 De-identification of files. 23

2.3.2 File contents. 23

2.3.3 Outcome measure 24

2.3.4 Training workshop. 24

2.3.5 Practice files. 25

Chapter 3: Results 27

3.1 Practice Files 27

3.2 SAVRY Summary Risk Ratings 27

3.2.1 Concordance of file-only ratings with expert ratings. 28

3.3 Individual SAVRY Risk and Protective Factors Ratings 29

3.3.1 Concordance of file-only ratings with expert ratings. 29

3.4 Performance of File-Only Raters 30

Chapter 4: Discussion 32

4.1 Causes of Missing Data 33

4.2 Lack of Congruency between File-Only and Expert Raters 36

4.3 Implications and Future Research 39

4.4 Limitations 42

Chapter 5: List of References 45

Appendix A: SAVRY Items 52

Vita..64

List of Tables

1. Previous Research on the SAVRY 54
2. Demographic Data 56
3. Examples of Decision Rules 58
4. SAVRY Summary Risk Ratings 59
5. Percentage of Individual SAVRY Risk Factor Item Ratings 60
6. Percentage of Individual SAVRY Protective Factor Item Ratings 62
7. Agreement with Experts by Rater 63

Abstract

Are File Review-Based SAVRY Ratings of Violence Risk Reliable?

Jeffrey Burl, M.A.

David DeMatteo, J.D., Ph.D.

Since its publication a decade ago, the Structured Assessment for Violence Risk in Youth (SAVRY) has gained acceptance as a strong predictor of future violence in adolescent populations. Clinicians scoring the SAVRY use their professional judgment to code a structured protocol of risk and protective factors based on clinical interviews, a review of the juvenile's records, and other sources of information. Much of the SAVRY validation research, however, has relied upon retrospective ratings obtained solely through file review. To date, no study has examined the reliability of file review-based SAVRY ratings. This study examined whether file-only SAVRY ratings are comparable to expert clinical ratings obtained through standard SAVRY administration procedures. Results indicate that file-only raters were unable to provide Summary Risk Ratings for 43% of the files and were unable to rate 53% of the SAVRY's individual items. The ratings that were coded by the file-only raters had low to moderate levels of agreement with the expert ratings. These results suggest that file-only SAVRY coding is not a reliable manner in which to obtain risk assessment ratings, but the current findings conflict with low rates of missing data in previous file-only SAVRY research. Further research should therefore be undertaken to provide greater clarity as to whether file-only SAVRY ratings of violence risk are reliable.

Are File Review-Based SAVRY Ratings of Violence Risk Reliable?
Chapter 1: Introduction and Literature Review

1.1 Background

In the United States, adolescents are responsible for a significant proportion of the overall number of violent crimes committed each year (Puzzanchera, 2009). Although rates of nonviolent juvenile crime have fallen over the last decade, many types of violent criminal behavior by juveniles have remained constant or increased. For example, since the year 2000, the arrest rate for murder has not decreased and the arrest rate for robbery has risen 15% (Puzzanchera & Adams, 2011). These and other violent activities cause profound physical, financial, and emotional loss to victims and are a major societal problem for many communities (Meyers & Schmidt, 2008).

Young perpetrators of violent crime are also subject to life-altering consequences, including many legal sanctions that equal what an adult would receive for the same crime due to punitive juvenile sentencing laws (Nellis, 2012). Originally, with the creation of the juvenile court in 1899, young offenders were processed in a rehabilitative-based system founded on the recognition of adolescence as a unique stage of human development (Scott, 2000). However, over the last 45 years, state laws have been rewritten to hold juveniles to a higher degree of responsibility for their actions. Several factors were responsible for this shift, including a series of Supreme Court decisions (e.g., *In re Gault*, 1967; *Kent v. United States*, 1966) and a public that grew increasingly concerned about the rising rate of juvenile crime (Melton, Petrila, Poythress, & Slobogin, 2007). In response, states began to focus more on public safety and punishment and increased the number of juveniles eligible for processing in the adult criminal court by

modifying the transfer[1] laws that determine whether the adult or juvenile system retains jurisdiction over an arrested youth (Redding, 2008; Salekin, Yff, Neumann, Leistico, & Zalot, 2002).

States have enacted several common mechanisms through which a youth can be waived out of the jurisdiction of the juvenile court and into the jurisdiction of the criminal court (Melton et al., 2007). For example, certain violent offenses are required by law to be filed in criminal court, or concurrently in the juvenile and adult systems so that the prosecutor can determine where the case will be ultimately filed. In another transfer mechanism called a judicial waiver, a hearing is conducted to determine whether transfer of the juvenile to the adult court is appropriate.

To determine whether it is appropriate to waive an adolescent defendant out of the juvenile system, the court examines a number of factors related to characteristics of the offender (e.g., the juvenile's age) and the offense (e.g., the impact of the offense on the community) (Heilbrun, Leheny, Thomas, & Huneycutt, 1997). During these hearings, clinicians may be asked to provide the court with information and opinions about other questions, such as the offender's treatment and rehabilitation needs. One of the most important factors that falls under the domain of the evaluating clinician is the adolescent's risk for future violent behavior (Heilbrun, Marczyk, & DeMatteo, 2002).

1.2 Risk Assessment

Psychologists and other mental health professionals involved in transfer proceedings frequently conduct violence risk assessments to determine a youth's risk for future violent behavior. Clinicians are currently assisted in this task by best practice

[1] Transfer is also referred to as certification and waiver in different states.

literature (e.g., Hoge & Andrews, 2010) and several validated risk assessment instruments for adolescent offenders. Prior to the development of these tools, however, juvenile risk assessments were often inconsistent or inaccurate. Causes of these weaknesses in juvenile risk prediction included the use of unstructured assessment methods, the reliance on adult risk factors when evaluating young offenders, and the lack of adequate training in the assessors.

1.2.1 Unstructured clinical judgment.

Mental health professionals have made significant developments in the science of risk assessment over the last three decades (Heilbrun, Yasuhara, & Shah, 2010). However, prior to these advances, evaluations of future risk, for both juveniles and adults, were often based on unstructured decision-making. Unstructured decision-making is a flexible decision-making process that is not defined by any one formula or measure. Instead, when predicting an outcome of future risk, the evaluator relies on clinical skills and knowledge, combining whatever factors that he or she considers important "in his or her head" (Dawes, Faust, & Meehl, 1989, p. 1668). However, research has identified unstructured clinical risk assessments to be prone to human error, subjectively biased, and of poor predictive accuracy (Monahan, 1981). Best practice guidelines now suggest that risk assessments include a thorough and methodical inquiry of criminogenic risks, criminogenic needs, and responsivity characteristics. Unstructured risk assessments have been criticized for not systematically incorporating and appropriately weighing these factors (Andrews, Bonta, & Hoge, 1990; Hoge, 2002; Towberman, 1992).

There are additional factors specific to juvenile evaluations that limit the predictive utility of unstructured risk assessments. Research has identified unique risk

factors associated with juvenile violence that, when not incorporated into the risk assessment, limit the clinician's predictive accuracy (DeMatteo & Marczyk, 2005). Clinicians were found to be lacking the relevant education and training in these particular risk factors and, as a result, not integrating them into their unstructured judgments, leading to inconsistent evaluations (Marczyk, Heilbrun, Lander, & DeMatteo, 2003). In addition, evaluators who rely on factors more associated with adult violence than those that have a demonstrated relationship with juvenile violence may draw erroneous conclusions on future risk (Borum, 1996). Due to the inaccuracy of unstructured risk assessment techniques and the importance of integrating juvenile-specific risk factors, researchers began to focus on developing structured assessments, and then extending those measures to juveniles (Borum, 2000).

1.2.2 Subsequent developments in risk assessment.

Despite the documented limitations in unstructured violence risk prediction, courts continued to request evaluations of future risk from mental health professionals (*Barefoot v. Estelle*, 1983; Monahan, 1981). To reconcile the demand by the courts for violence risk assessments with the poor predictive utility of available risk assessment techniques, researchers set out to empirically identify methods that would increase the predictive power of identifying those who were at a higher risk of criminal behavior. Significant advances in the understanding and practice of risk assessment have resulted from the development of a number of measures that utilize actuarial and structured professional judgment approaches to risk assessment.

Actuarial assessments utilize mechanistic formulas based on the presence of certain predictor variables, which are identified through research as being associated with

known risk outcomes, to determine the likelihood of future risk (Litwack, 2001). In contrast to decisions based on unstructured reasoning, actuarial outcomes are fixed and rule-bound (Heilbrun et al., 2010). Although actuarial measures have consistently been found to be more accurate than unstructured clinical decision-making (Grove, Zald, Lebow, Snitz, & Nelson, 2000; Mossman, 1994), they have also been criticized on a number of grounds (Slobogin, 2006). One major limitation of actuarial methods is their inability to capture factors that may have contributed to an offense, but which were not included as items in the assessment by the test developers. Another significant drawback is that actuarial instruments typically do not incorporate dynamic variables that may change after the point of the assessment.

In contrast to formula-based actuarial measures, structured professional judgment instruments require clinical expertise and experience in weighing a checklist of factors, each of which has been identified in the literature as being associated with violent behavior, to produce an overall risk rating. Structured professional judgment tools are typically designed to help clinicians focus on preventing or managing, rather than predicting, the occurrence of future violence through the integration of dynamic risk factors that might indicate a reduction or increase in the stability of one's risk (Vincent, 2006). The inclusion of dynamic risk factors emphasizes an ongoing process to risk assessment as opposed to an effort at making a single yes-or-no decision (Douglas & Kropp, 2002).

Disagreement remains as to whether actuarial or structured professional judgment models offer a predictive advantage (Webster, Hucker, & Bloom, 2002). Both actuarial and structured professional judgment approaches to risk assessment involve the use of a

set of predetermined risk factors that have been identified in the literature as being related to a future behavior. These approaches differ, however, in how the clinician makes a final summary judgment. In the actuarial approach, the instrument provides a risk rating based on an algorithm designed by the test developers. In contrast, a clinician using a structured professional judgment instrument will determine the final judgment through the evaluation and weighing of the risk factors using his or her skills and experience. Although individual studies differ on which method offers a predictive advantage (e.g., de Vogel, de Ruiter, van Beek, & Mead, 2004; Douglas, Yeomans, & Boer, 2005), a recent survey of the existing research comparing actuarial and structured professional judgment outcomes found that both methods appear to be, at present, comparable in their predictive accuracy (Heilbrun et al., 2010).

1.3 Juvenile Risk Assessment

Actuarial and structured professional judgment methods led to an improvement in violence risk classification, but were largely developed for adult populations. These adult assessment instruments were incompatible with juvenile offenders because they did not account for the risk factors that had been demonstrated to have an empirical relationship to violence in youth (Borum, 2000). The need for juvenile risk assessment instruments has led to the development over the last decade of several well-validated instruments that measure various levels of violent behavior.

1.3.1 Risk factors for juvenile violence.

There is a large amount of research on the unique risk factors that predict both violent and non-violent antisocial behavior among juveniles. For example, Cottle, Lee, and Heilbrun (2001) examined 23 studies (N = 15,265 juveniles) on juvenile recidivism

and found that offense history was the strongest predictor of general reoffending. They also concluded that family problems, ineffective use of leisure time, delinquent peers, conduct problems, and non-severe psychopathology were strong predictors of juvenile reoffending. Lipsey and Derzon (1998) found that prior antisocial behavior is an important risk factor as well, but that relationships with antisocial peers and a lack of prosocial connections were also important predictors. In a review of significant risk factors of juvenile violence, DeMatteo and Marczyk (2005) also identified early aggressive behavior, antisocial beliefs and attitudes, physically aggressive parents, frequent school transitions, impoverished home environments, and frequent exposure to violence as important contributors to later aggressive and violent behavior.

Research has also found that predictors of violent behavior vary according to developmental stages. According to a meta-analysis of longitudinal studies examining factors that predicted violent behavior, risk factors change as a child grows into adolescence (Lipsey & Derzon, 1998). The best predictor of future violence among youth between the ages of 6 and 11 years was having a general offense and substance use, but by the time youth are between the ages of 12 and 14, the strongest predictors of later violence are factors related to social relationships. Similarly, for these latter youth, substance use also drops from a first-order predictor to a fifth-order predictor. Research has also found that individual and family risk factors are important during childhood, but that environmental factors, such as peer influences and school-related problems, become more important in adolescence (Howell, 1997).

In addition to unique risk factors that vary across the span of childhood and adolescence, there are other risk-related aspects to juvenile development. For example,

the likelihood of engaging in delinquent behavior is at its highest levels during adolescence, therefore base rates of adolescent antisocial behavior should be considered in juvenile risk assessment (Borum, 2000; Moffit, 1993). Another important aspect of juvenile risk assessment relates to the psychosocial immaturity and decision-making capabilities of adolescents (Grisso, 1996). Adolescents are by definition in a period of growth across all domains of functioning and cannot therefore be characterized by a single observation. Risk assessments that do not incorporate dynamic variables are therefore less likely to be able to differentiate between the fluctuating risk levels presented by adolescents.

1.3.2 Juvenile risk assessment measures.

Reliable risk assessment measures that incorporated juvenile risk factors and developmental considerations were generally unavailable until recently. Although a significant body of research on characteristics that raise juveniles' risk for violence had amassed in the literature, few attempts at focusing this material into a usable instrument had been attempted (Lodewijks, Doreleijers, de Ruiter, & Borum, 2008b). However, over the past decade, considerable advances have been made in the practice and research of juvenile risk assessment (Grisso, Vincent, & Seagrave, 2005; Heilbrun, Goldstein, & Redding, 2002). During this period, risk assessment instruments have been developed that incorporate the unique risk factors and important developmental considerations presented by this population. These instruments have led to more valid and consistent risk judgments than achieved previously (Hoge, 2002). Previously used techniques, such as unstructured judgments, have been replaced with what are now considered more valid and, some argue, essential measures (Gacono, 2000). One such measure, which

examines the risk of general recidivism among youth aged 12 to 17, is the Youth Level of Service/Case Management Inventory (YLS/CMI; Hoge & Andrews, 2002). The YLS/CMI is based on a similar risk prediction instrument for adults, the Level of Service Inventory-Revised (LSI-R; Andrews & Bonta, 1995; now the Level of Service/Case Management Inventory, or LS/CMI; Andrews, Bonta, & Wormith, 2004). The YLS/CMI is a checklist of 42 criminogenic risk and need factors organized into eight domains: prior/current offenses/dispositions, family circumstances/parenting, education/ employment, peer relations, substance abuse, leisure/recreation, personality/behavior, and attitudes/orientation. Similar to its actuarial parent instrument, individual items on the YLS/CMI are summed to provide an overall score that falls into one of several risk summary ranges. The YLS/CMI is designed to differentiate high risk youth from lower risk youth under the premise that appropriate interventions with such youth can be effective in reducing recidivism (Catchpole & Gretton, 2003).

An accumulating body of literature indicates that the YLS/CMI has excellent predictive validity for juvenile reoffending and sound reliability over both short and longer time periods (Schmidt, Hoge, & Gomes, 2005; Welsh, Schmidt, McKinnon, Chattha, & Meyers, 2008). A recent meta-analysis by Olver, Stockdale, and Wormith (2009) of 22 studies incorporating the YLS/CMI or another youth version of the LSI found that the LSI instruments performed better in predictions of general recidivism (r_w = .32) than in violent recidivism (r_w = .26).

Another assessment that has been found to have use as a predictor of recidivism is the Psychopathy Checklist: Youth Version (PCL:VY; Forth, Kosson, & Hare, 2003). Like the YLS/CMI, the PCL:YV is modeled on a previously developed adult instrument,

the Psychopathy Checklist Revised (PCL-R; Hare, 1991, 2003). The PCL:YV is scored by rating 20 items related to behavioral and personality factors, with a total score ranging from 0 to 40.

Although higher PCL:YV scores have similar predictive capabilities as other juvenile risk assessment measures (Olver et al., 2009), considerable differences set this instrument apart. First, the PCL:YV was designed to measure psychopathy, not to be an assessment for future antisocial behavior. Although psychopathy is reliably associated with future antisocial behavior, the PCL measures are not risk assessment measures per se. Second, there is an active debate about the construct of adolescent psychopathy tapped by the PCL:YV (Lynam, 2002; Skeem & Petrilla, 2004). Disagreement remains among mental health professionals regarding the developmental pathways of psychopathy and whether it is a form of pathology that can accurately assessed during adolescence (Salekin & Frick, 2005). Finally, there is concern that the negative connotations associated with "psychopath" make the label too damaging to be used with developmentally immature juveniles (Seagrave & Grisso, 2002). Research also indicates that juveniles with this label may receive more negative consequences than juveniles not identified as a psychopath, such as the imposition of more punitive court sanctions (Boccaccini, Murrie, Clark, & Cornell, 2008).

Despite concerns about the use of a psychopathy measure with youth, assessments with the PCL:YV in juvenile legal proceedings have only increased since the instrument's development (Viljoen, MacDougall, Gagnon, & Douglas, 2010). Further adding to its use in juvenile justice settings is a body of research demonstrating it to be comparable to other risk assessment instruments in predicting both general and violent

recidivism (Edens, Campbell, & Weir, 2007). Psychopathic characteristics in juveniles have been linked to both delinquent and violent behavior; a recent meta-analysis of 28 studies found an r_w of .18 for general recidivism and .25 for violent recidivism (Olver et al., 2009). Another meta-analysis of 21 studies using the PCL:YV or the PCL modified for use with adolescents found that psychopathy was significantly associated with both general and violent recidivism, with weighted mean correlation coefficients of .24 (n = 2,787) and .25 (n = 2,067), respectively (Edens et al., 2007).

Other juvenile violence risk assessment instruments have also been developed and validated. For example, the Early Assessment Risk List for Boys (EARL-20B; Augimeri, Webster, Koegl, & Levene, 2001) is a 20-item measure that assesses risk of violence in males under the age of 12 years. Similar to the development of the YLS/CMI and PCL:YV, the EARL-20B was modeled on an adult risk assessment measure, the Historical Clinical Risk 20 (HCR-20; Webster, Douglas, Eaves, & Hart, 1997), and utilizes items that have an associated relationship with general recidivism in young males. Research suggests that the EARL-20B is a useful tool in the assessment of antisocial behavior among adolescent males (Enebrink, Långström, & Gumpert, 2006). Psychometric properties have not yet been established for the Early Assessment Risk List for Girls (EARL-21G; Levene et al., 2001), which is modeled similarly to the EARL-20B but based on known risk factors for female adolescents, such as maternal caregiver and daughter interactions and sexual development (Odgers, Moretti, & Repucci, 2005).

1.4 The SAVRY

The YLS/CMI, PCL:YV, and other juvenile instruments have contributed to significant improvements in the accuracy of violence risk predictions with youth.

Another juvenile risk measure that has accumulated a strong base of support is the Structured Assessment of Violence Risk in Youth (SAVRY; Borum, Bartell, & Forth, 2003). The SAVRY is a violence risk assessment instrument modeled after a measure normed on adults (i.e., HCR-20), but consisting of items associated with risk for future violence in adolescents. The SAVRY has several strengths as a measure of juvenile violence risk. First, it integrates dynamic risk factors, which are an integral part of juvenile risk assessment because of their importance in identifying aspects of a youth's life that may change and affect risk potential. Second, the SAVRY is modeled under the structured professional judgment approach to risk assessment. Although the risk assessment literature indicates an overall equal level of accuracy across actuarial and structured professional judgment approaches with adults, research solely with juveniles suggests that structured professional judgment instruments may be most accurate in the prediction of future violence (Catchpole & Gretton, 2003). Third, the SAVRY has extensive information on its predictive and incremental validity, properties which not all measures developed over the last 10 years possess (Hannah-Moffat & Maurutto, 2003; Skeem & Cauffman, 2003).

Prior to the SAVRY's development, the instrument's developers were initially working on separate violence risk assessment measures, the Youth Risk Checklist, by Randy Borum, and the Adolescent Violence Risk Assessment, by Patrick Bartell and Adelle Forth (Borum, Bartell, & Forth, 2005). Upon learning of each project, the researchers joined together and the items from the two assessments were pooled and evaluated for fitness in the SAVRY (Bartel, Borum, & Forth, 2000). The SAVRY

initially was released in a consultation edition, and has since undergone other minor re-editions (Borum, Bartell, & Forth, 2001, 2002, 2003).

1.4.1 Standard SAVRY coding procedure.

The SAVRY consists of 24 risk factors and 6 protective factors that have a demonstrated relationship with future violent behavior (see Appendix A). The risk factors are categorized into three categories. The first category, Historical Risk Factors (10 items), consists of generally static items based on past behavior or experiences (e.g., "Exposure to Violence in the Home"). The second category, Social/Contextual Risk Factors (6 items), includes items that describe the youth's relationships to other people, other institutions, and his or her environment (e.g., "Community Disorganization"). The third category is Individual/Clinical Risk Factors (8 items), which is composed of items that examine attitudes as well as psychological and behavioral variables (e.g., "Anger Management Problems"). The six protective factors assess individual and contextual elements that may reduce the likelihood of future violence. An example of a SAVRY protective factor is "Strong Commitment to School."

Each risk factor on the SAVRY is scored as "Low," "Moderate," or "High." The SAVRY manual (Borum et al., 2003) indicates that a risk factor should be scored "Low" when the risk factor is absent, "Moderate" when the risk factor is minimally present and/or is the cause of minor impairment in functioning, and "High" when the risk factor is prominent in the youth's history or present life and causes significant impairment. The six protective factors are coded as "Present" or "Absent." The manual provides item-specific guidance for rating each risk and protective factor. In addition to the rating,

each item on the SAVRY can be denoted as a "Critical Item" if the clinician feels it is especially significant in understanding the level of risk presented by the juvenile.

The SAVRY was designed to provide clinicians with a structured method of examining the important risk and protective factors to allow for a professional judgment of a youth's risk for future violence. According to the manual, clinicians scoring the SAVRY should not attempt to create a final numerical score by adding the number of existing risk items and subtracting the relevant protective items. Instead, clinicians are advised to make a final Summary Risk Rating of either "Low," "Moderate," or "High" based upon the clinical assessment of the various risk and protective factors present in each juvenile's case.

After the publication of the initial version of the SAVRY, a second version was published with one primary change resulting from the modification of an item that tapped psychopathic characteristics. In the first version of the SAVRY, the PCL:YV was required to assess Item 21, "Psychopathic Traits." According to the SAVRY manual, this item on the SAVRY was changed to "Low Empathy/Remorse" for three reasons. First, for purposes of scoring the SAVRY, the test developers were more interested in tapping the relevant traits that constituted a risk for future violence, not the construct of psychopathy. Second, requiring the score from the PCL:VY created an imbalance in user qualifications between the SAVRY, which is a risk assessment tool, and the PCL:YV, which is a diagnostic test. Finally, the SAVRY developers were concerned that the label "psychopathy" would carry an overwhelmingly negative connotation and endorsement of this item might overshadow any other information generated by the SAVRY.

1.4.2 Psychometric data on the SAVRY.

A growing body of literature supports the SAVRY as a valid and reliable measure of adolescents' risk for future violence and aggression. In examining the predictive accuracy of overall ratings, Lodewijks and colleagues (2008b) found that the SAVRY Summary Risk Ratings had excellent predictive validity for violent behavior (AUC = .86). Welsh and colleagues (2008) also found the SAVRY to be an accurate predictor of nonviolent aggression as well (AUC = .77). In examining the three risk categories, the highest predictive values have been established for the Individual/Clinical and Contextual domains (Lodewijks et al., 2008b). A recent meta-analysis of nine available studies on the SAVRY found that its predictive accuracy for both general and violent recidivism (r_w = .32 and .30, respectively) was superior to the LSI/CMI and PCL:YV (Olver et al., 2009). Similarly, when examined together with the LSI/CMI and PCL:YV, the SAVRY has been found to have strong incremental validity that improves the predictive power of both institutional aggression and serious aggressive behavior more than the other instruments alone (Borum et al., 2003).

1.4.3 Use of multiple sources in coding the SAVRY.

To complete the 30 individual item ratings on the SAVRY, the manual advises clinicians to use a multi-source approach. Obtaining information from multiple sources is an established principle of forensic mental health assessment (Heilbrun et al., 2002). As opposed to therapeutic relationships, forensic settings are often adversarial and contain incentives for exaggeration or minimization. Although self-report is an essential element of a forensic assessment, the validity of the evaluation's recommendations is improved by the use of collateral sources, including interviews with family members,

friends, or teachers, police and probation reports, school records, previous mental health evaluations, and other records. Similarly, the SAVRY authors suggest the use of multiple sources when coding the SAVRY items.

1.5 The Current Study

The SAVRY was developed as an instrument to be completed on the basis of information obtained from multiple sources, including a clinical interview and record review. However, clinical situations may arise when standard administration is not feasible. For example, a juvenile who chooses not to participate in, or is not available for, a forensic mental health evaluation may require the clinician to rely more heavily on available records. Borum and colleagues (2005) also note that certain scenarios (e.g., situations in which a rapid assessment of future risk is needed) may require a nonstandard administration of the SAVRY. Although such situations are rare, when they do occur a clinician may only be able to utilize existing records. No empirical data exist, however, to support the validity of the results obtained from such an assessment. No studies could be located to date that have examined whether SAVRY ratings based solely on records are comparable to those developed through the standard administration procedure.

Another, and perhaps more, important reason to examine the reliability of file review-based SAVRY ratings is that much of the psychometric research on the SAVRY is based on file-only SAVRY scores. File-only risk assessment ratings have often been used in the research literature as a proxy for standard administration scores, which are not always practical to obtain (Campbell, Porter, & Santor, 2004). For instance, although a researcher may wish to examine a given instrument using a prospective design that mimics a “real-world” setting, the researcher may not have resources to follow a sample

longitudinally, and the desired instrument may never have been originally administered to the sample that is available to the researcher. As a result, researchers interested in studying a given instrument often will use sample file information to "go back" and score the instrument, and then examine the validity of the instrument in measuring the outcome to which they have access. This postdictive design allows researchers to examine important psychometric properties of risk assessment instruments even when ratings based on standard administration procedures were not initially recorded (Grann, Långström, Tengström, & Stålenheim, 1998).

Research lends some support for the use of retrospective ratings in postdictive studies. For example, researchers have found similarities between file-only scores and standard administration scores on another instrument used for risk assessment, the PCL-R. The results of the PCL-R studies suggest that scores based solely on file review are as reliable as scores obtained through the use of a clinical interview, provided that the rater was provided with an adequate source of records (Grann et al., 1998; Wong, 1988). Although this research provides general support for the use of file-only retrospective scores, no studies have specifically examined such use with the SAVRY.

As Table 1 demonstrates, most of the studies on the SAVRY have utilized a retrospective design to examine the SAVRY's ability to predict future violent behavior in adolescence. Catchpole and Gretton (2003), for instance, examined the predictive validity of the instrument in a sample of violent youth. As the youth had "already been discharged from their respective facilities," the authors relied on "extensive information collected in the forensic files" to retrospectively code the SAVRY (p. 693). Meyers and Schmidt (2008) also examined the predictive validity of the SAVRY to identify youth at

risk for future violent behavior. These authors relied on "existing file information" collected on each juvenile to rate the items on the SAVRY (p. 349). Another study examined the relationship between rates of violent reoffending and SAVRY scores that were coded according to "the same file information as was available at the time" of the initial clinical referral (Lodewijks, Doreleijers, & de Ruiter, 2008a, p. 699).

Postdictive studies have helped build an empirical knowledge base for the SAVRY, but they have been conducted in the absence of research demonstrating that file-only and standard administration (i.e., utilizing clinical interviews, collaterals, and records) SAVRY ratings are comparable. Although several studies have examined the predictive validity of the SAVRY using a prospective design and standard rating administration, these studies did not obtain file-only ratings to allow for a comparison of the two coding procedures. An examination of the reliability of file review-based SAVRY ratings is therefore needed to determine if such ratings are equivalent to standard administration ratings. One cannot know whether the file-only ratings obtained in the studies cited in Table 1, even when agreed upon by multiple raters, are an acceptable substitute for SAVRY ratings based on clinical interviews and other sources. At this point, file-only SAVRY ratings have not been established as reliable. This study will assess whether file review-based ratings are reliable by examining whether such ratings are equivalent to ratings based on both an interview and file information.

1.6 Hypotheses

It was hypothesized that file review-based SAVRY risk factor ratings, protective factor ratings, and Summary Risk Ratings would be highly concordant with ratings obtained via standard administration ("expert" ratings). Research on other measures used

to assess risk (e.g., PCL-R) has found that reliable ratings can be obtained without conducting a clinical interview. Accordingly, it was expected that file-only SAVRY ratings would be similar to expert ratings.

Chapter 2: Methods

The objective of this study was to examine the reliability of file review-based SAVRY ratings. To conduct this comparison, a sample of archived decertification files was obtained from a university-based forensic clinic in Philadelphia, Pennsylvania. The juveniles in this sample had been referred to the clinic for an evaluation related to their request to be decertified from criminal court. During the decertification evaluation, each juvenile was assessed with the SAVRY according to the manual's standard scoring procedure. After the juvenile had been evaluated by the clinic, the file was archived and subsequently used in this study to retrospectively code the SAVRY by raters blind to the standard administration scores.

2.1 Setting

In Pennsylvania, adolescent offenders 14 years of age and older are criminally charged (a "direct file") if they are charged with the use of a deadly weapon while committing rape; involuntary deviate sexual intercourse; aggravated assault; robbery; robbery of a motor vehicle; aggravated indecent assault; kidnapping; voluntary manslaughter; or attempting, conspiring, or soliciting to commit any of these offenses or to commit murder (The Juvenile Act, 42 Pa. Cons. Stat. Ann. § 6355). If a juvenile offender meets these offense criteria, he or she is statutorily excluded from the jurisdiction of the juvenile court. A mechanism remains in place, however, by which the juvenile can request to be decertified into the juvenile court system. In requesting to be considered for decertification, the juvenile must prove by a preponderance of the evidence that the public interest will best be served by such a reverse transfer (Pennsylvania Juvenile Court Judges' Commission, 2008).

As the juvenile defendant is responsible for demonstrating that the public interest will best be served by decertification to juvenile court, decertification evaluation referrals to mental health professionals typically come from defense attorneys. A local university-based forensic clinic conducts such evaluations at the request of the Defender Association, private attorneys, and court-appointed attorneys. When a juvenile is referred to this clinic to be evaluated for decertification purposes, the attorney provides the evaluating clinicians with a file of available records. The clinicians review the file information, complete a clinical interview with the juvenile, administer several psychological tests to the juvenile (e.g., the SAVRY), and conduct collateral interviews with family members and other persons familiar with the juvenile's background. The evaluation is aimed at better understanding the behaviors and capacities underlying the juvenile offender's risk to public safety, treatment needs, and amenability to treatment (Heilbrun et al., 2002). A psychological report is then produced by the evaluating clinicians and provided to the referring attorney.

2.2 Participants

This study sampled a continuous number of juvenile decertification referrals to the forensic clinic since its adoption of the SAVRY as a risk assessment instrument in 2008 (K. Heilbrun, personal communication, August 17, 2010) to July 2011, the date when file review coding of the SAVRY began. During this time period, 38 juveniles were referred to the clinic. In addition, two other juveniles were referred to the clinic for reasons other than decertification[2] and were included in the study sample because they

[2] One juvenile was referred for an evaluation to determine treatment needs and amenability in the context of state sentencing, and the other juvenile was referred for an evaluation to determine treatment needs and to identify treatment programs that could best address those needs.

were assessed with the SAVRY to determine risk for future violence in the context of their mental health evaluations. Of these 40 juveniles, 2 were female and were excluded from the sample to eliminate variability caused by the potential differences in risk assessment factors between males and females (Odgers et al., 2005). Two additional juvenile files were excluded because the files did not contain the scored expert SAVRY assessment and thus could not contribute to the examination of the reliability of file-only ratings. Lastly, one other file was excluded because the expert SAVRY evaluation was missing information (i.e., lacked a Summary Risk Rating). This resulted in a final sample of 35.

Table 2 provides demographic information on the sample. The youth were primarily African American (n = 26) and Hispanic (n = 7), with fewer numbers of Caucasian (n = 1) and "other" (n = 1). The sample's mean age was 16.69 years (SD = .796, range = 15 to 18). At the time of their arrests for the index offenses, 37.1% (n = 13) of the youth engaged in daily drug or alcohol use and another 22.9% (n = 8) engaged in weekly or monthly drug or alcohol use (14.3% and 8.6%, respectively). The mean number of prior arrests in this sample was 1.97 (SD = 2.14), with a range between 0 and 9 arrests. The majority of the sample was referred to the clinic by a private or court appointed attorney (n = 21), and the remaining juveniles (n = 14) were represented by the Public Defender. Independent sample t tests revealed no significant differences in age, level of pre-offense drug use, or number of prior arrests (ps > .05) between the two referral groups.

2.3 Procedure

2.3.1 De-identification of files.

During the original decertification evaluation, the juvenile was administered the SAVRY to measure his risk of future violent behavior. These expert SAVRY ratings were based on a review of the file information, an interview with the defendant, and other sources (e.g., collateral interviews). The juvenile's file, including the assessment ratings, was then archived in the university clinic's storage. Upon selection for use in this study, the files were de-identified by the author of this study and clinic staff members. All data that were collected during the evaluation, including assessments (e.g., the SAVRY), clinical interview notes, collateral interview notes, and other material not provided by the attorney at the time of the original referral, were also removed from the file. Additionally, any clinical findings and opinions from the evaluation, including copies of reports sent to the attorney, were removed from the file.

2.3.2 File contents.

The files provided to the clinic from the referring attorneys ranged in volume from 1 document to 58 documents, with a mean of 21.63 (*SD* = 16.05) documents across all files. The variability in file volume and content is influenced by a number of factors, including the complexity of the case, the availability of records at the time of referral, and the criminal and mental health histories of the juveniles. Although there was not uniformity across files, they all typically contained one or more documents from the following categories: psychosocial information, reports from previous mental health evaluations, medical and mental health records, academic documents, police investigation

materials, and court transcripts. An independent sample *t* test revealed no differences in the number of documents per file between the two referral sources ($p > .05$).

2.3.3 Outcome measure.

As described above, the SAVRY is a violence risk assessment instrument based on the structured professional judgment model. File-only raters scored the SAVRY using the same structured professional judgment approach as the expert clinicians, but they were only given access to the de-identified files originally provided by the referring attorneys. Raters were instructed to mark any items lacking sufficient information to reliably code as "Unable to Rate." In addition, file-only raters were not asked to indicate Critical Items, but they were asked to document any important reasoning behind their item rankings.

2.3.4 Training workshop.

The raters in this study were five individuals involved in a psychology and law graduate program or program-affiliated research lab. The SAVRY manual indicates that users of the instrument "should have expertise (i.e., knowledge, training, experience) in conducting individual assessments, in child/adolescent development, and in youth violence" (Borum et al., 2003, p. 12). The manual authors suggest that psychologists, psychiatrists, probation officers, and social workers with the necessary training are qualified to use the SAVRY. To ensure that the five raters had the required expertise to code the SAVRY, they participated in a two-part training workshop prior to conducting any coding. In the first portion of training, the raters learned about the decertification process, the importance of a thorough risk assessment, and the use of structured professional judgment models in general and the SAVRY in particular. Raters were also

taught about the various types of file information contained in decertification files. This 1-hour workshop was taught by this author, who has supervised training and experience in conducting forensic evaluations with adolescents, including decertification evaluations in which the SAVRY was administered. Next, the raters participated in another 1-hour workshop on the SAVRY led by a senior clinical psychologist who is board-certified in both forensic and clinical psychology. Raters learned about the instrument's psychometric properties, were instructed on how to apply the file information to the various SAVRY items, and conducted a supervised simulated rating of the SAVRY using a mock decertification file. Raters were given access to the SAVRY manual for the duration of the study and were also provided a video recording of the training workshop for their reference.

2.3.5 Practice files.

In addition to participating in the training workshop, the raters were provided an opportunity to practice coding the SAVRY with five practice files that were distinct from the files used in the main analysis. The practice files were selected from a group of files in which a SAVRY was not administered at the time of the original evaluation. The practice phase was designed to allow the raters to gain familiarity with the different types of file information and to establish interrater agreement. To achieve that purpose, thorough and extensive files were selected for this practice phase. The practice files were de-identified and cleared of any information collected during the original evaluation (as described in section 2.3.1). Although these practice files allowed a determination of the level of agreement between file-only raters, no pre-study check could be made of the

level of agreement between expert and file-only raters because the practice files did not contain an expert SAVRY assessment.

As the raters completed each practice file, they were provided follow-up group and individual feedback sessions by this author to address interrater discrepancies. Items that caused interrater disagreement were discussed in relation to the manual's instructions. For example, initial practice ratings on item 9, "Early Caregiver Disruption," did not exhibit a uniform cut-off age for what constituted a youth's "early" life. Raters were then pointed to the instructions in the manual that state that the item refers to the period from birth to 12 years of age. Many other rating situations arose, however, that were not addressed by the SAVRY manual. In response, a set of decision rules was established to guide the raters in these conflicting coding decisions. For example, raters demonstrated an understanding that item 16, "Community Disorganization," assessed the levels of crime, poverty, and violence in the juvenile's neighborhood. Despite this understanding, raters initially showed high disagreement when coding this item. Some raters believed they could make a strong assumption about the community in which the juvenile lived based upon indirect information in the file (e.g., where the crime took place); other raters argued that unless the file specifically mentioned the youth's area of residence, one could not know where the juvenile lived or what the neighborhood was like. Decision rules provided the necessary guidance for the raters when they came across problematic coding situations such as this one (see Table 3 for the decision rule for this item and for a sample of other decision rules).

Chapter 3: Results

Interrater reliability was assessed using intraclass correlation coefficients (ICC; Shrout & Fleiss, 1979). ICCs provide a measure of the reliability of the coding judgments made by the different raters. According to Fleiss (1986), ICC values above .75 are excellent, ICCs between .60 and .75 are good, ICCs between .40 and .60 are moderate, and ICCs below .40 are poor. ICCs were calculated using a two-way mixed, absolute agreement model. Frequency analyses were used to determine the number and percentage of each type of rating assigned to individual items and Summary Risk Ratings.

3.1 Practice Files

File-only raters were able to provide ratings for an average of 80.6% of the individual items and coded the remaining 20.4% of individual items as "Unable to Rate." Individual item completion rates ranged between 87.7% and 74.2% across the five raters. The file-only raters also provided an almost complete set of Summary Risk Ratings (94%). One rater coded one file's Summary Risk Rating as "Unable to Rate." As noted in the Methods, raters initially had difficulty providing uniform ratings in the practice phase until a set of decision rules were adopted. Interrater agreement increased as raters gained familiarity with coding the files, and at the conclusion of the practice phase raters demonstrated an excellent level of reliability (.914).

3.2 SAVRY Summary Risk Ratings

During their original decertification evaluations at the forensic clinic, approximately half of the sample (n = 18) was identified by the expert clinicians as a moderate risk for future violence. Of the other half of participants, 37.1% (n = 13) were

classified as a low risk and 11.4% (*n* = 4) were identified as a high risk of future violent behavior by the expert clinicians. In contrast, file-only raters classified a quarter of the sample as a moderate risk for future violence (*n* = 9). Of the remainder of the sample, 14.3% (*n* = 5) were determined to be a low risk, 17.1% (*n* = 6) were categorized as a high risk, and 42.9% (*n* = 15) were coded as "Unable to Rate." The percentage of each summary risk level coded by expert and file-only raters is presented in Table 4.

3.2.1 Concordance of file-only ratings with expert ratings.

The interrater reliability of the expert and file-only SAVRY Summary Risk Ratings was in the moderate range (.466). This value is based on the 20 files in which a comparison was possible between expert and file scores (i.e., the files in which the file-only raters did not code the Summary Risk Rating as "Unable to Rate"). In these files, there were higher rates of overlap between expert and file-only raters when expert raters coded the juvenile defendants as a low or moderate risk for future violence than when expert raters categorized the juveniles as a high risk for future violence. Juveniles categorized as a low risk by expert raters were correctly classified as a low risk by file-only raters in 60% of the files (3 out of 5 files). The remaining low-risk files (*n* = 2, or 40%) were determined to be a moderate risk by the file-only raters. Of the 9 juveniles categorized as a moderate risk by expert raters, 6 (67%) were also categorized as a moderate risk by the file-only raters. The remaining 3 moderate risk files were categorized as either a low risk (*n* = 2, or 22%) or high risk (*n* = 1, or 11%) by file-only raters. In contrast to low and moderate Summary Risk Ratings, file-only raters had a very low level of agreement with expert raters when the expert raters determined a juvenile to be a high risk of future violence. In these 6 high risk files, file-only raters also

categorized the juvenile as high risk in only 1 instance (17%); the remaining 5 files (83%) were categorized as a moderate risk by file-only raters.

3.3 Individual SAVRY Risk and Protective Factors Ratings

Tables 5 and 6 list the percentages of both expert and file-only ratings on each risk level of the SAVRY individual items (items 1-30). Similar to the Summary Risk Ratings, there were high rates of missing data in the individual items rated by file-only raters; these raters indicated they were unable to rate 53.3% of the individual SAVRY items. Items coded most often as "Unable to Rate" by file-only raters were items 6, "Exposure to Violence in the Home" (85.7% of files), and 12, "Peer Rejection" (77.1% of files). No file was completely rated on all 30 individual items by file-only raters; files ranged between 5 (16.7%) and 29 (96.7%) of the items coded as "Unable to Rate," with a mean of 15.9 items unrated.

In contrast to the file-only raters, expert raters left few items unrated (n = 27, or 2.5% of all individual items). Of these 27 missing items, 29.6% (n = 8) were item 4, "Past Supervision/Intervention Failure," left both unrated and noted by the expert rater as "Not Applicable." The other 70.4% (n = 19) of unrated items were left unrated with no notation as to the reason. Item 16, "Community Disorganization," was the most common item in this group of unrated items that lacked explanative notation (n = 8).

3.3.1 Concordance of file-only ratings with expert ratings.

The agreement between file-only and expert SAVRY risk factor and protective factor ratings was in the poor to moderate range. Individual risk items (items 1-24) on the SAVRY had a mean ICC of .344 and individual protective items (items 25-30) had a

mean ICC of .321. When all expert and file-only individual items (items 1-30) were examined together, the ICC value was moderate (.404).

3.4 Performance of File-Only Raters

To examine the performance of the individual file-only raters, each rater's level of agreement with the expert raters was also assessed. Raters were individually examined according to their ratings on the 30 individual items, the Summary Risk Ratings, and the individual items and Summary Risk Ratings combined. Table 7 presents the ICC values for each of these rating areas by the five raters, which ranged from poor to excellent. Although a statistical test to examine differences between the raters' ICCs was not possible ("Compare Intraclass Correlations," n.d.), the data reveal that raters demonstrated the greatest variability in their overall summary judgments. Each rater was not able to provide a Summary Risk Rating for every file they were assigned, but raters varied between poor and excellent in the files for which they did provide a summary rating. Also of note is the observation that only one rater achieved an ICC of .5 or above across all three ratings areas.

The interrater reliability between the five file-only raters was not assessed during the study. Conducting checks of interrater reliability throughout the study would have allowed for an examination of how reliably file-only raters coded the SAVRY as compared to each other, as well allowed for a determination of whether any interrater drift occurred. However, these concerns were negated by the high level of interrater agreement observed during practice coding and the pilot study-nature of this project, which is focused on the comparison between file-only and expert raters, and not on the agreement across different file-only raters. In addition, assessing for interrater agreement

throughout the study would have reduced the already small number of files available for examination. To assess interrater reliability, a subsample of the 35 files would have needed to be rated by each of the 5 raters, and as a result, that subsample of files would not be available for a comparison between the expert's ratings and a single set of file-only ratings.

Chapter 4: Discussion

The purpose of this study was to empirically examine whether the SAVRY can be reliably scored solely on the basis of file information. Much of the previous research on the SAVRY has been based on the use of retrospective file-only ratings, but limited data exist to support such an approach with the SAVRY. Using expert clinical ratings obtained via standard administration as the comparison, this study examined the ability of file-only raters to reliably and accurately code the SAVRY. Given past results indicating the reliability of file-only PCL-R scores, it was hypothesized that file-only SAVRY ratings would be congruent with expert SAVRY ratings.

The file-only raters began the study by coding a set of practice files. Although they were not able to rate all individual items, the raters were able to code between approximately 75 – 90% of these items. The raters also provided a near-complete set of Summary Risk Ratings and demonstrated a high level of interrater agreement. In the study's main analysis, however, file-only raters indicated that they were unable to rate more than half of the SAVRY individual items and that they could not determine a Summary Risk Rating for 43% of the files. Indeed, of the 35 files examined in this study, not one was completely rated on all 30 items. Of the items file-only raters felt they could code, their level of accuracy, as compared to expert ratings based on standard administration of the SAVRY, was only in the low to moderate range.

These results are not consistent with previous research that supports the general use of risk assessment file-based scoring. The risk assessment literature holds that file-only ratings may be used in place of ratings based on standard administration when such scores are not available (Campbell et al., 2004). Studies with the PCL-R have also found

that file-based ratings are similar to ratings obtained through the use of both a clinical interview and file review (Grann et al., 1998; Wong, 1988). In this study, however, file-only raters were often unable to fully rate the instrument, and ratings that were obtained did not accurately reflect the defendants' level of risk as measured by expert ratings based on standard administration of the instrument.

These results also conflict with prior research with the SAVRY specifically in which file-only raters were able to successfully obtain complete sets of SAVRY ratings. In comparison to the percentage of file-only Summary Risk Ratings coded as "Unable to Rate" in this study (43% of files), raters in previous file-only SAVRY research had much lower rates of incomplete overall SAVRY ratings. For example, the percentage of Summary Risk Ratings that could not be coded due to incomplete file information in a sample of 3 previous studies ranged from 2% to 10%: Meyers and Schmidt (2008) could not use 3 of 133 files due to limited file information, Lodewijks and colleagues (2008a) found that 5 of 130 files could not be rated due to inadequate file information, and Catchpole and Gretton (2003) reported that 8 of 74 files lacked adequate information to code the SAVRY.

4.1 Causes of Missing Data

Why were the file-only raters in this study substantially less often able to rate the SAVRY than previous file-only researchers? According to the comments provided by the raters in which they documented their reasoning when scoring the files, the higher levels of missing data in this study were due to the lack of sufficient data in the files. The file-only raters reported, in effect, that due to the lack of information in the files, they could not effectively use the structured professional judgment approach to risk

assessment. Raters were unable to form an overall Summary Risk Rating because they did not have enough information to provide ratings for each or even most of the items. File-only raters attested to this lack of data in these representative comments:

- "The file only has information around the current offense…not enough information to determine amount or degree for ratings."
- "Limited information. Not enough background to score many items…or to produce a confident summary risk rating."
- "No background information available. The most that is known is that the individual was allegedly present for a simple assault."
- "This file contains only police records. It did not contain the kind of info needed to complete this assessment."
- "File was extremely limited: no information regarding the history of the individual (mental health, violence, substance use, prior offending, etc.) was provided. As a result, I am unable to assign a summary risk rating."

The lack of file information did not affect raters' ability to score every item. Significant amounts of data are not necessary for all SAVRY items; some items, such as item 1, "History of Violence," are relatively clear and straightforward to rate. This item requires the rater to calculate the number of prior times the youth has been previously arrested. Once basic decision rules were established (e.g., "Do not count arrests that resulted in charges that were withdrawn."), this item was typically answered across files (and coded as "Unable to Rate" in only 3 of the 35 files).

However, several other items were more difficult to rate given the limited file information. Due to the lack of sufficient file information, many SAVRY items were

coded as "Unable to Rate" because raters did not have evidence in either support or disconfirmation of the item. For example, item 6, "Exposure to Violence in the Home," required the rater to determine whether the youth witnessed or experienced physical aggression or violence. Raters indicated that although files sometimes noted if such violence was present, files never indicated when it was not present. The following statements are some of the comments provided by raters attempting to code this item:

- "There are some allusions to abusive behaviors/neglect, however this seems to be speculation based on the fact that both of defendant's parents were drug addicts and it is unclear whether or not there is further evidence to this effect."
- "Step-father was 'mentally abusive' to mother; unclear whether physical abuse occurred and whether defendant witnessed it."
- "No discussion of defendant's home life."

The lack of definitive information about exposure to violence in the home makes intuitive sense; if such information was available, a writer might document it, but unless prompted by a checklist such as the SAVRY, the writer would have no reason to document that such a characteristic was *not* present. To determine the rating for this item, an evaluator would need to ask the youth about this factor in a clinical interview. When an in-person interview was not possible and the file did not speak to the item, the evaluator would likely have no way to ascertain the answer to this item and, as a result, would code it as "Unable to Rate." Evidence lacking in either direction caused rating difficulties for raters on many items.

Why did the files used in this study lack sufficient information? One possible cause may be related to the files themselves – that is, the files of juveniles seeking

decertification from the criminal court. Decertification files may be different from files used in previous research (e.g., files maintained in residential treatment facilities) that limit their usefulness when coding the SAVRY. Juveniles undergoing decertification evaluations may not have accumulated an adequate amount of relevant historical documentation; for example, a 15-year-old may not have certain sources of information on record, such as probation reports or mental health documents, which would allow an evaluator to reliably rate the SAVRY. The limited file information may also be related to attorney variables; it is possible, for example, that attorneys representing juveniles for decertification purposes may only have access to limited sources of information, or only have limited means to obtain records.

4.2 Lack of Congruency between File-Only and Expert Raters

The second major finding in this study was the low level of agreement between file-only and expert raters. Initially, based on the level of agreement between file-only raters during the practice phase, SAVRY ratings appeared to be highly reliable. However, the finding that file-only ratings were dissimilar to expert ratings suggests that file-only ratings may be reliable with each other, but not with ratings based on standard administration. Given the assumption in this study that the expert ratings were clinically accurate, the file-only ratings may have been reliable with one another, but not a valid reflection of the youths' actual risk.

These results do not align with previous research upon which the hypothesis for this study was based that found instruments used for risk assessment purposes could be reliably coded regardless of whether the rater used file-only information or standard administration procedures. Similar to the cause of the high number of missing items, one

possible cause of the lack of congruency between file-only and expert ratings was the lack of sufficient information in the file. Earlier reliability studies with the PCL-R assessed adults who likely had amassed a thicker set of historical records than a juvenile who was newer to the legal and mental health systems. Again, the lack of accuracy in ratings made with an incomplete picture of the juvenile's presenting risk makes intuitive sense. Whereas file-only assessments with adults may be accurate given the larger documented historical record, juvenile risk assessments may require a clinical interview to gain enough information on the juvenile's risk level. These findings may suggest that adult instruments can be reliably rated with files, but juvenile measures, such as the SAVRY, cannot.

Based on the rater feedback, another cause of the low level of agreement between raters and experts was the interrater variability that arose when raters were presented with ambiguous file information. Raters found that when applying the SAVRY items to the myriad backgrounds presented by the juveniles under evaluation, several items allowed for wide latitude in determining a rating. Although the SAVRY items may serve as operational definitions of established risk and protective factors, some items were not as clearly defined as others.

Several steps were taken to ensure that variability in rater judgments was minimized and that raters were coding the items in a manner consistent with the test developers' intentions. First, raters were selected for this study that possessed a strong clinical and research background in forensic psychology. Next, the raters participated in a training workshop and completed a series of practice files. The raters were provided substantial feedback throughout the practice phase using the manual's instructions and a

set of decision rules that provided guidance for situations not directly addressed by the manual. Finally, at the conclusion of the training, raters demonstrated a high level of interrater agreement.

Despite the raters' training and excellent interrater reliability, they encountered challenging coding situations that resulted in decision-making processes unique to each rater. Several items were unclear to raters in the contexts of particular files, and raters indicated that they could not find guidance for the coding decisions in the manual or decision rules. As an example, several SAVRY items assess behavior in school. Due to the lack of a complete definition of "school" in the manual or decision rules, raters needed to decide whether schools included those that were attended during an adjudication to a residential or detention facility. Some raters coded this item only using the youth's performance at schools in the community, arguing that a youth's behavior in a residential or detention setting would likely be different than in the community because of the greater level of security. These raters also held that any review of a juvenile's behavior in a restricted setting was not based on a representative sample because it would be based on a comparison to other disruptive and delinquent students. Other raters, however, did consider a youth's participation in schools that were attended while in a residential or detention facility. Because such schools technically were schools, in that they were accredited, had curricula, and taught the same concepts the juvenile would learn in the community, raters held that such settings were an important way to assess items measuring school conduct. Choosing one way or the other about schools affected the relevant items' ratings. The disagreement across raters on school items is

representative of a larger study-wide level of variability that likely affected the level of agreement with expert raters.

4.3 Implications and Future Research

Despite an impressive body of research supporting the use of the SAVRY in juvenile risk assessment, much of the research has been based on retrospective file-only scores that have not been established as reliable. This study is the first to examine whether file review-based ratings of violence risk on the SAVRY are congruent with those established through standard administration. The lack of agreement between file-only and expert raters in the present study raises several important implications.

First, the present findings, interpreted in light of the many previous file-only SAVRY studies, suggests that to reliably score the SAVRY solely with file information, the evaluator must have access to an adequate source of information on the defendant. The file-only raters appeared to be unable to complete their ratings because the files they were provided lacked sufficient records. Past file-based SAVRY research may have been successful because the file-only rater had access to more meaningful data; for example, many studies were based on the records juveniles amassed while adjudicated at a residential facility. Such documents would likely provide the needed information to code Individual/Clinical risk items, which were often coded as "Unable to Rate" in the present study.

The need for adequate information is an established principle of forensic mental health evaluations (Heilbrun et al., 2002), but the results of this study raise a further question of how a file-based evaluator can determine when enough information has been obtained. When does one have enough information in a file to reliably rate the SAVRY?

The file-only raters in this study did not have enough information, but perhaps there is a threshold level of data that can be reached that allows one to accurately code the SAVRY. Further research will need to be undertaken to address this question. Other research may involve raters providing overall summaries of each file's adequacy in rating the SAVRY.

An alternative aspect of this first implication is the effect of the information that was available in the file-only records. Specifically, file-only raters relied on information in the file that may have not have been balanced or accurate, which would have skewed their understanding of the juvenile's level of risk. For example, in each file provided by one of the three referral sources identified in this study (i.e., public-defender-referred evaluations), a psychosocial summary written by a staff social worker was included. Although these reports include important historical information (e.g., number of prior arrests), evidence based upon the primary investigator's comparison of the psychosocial summary to the expert scores indicates that they also contain a biased viewpoint representing the attorney's interest (i.e., that the juvenile is a low risk and is suitable for decertification from the criminal court). Given the relative lack of other information, file-only raters may have utilized such information and, in contrast to expert raters, not been able to "challenge" it during a clinical interview. Although the reliability and validity of file-only ratings was affected by the lack of file information, the presence of certain records may also have incorrectly influenced their ratings as well.

The second implication of this study is that raters carry specific and individual idiosyncratic differences that may affect their scoring of structured risk assessments. When rating situations were ambiguous, raters relied on unique decision making

processes based on their experience and rating style. Despite having received thorough training prior to coding the primary files in this study, the raters coded certain items differently based on their interpretation of the items' meaning and value.

One aspect of the practice phase, the creation of the decision rules document, was a direct attempt to provide raters with greater clarity about the SAVRY items' meanings. As discussed in the Methods, raters were inconsistent in coding identical file information based upon their individual interpretations of certain item descriptions in the SAVRY manual. Although the decision rules slightly reduced the study's ecological validity, this document proved necessary based on the raters' performance and feedback during the practice phase. No other research was encountered during the preparation of this study that indicated that the SAVRY manual item descriptions were not specific enough, but the results of this study suggest that, in certain file-only coding situations, the SAVRY manual may not be adequate. Further research on both expert and file-only SAVRY ratings could examine whether a revised SAVRY manual or added decision rules document aids in improving risk classification.

The implication that the raters interpreted individual SAVRY items according to idiosyncratic differences is supported by a growing body of research suggesting that the reliability of scores on risk assessment instruments may be affected by rater error, including individual rater differences (Boccaccini, Turner, & Murrie, 2008). Boccaccini and colleagues (2008) found that individual rating styles, such as variations in adherence to scoring procedures or tendencies to use certain ends of the rating scales, contributed to 34% of the variance in a sample of PCL-R scores. The current results suggest that

judgments using structured assessments may vary according to idiosyncratic differences between raters, particularly when data sources are limited.

4.4 Limitations

Certain aspects of this study may have contributed to the lack of agreement between file-only and expert raters, and may limit the degree to which the outcomes will generalize to other situations. One potential contributor to the percentage of unrated items and low level of agreement with the expert clinicians is the use of graduate students and graduate-level researchers as raters. Although these raters were involved in a psychology and law program and were well-trained in the SAVRY, their level of experience in risk assessment was qualitatively less than that of experienced clinicians. For example, an evaluator with years of experience may have been able to identify subtle patterns of behavior even in scant data that may have been missed by less experienced raters. Mitigating this limitation, however, are two arguments. One, the raters completed a thorough training workshop and were well prepared following the training period as based on their performances in the practice phase. Second, the graduate raters were well-situated to provide accurate ratings given that they were functioning as independent raters and were free of any bias that could have developed during a clinical interview.

Another study limitation is related to the amount of missing data in the study. Given the number of items that were left unrated, the level of agreement between expert and file-only raters on individual items could not be determined. Such an analysis limits the potential of this study because that data could inform clinicians as to whether certain items on the SAVRY are more difficult to rate with file information, and thus should be especially assessed during a clinical interview.

Several aspects related to the files used in this study also present potential limitations. During the practice phase, the raters were provided files with an extensive amount of records. Practice files were selected on the basis of their volume and length in order to provide file-only raters with exposure to the different types of data that may have been encountered during the study. Although these comprehensive practice files were determined to be necessary in order to adequately train the raters, they were not reflective of the type of files that the raters received in the primary phase of the study. The file-only raters may have had difficulty transitioning from practicing on files that allowed for a near-complete rating of the SAVRY to files that contained relatively less information.

In addition, due to the types of files utilized in both the practice and primary phases of this study, the results may not generalize to other applications or settings. The juvenile files were all referred from attorneys in one city for primarily one legal question (i.e., decertification from the court). The adolescents in this study had been arrested and directly filed in the criminal court based on the severity of their charges and represent only one subgroup of youth that may be assessed with the SAVRY. As the manual indicates, the SAVRY may be used with any youth who presents with a risk of future aggression or violence. These findings may not be consistent with those obtained with adolescents in different types of settings, such as youth seeking decertification in other jurisdictions or that have been referred for evaluation related to aggressive behavior in a school setting.

Two other limitations should be noted. First, as described in the results, the interrater reliability of the five raters was not measured throughout the study. It is possible that the raters drifted in their observance to coding rules or reduced over time the

effort they put into synthesizing the available, even if limited, information presented in the file. Agreement between file-only raters was not assessed throughout the study due to the pilot-like nature of the study, the focus on the agreement between file-only and expert raters, and the concern that assessing inter-rater reliability would have reduced the already small number of available files. However, given the potential benefits to assessing IRR, other methods of analysis could have been employed. For example, all file-only raters could have rated a sample of the files, and a particular rater's scores could have been assigned to be the representative score for each commonly rated file, thus eliminating the concern that the file would not be able to be used in the primary analysis. Assessing IRR during the study may have provided further information helpful to understanding the current findings.

The other limitation of note is the lack of females in this sample. Although the SAVRY manual indicates that females may be validly assessed with the SAVRY, the authors of the instrument also recently acknowledged that distinct differences have been identified in risk factors for female adolescents (Borum, Lodewijks, Bartel, & Forth, 2010). Given these differences and the low number of females referred to the clinic for decertification (n = 2), females were not included in this study. Future research, however, should strive to include females to determine whether files from female youth show different levels of reliability from males.

Chapter 5: List of References

Andrews, D. A., & Bonta, J. (1995). *Level of Supervision-Revised (LSI-R): An offender assessment system. User's guide*. Toronto: Multi-Health Systems.

Andrews, D. A., Bonta, J., & Hoge, R. D. (1990). Classification for effective rehabilitation: Rediscovering psychology. *Criminal Justice and Behavior, 17*, 19-52.

Andrews, D. A., Bonta, J., & Wormith, J. S. (2004). *Level of Supervision/Case Management Inventory (LS/CMI): An offender assessment system. User's guide*. Toronto: Multi-Health Systems.

Augimeri, L. K., Koegl, C. J., Webster, C. D., & Levene, K. S. (2001). *Early assessment risk list for boys: EARL-20B (Version 2)*. Toronto, ON: Earlscourt Child and Family Centre.

Barefoot v. Estelle, 463 U.S. 880 (1983).

Bartel, P., Borum, R., & Forth, A. (2000). *Structured Assessment for Violence Risk in Youth (SAVRY)* (Consultation Ed.). Tampa, FL: Louis de la Parte Florida Mental Health Institute, University of South Florida.

Boccaccini, M. T., Murrie, D. C., Clark, J. W., & Cornell, D. G. (2008). Describing, diagnosing, and naming psychopathy: How do youth psychopathy labels influence jurors? *Behavioral Sciences and the Law, 26*, 487-510.

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*, 262-283.

Borum, R. (1996). Improving the clinical practice of violence risk assessment: Technology, guidelines and training. *American Psychologist*, *51*, 945−956.

Borum, R. (2000). Assessing violence risk among youth. *Journal of Clinical Psychology, 56*, 1263-1288.

Borum, R., Bartel, P., & Forth, A. (2001). *Manual for the Structured Assessment for Violence Risk in Youth (SAVRY), (Consultation Ed.).* Tampa, FL: University of South Florida.

Borum, R., Bartel, P., & Forth, A. (2002). *Manual for the Structured Assessment for Violence Risk in Youth (SAVRY), Version 1*. Tampa, FL: University of South Florida.

Borum, R., Bartel, P., & Forth, A. (2003). *Manual for the Structured Assessment for Violence Risk in Youth (SAVRY), Version 1.1.* Tampa, FL: University of South Florida.

Borum, R., Bartel, P. A., & Forth, A. E. (2005). Structured Assessment for Violence Risk in Youth. In T. Grisso, G. Vincent, & D. Seagrave (Eds.), *Mental health screening and assessment in juvenile justice* (pp. 311-323). New York: The Guilford Press.

Borum, R., Lodewijks, H., Bartel, P., & Forth, A. (2010). Structured Assessment of Violence Risk in Youth. In R. K. Otto & K. S. Douglas (Eds.), *Handbook of violence risk assessment* (pp. 63-80). New York: Routledge.

Campbell, M. A., Porter, S., & Santor, D. (2004). Psychopathic traits in adolescent offenders: An evaluation of criminal history, clinical, and psychosocial correlates. *Behavioral Sciences and the Law, 22,* 23-47.

Catchpole, R. E. H., & Gretton, H. M. (2003). The predictive validity of risk assessment with violent young offenders: A 1-year examination of criminal outcome. *Criminal Justice and Behavior, 30,* 688-707.

Cottle, C. C., Lee, R. J., & Heilbrun, K. (2001). The prediction of criminal recidivism in juveniles: A meta-analysis. *Criminal Justice and Behavior, 28,* 367–394.

Dawes, R. M., Faust, D., & Meehl, P. E. (1989). Clinical versus actuarial judgment. *Science, 243,* 1668-1674.

DeMatteo, D., & Marczyk, G. (2005). Risk factors, protective factors, and the prevention of antisocial behavior among juveniles. In K. Heilbrun, N. E. S. Goldstein, & R. E. Redding (Eds.), *Juvenile delinquency: Prevention, assessment, and intervention* (pp. 19-44). New York: Oxford University Press.

de Vogel, V., de Ruiter, C., van Beek, D., & Mead, G. (2004). Predictive validity of the SVR-20 and the Static-99 in a Dutch sample of treated sex offenders. *Law and Human Behavior, 28,* 235-251.

Dolan, M. C., & Rennie, C. E. (2008). The Structured Assessment of Violence Risk in Youth as a predictor of recidivism in a United Kingdom cohort of adolescent offenders with conduct disorder. *Psychological Assessment, 20,* 35-46.

Douglas, K. S., & Kropp, P. R. (2002). A prevention-based paradigm for violence risk assessment: Clinical and research applications. *Criminal Justice and Behavior, 29,* 617-657.

Douglas, K., Yeomans, M., & Boer, D. (2005). Comparative validity analysis of multiple measures of violence risk in a sample of criminal offenders. *Criminal Justice and Behavior, 32,* 479-510.

Edens, J. F., Campbell, J. S., & Weir, J. M. (2007). Youth psychopathy and criminal recidivism: A meta-analysis of the Psychopathy Checklist measures. *Law and Human Behavior, 31,* 53-75.

Enebrink, P., Långström, N., & Gumpert, C. H. (2006). Predicting aggressive and disruptive behavior in referred 6- to 12-year old boys: Prospective validation of the EARL-20B risk/needs checklist. *Assessment, 13,* 356-367.

Fleiss, J. L. (1986). *The design and analysis of clinical experiments*. New York: Wiley.

Forth, A. E., Kosson, D. S., & Hare, R. D. (2003). *The Psychopathy Checklist: Youth Version*. Toronto, Ontario, Canada: Multi-Health Systems.

Gacono, C. B. (2000). Suggestions for implementation and use of the Psychopathy Checklists in forensic and clinical practice. In C. B. Gacono (Ed.), *The clinical and forensic assessment of psychopathy: A practitioner's guide* (pp. 175-201). Mahwah, NJ: Lawrence Erlbaum Associates.

Gammelgard, M., Koivisto, A., Eronen, M., & Kaltiala-Heino, R. (2008). The predictive validity of the Structured Assessment of Violence Risk in Youth (SAVRY) among institutionalised adolescents. *The Journal of Forensic Psychiatry and Psychology, 19,* 352-370.

Grann, M., Långström, N., Tengström, A., & Stålenheim, E. G. (1998). Reliability of file-based retrospective ratings of psychopathy with the PCL-R. *Journal of Personality Assessment, 70,* 416-426.

Grisso, T. (1996). Society's retributive response to juvenile violence: A developmental perspective. *Law and Human Behavior, 20,* 229-247.

Grisso, T., Vincent, G., & Seagrave, D. (2005). *Mental health screening and assessment in juvenile justice*. New York: Guilford Press.

Grove, W., Zald, D., Lebow, B., Snitz, B., & Nelson, C. (2000). Clinical versus mechanical prediction: A meta-analysis. *Psychological Assessment, 12,* 19-30.

Hannah-Moffat, K., & Maurutto, P. (2003). *Youth risk/need assessment: An overview of issues and practices.* Ottawa: Department of Justice Canada, Research and Statistics Division.

Hare, R. D. (1991). *The Hare Psychopathy Checklist–Revised (PCL-R)*. Toronto, Ontario: Multi-Health Systems.

Hare, R. D. (2003). *The Hare Psychopathy Checklist–Revised* (2nd ed.). Toronto, Ontario: Multi-Health Systems.

Heilbrun, K., Goldstein, N. E. S, & Redding, R. E. (Eds.) (2005). *Juvenile delinquency: Prevention, assessment, and intervention*. New York: Oxford University Press.

Heilbrun, K., Leheny, C., Thomas, L., & Huneycutt, D. (1997). A national survey of U.S. statutes on juvenile transfer: Implications for policy and practice. *Behavioral Sciences and the Law, 15*, 125-149.

Heilbrun, K., Marczyk, G. R., & DeMatteo, D. (2002). *Forensic mental health assessment: A casebook*. New York: Oxford University Press.

Heilbrun, K., Yasuhara, K., & Shah, S. (2010). Violence risk assessment tools: Overview and critical analysis. In R. K. Otto & K. S. Douglas (Eds.), *Handbook of violence risk assessment* (pp. 1-14). New York: Routledge.

Hoge, R. (2002). Standardized instruments for assessing risk and need in youthful offenders. *Criminal Justice and Behavior, 29*, 380−396.

Hoge, R. D., & Andrews, D. A. (2001). *The Youth Level of Service/Case Management Inventory (YLS/CMI) Screening Version.* Ottawa, Ontario, Canada: Carleton University.

Hoge, R. D., & Andrews, D. A. (2002). *The Youth Level of Service/Case Management Inventory*. Toronto, ON, Canada: Multi-Health Systems.

Hoge, R. D., & Andrews, D. A. (2010). *Evaluation for risk of violence in juveniles*. New York: Oxford.

How to compare 2 intraclass correlation coefficients or Cronbach alphas from two independent samples. (n.d.). Retrieved April 28, 2012, from http://www-01.ibm.com/support/docview.wss?uid=swg21498525

Howell, J. (1997). *Juvenile justice and youth violence*. Thousand Oaks, CA: Sage.

In re Gault, 387 U.S. 1 (1967).

Kent v. United States, 383 U.S. 541 (1966).

Levene, K. S., Augimeri, L. K., Pepler, D. J., Walsh, M. M., Webster, C. D., & Koegl, C. J. (2001). *Early assessment risk list for girls (EARL-21G)*. Toronto: Earlscourt Child and Family Centre.

Lipsey, M. W., & Derzon, J. H. (1998). Predictors of violence and serious delinquency in adolescence and early adulthood: A synthesis of longitudinal research. In R. Loeber & D. P. Farrington (Eds.), *Serious and violent juvenile offenders: Risk factors and successful interventions* (pp. 86-105). Thousand Oaks, CA: Sage.

Litwack, T. R. (2001). Actuarial versus clinical assessments of dangerousness. *Psychology, Public Policy, and Law, 7*, 409-443.

Lodewijks, H. P. B., Doreleijers, A. H., & De Ruiter, C. (2008a). SAVRY risk assessment in violent Dutch adolescents: Relation to sentencing and recidivism. *Criminal Justice and Behavior, 35*, 696-709.

Lodewijks, H. P. B., Doreleijers, T. A. H., de Ruiter, C., & Borum, R. (2008b). Predictive validity of the Structured Assessment of Violence Risk in Youth (SAVRY) during residential treatment. *International Journal of Law and Psychiatry, 31*, 263-271.

Lynam, D. R. (2002). Fledgling psychopathy: A view from personality theory. *Law and Human Behavior, 26*, 255-259.

Marczyk, G. R., Heilbrun, K., Lander, T., & DeMatteo, D. (2003). Predicting juvenile recidivism with the PCL:YV, MAYSI, and YLS/CMI. *International Journal of Forensic Mental Health, 2*, 7-18.

Melton, G., Petrila, J., Poythress, N., & Slobogin, C. (2007). *Psychological evaluations for the courts* (3rd ed.). New York: Guilford Press.

Meyers, J. R., & Schmidt, F. (2008). Predictive validity of the Structured Assessment for Violence Risk in Youth (SAVRY) with juvenile offenders. *Criminal Justice and Behavior, 35*, 344-355.

Moffitt, T. E. (1993). Adolescence-limited and life-course-persistent antisocial behavior: A developmental taxonomy. *Psychological Review, 100*, 674-701.

Monahan, J. (1981). *The clinical prediction of violent behavior*. (DHHS Publication No. ADM 89-92). Rockville, MD: National Institute of Mental Health.

Mossman, D. (1994). Assessing predictions of violence: Being accurate about accuracy. *Journal of Consulting and Clinical Psychology, 62*, 783-792.

Nellis, A. (2012). The Lives of Juvenile Lifers: Findings from a National Survey. Washington D.C.: The Sentencing Project. Downloaded from http://sentencingproject.org/doc/publications/jj_The_Lives_of_Juvenile_Lifers.pdf

Odgers, C. L., Moretti, M. M., & Repucci, N. D. (2005). Examining the science and practice of violence risk assessment with female adolescents. *Law and Human Behavior, 29*, 7-27.

Olver, M. E., Stockdale, K. C., & Wormith, J. S. (2009). Risk assessment with young offenders: A meta-analysis of three assessment measures. *Criminal Justice and Behavior, 36*, 329-353.

Pennsylvania Juvenile Court Judges' Commission. (2008). *The Juvenile Act* (42 Pa. C.S. Sec. 6301 et seq.). Harrisburg, PA: Author.

Puzzanchera, C. (2009). *Juvenile arrests 2007*. Washington, DC: U.S. Department of Justice, Office of Juvenile Justice and Delinquency Prevention.

Puzzanchera, C., & Adams, B. (2011). *Juvenile arrests 2009*. Washington, DC: U.S. Department of Justice, Office of Juvenile Justice and Delinquency Prevention.

Redding, R. (2008). *Juvenile transfer laws: An effective deterrent to delinquency?* Washington, DC: Department of Juvenile Justice and Delinquency Prevention.

Salekin, R. T., & Frick, P. J. (2005). Psychopathy in children and adolescents: The need for a developmental perspective. *Journal of Abnormal Child Psychology, 33*, 403-409.

Salekin, R. A., Yff, R. M. A., Neumann, C. S., Leistico, A. M. R., & Zalot, A. A. (2002). Juvenile transfer to adult courts: A look at the prototypes for dangerousness, sophistication-maturity, and amenability to treatment through a legal lens. *Psychology, Public Policy, and Law, 8*, 373-410.

Schmidt, F., Hoge, R. D., & Gomes, L. (2005). Reliability and validity analyses of the Youth Level of Service/Case Management Inventory. *Criminal Justice and Behavior, 32*, 329-344.

Schmidt, F., Campbell. M. A., & Houlding, C. (2011). Comparative analyses of the YLS/CMI, SAVRY, and PCL:YV in adolescent offenders: A 10-year follow-up into adulthood. *Youth Violence and Juvenile Justice, 9*, 23-42.

Scott, E. S. (2000). The legal construction of adolescence. *Hofstra Law Review, 29*, 547-598.

Seagrave, D., & Grisso, T. (2002). Adolescent development and the measurement of juvenile psychopathy. *Law and Human Behavior, 26*, 219-239.

Shrout, P. E., & Fleiss, J. L. (1979). Intraclass correlations: Uses in assessing rater reliability. *Psychological Bulletin, 86*, 420-428.

Skeem, J. L., & Cauffman, E. (2003). Views of the downward extension: Comparing the youth version of the Psychopathy Checklist with the Youth Psychopathic Traits Inventory. *Behavioral Sciences and the Law, 21*, 737-770.

Skeem, J. L., & Petrila, J. (2004). Juvenile psychopathy: Informing the debate. *Behavioral Sciences and the Law, 22*, 1-4.

Slobogin, C. (2006). Dangerousness and expertise redux. *Emory Law Journal, 56*, 275-325.

The Juvenile Act, 42 Pa. Cons. Stat. Ann. § 6355 (2008).

Viljoen, J.L., MacDougall, E. A. M., Gagnon, N. C., & Douglas, K. S. (2010). Psychopathy evidence in legal proceedings involving adolescent offenders. *Psychology, Public Policy, and Law,* 16, 254-283.

Viljoen, J. L., Scalora, M., Ullman, D., Cuadra, L., Bader, S., Chavez, V., & Lawrence, L. (2008). Assessing risk for violence in adolescents who have sexually offended: A comparison of the J-SOAP-II, SAVRY, and J-SORRAT-II. *Criminal Justice and Behavior, 35,* 5-23.

Vincent, G. M. (2006). Psychopathy and violence risk assessment in youth. *Child and Adolescent Psychiatric Clinics of North America, 15,* 407-428.

Vincent, G. M., Chapman, J., & Cook, N. E. (2011). Risk-needs assessment in juvenile justice: Predictive validity of the SAVRY, racial differences, and the contribution of needs factors. *Criminal Justice and Behavior, 38,* 42-62.

Webster, C. D., Douglas, K. S., Eaves, D., & Hart, S. D. (1997). *HCR-20: Assessing risk for violence* (Version 2). Burnaby, BC: Mental Health, Law, and Policy Institute, Simon Fraser University.

Webster, C. D., Hucker, S. J., & Bloom, H. (2002). Transcending the actuarial versus clinical polemic in assessing risk for violence. *Criminal Justice and Behavior, 29,* 659-665.

Welsh, J. L., Schmidt, F., McKinnon, L., Chattha, H. K., & Meyers, J. R. (2008). A comparative study of adolescent risk assessment instruments: Predictive and incremental validity. *Assessment, 15,* 104-115.

Wong, S. (1988). Is Hare's Psychopathy Checklist reliable without the interview? *Psychological Reports, 62,* 933-934.

Appendix A: SAVRY Items

Historical items

1. History of violence
2. History of non-violent offending
3. Early initiation of violence
4. Past supervision/intervention failures
5. History of self-harm or suicide attempts
6. Exposure to violence in the home
7. Childhood history of maltreatment
8. Parental/caregiver criminality
9. Early caregiver disruption
10. Poor school achievement

Social/contextual items

11. Peer delinquency
12. Peer rejection
13. Stress and poor coping
14. Poor parental management
15. Lack of personal/Social support
16. Community disorganization

Individual items

17. Negative attitudes
18. Risk taking/impulsivity

19. Substance use difficulties

20. Anger management problems

21. Low empathy/remorse

22. Attention deficit/hyperactivity difficulties

23. Poor compliance

24. Low interest/Commitment to school or work

Protective items

P1. Prosocial involvement

P2. Strong social support

P3. Strong attachments and bonds

P4. Positive attitude towards intervention and authority

P5. Strong commitment to school or work

P6. Resilient personality

Table 1
Previous Research on the SAVRY

Authors	**Focus of Research**	**Research Design**
Catchpole & Gretton (2003)	Prediction of general and violent recidivism in a sample of Canadian adolescent offenders over a 1-year period.	Retrospective coding of SAVRY based on archived file information.
Dolan & Rennie (2008)	Prediction of general and violent recidivism in a sample of British adolescent offenders over a 1-year period.	Prospective coding based on transcripts of clinical interview and file information (no coding of, or comparison to, file-only scores).
Gammelgard et al. (2008)	Prediction of violent behavior in various residential treatment settings in a sample of Finnish adolescent offenders during first 6-months in placement.	Retrospective coding of SAVRY based on archived file information.
Lodewijks et al. (2008a)	Prediction of violent recidivism in a sample of Dutch adolescent offenders over a 3-year period.	Retrospective coding of SAVRY based on archived file information.
Lodewijks et al. (2008b)	Prediction of disruptive behavior in a residential treatment setting in a sample of Dutch adolescent offenders during placement (avg. placement = 22 months).	Prospective coding based on clinical interview and file information (no coding of, or comparison to, file-only scores).
Meyers & Schmidt (2008)	Prediction of general and violent recidivism in a sample of Canadian adolescent offenders over a 1- and 3-year period.	Retrospective coding of SAVRY based on archived file information.
Viljoen et al. (2008)	Prediction of violent, sexual, and non-sexual aggression in a sample of American adolescent sexual offenders during treatment (avg. placement = 1 year) and after discharge (avg. follow-up = 6.58 years).	Retrospective coding of SAVRY based on archived file information.

Table 1 (continued)
Previous Research on the SAVRY

Welsh et al. (2008)	Prediction of general and violent recidivism in a sample of Canadian adolescent offenders following discharge from placement (average follow-up = 35.8 months).	Retrospective coding of SAVRY based on archived file information
Schmidt et al. (2011)	Prediction of violent, nonviolent, sexual, and technical recidivism in a sample of Canadian adolescent offenders following discharge from placement (average follow-up = 10.4 years).	Retrospective coding of SAVRY based on archived file information
Vincent et al. (2011)	Prediction of violent and nonviolent recidivism in a sample of American adolescent offenders following discharge from placement (average follow-up = 5.1 years).	Prospective coding of SAVRY based on clinical interview and file information (no coding of, or comparison to, file-only scores)

Table 2
Demographic Data

Demographic Characteristic	***n***	**%**
Race/Ethnicity		
African American	26	74.3
Hispanic	7	20.0
Caucasian	1	2.9
Other (“Biracial”)	1	2.9
Age		
15	2	5.7
16	12	34.3
17	16	45.7
18	5	14.3
Substance Use		
Daily	13	37.1
Weekly	5	14.3
Monthly	3	8.6
Rare use or no longer using	11	31.4
No prior substance use	3	8.6
Number of prior arrests		
0	10	28.6

Table 2 (continued)
Demographic Data

1	8	22.9
2	6	17.1
3	5	14.3
4	3	8.6
6	1	2.9
7	1	2.9
9	1	2.9

Table 3
Examples of Decision Rules

Item #	**Coding Disagreement**	**Decision Rule**
1 – History of Violence	Should a charge for a previous violent offense that has been withdrawn or dismissed count under this item?	Charges that are withdrawn or dismissed should not count in the youth's offense history.
9 – Parental Criminality	Who is considered a "parent"? What if the youth lives with non-traditional parents?	Criminality by any current and past caregivers (including biological parents, stepparents, and other people acting in a parental role) is to be used to rate this item.
16 – Community Disorganization	Can an assumption be made about the juvenile's neighborhood even if the file is lacking specific information about the youth's area of residence?	Unless the juvenile's neighborhood is specifically described, or there is specific information about activity that takes place in the juvenile's community, do not make assumptions about community disorganization.
20 – Anger Management	Does a youth's participation in "Anger Management" classes indicate problems with anger management?	Because many "Anger Management" classes in detention settings are mandatory, do not count such classes towards this item – unless file indicates the youth has an actual anger management problem.
P4 – Positive Attitude to Intervention	Does an increase in positive behavior at home fall under this item?	This item does not include behavior at home (e.g., "listening to mother") unless the behaviors at home are directly related to an attempt to reduce the risk of youth's violence.

Table 4
SAVRY Summary Risk Ratings

SAVRY Summary Risk Rating ***n*** **(%)**	**Expert Ratings**	**File-Only Ratings**
Low	13 (37.1)	5 (14.3)
Moderate	18 (51.4)	9 (25.7)
High	4 (11.4)	6 (17.1)
Unable to Rate	0 (0)	15 (42.9)

Table 5
Percentage of Individual SAVRY Risk Factor Item Ratings

SAVRY Risk Factors	**Expert / File-Only**			
	Low	Moderate	High	Missing
Historical risk factors				
1. History of violence	14.3 / 5.7	42.9 / 51.4	42.9 / 34.3	0 / 8.6
2. History of nonviolent offending	40.0 / 20.0	57.1 / 37.1	2.9 / 8.6	0 / 34.3
3. Early initiation of violence	60.0 / 45.7	22.9 / 11.4	17.1 / 14.3	0 / 28.6
4. Past supervision/ intervention failures	34.3 / 14.3	31.4 / 17.1	5.7 / 17.1	28.6 / 51.4
5. History of self-harm or suicide attempts	82.9 / 28.6	8.6 / 5.7	8.6 / 2.9	0 / 62.9
6. Exposure to violence in the home	97.1 /11.4	2.9 / 0	0 / 2.9	0 / 85.7
7. Childhood history of maltreatment	80.0 / 20.0	20.0 / 17.1	0 / 0	0 / 62.9
8. Parental/care-giver criminality	42.9 / 14.3	45.7 / 5.7	11.4 / 17.1	0 / 62.9
9. Early caregiver disruption	60.0 / 20.0	31.4 / 25.7	8.6 / 2.9	0 / 51.4
10. Poor school achievement	8.6 / 11.4	57.1 / 17.1	34.3 / 34.3	0 / 37.1

Table 5 (continued)
Percentage of Individual SAVRY Risk Factor Item Ratings

Social/Contextual risk factors				
11. Peer delinquency	25.7 / 2.9	51.4 / 17.1	22.9 / 28.6	0 / 51.4
12. Peer rejection	68.6 / 11.4	25.7 / 5.7	5.7 / 5.7	0 / 77.1
13. Stress and poor coping	20.0 / 2.9	48.6 / 8.6	31.4 / 40.0	0 / 48.6
14. Poor parental management	68.6 / 8.6	25.7 / 14.3	2.9 / 17.1	2.9 / 60.0
15. Lack of personal/ social support	77.1 / 37.1	11.4 / 8.6	8.6 / 8.6	2.9 / 45.7
16. Community disorganization	5.7 / 2.9	31.4 / 0	40.0 / 17.1	22.9 / 80.0
Individual/Clinical risk factors				
17. Negative attitudes	65.7 / 0	34.3 / 14.3	0 / 20.0	0 / 65.7
18. Risk taking/Impulsivity	28.6 / 0	54.3 / 22.9	17.1 / 37.1	0 / 40.0
19. Substance use difficulties	40.0 / 14.3	37.1 / 11.4	22.9 / 25.7	0 / 48.6
20. Anger management problems	37.1 / 8.6	45.7 / 8.6	17.1 / 20.0	0 / 62.9
21. Low Empathy/ Remorse	77.1 / 5.7	20.0 / 5.7	2.9 / 14.3	0 / 74.3
22. Attention-deficit/ hyperactivity difficulties	57.1 / 11.4	34.3 / 20.0	8.6 / 8.6	0 / 60.0
23. Poor compliance	62.9 / 17.1	28.6 / 20.0	8.6 / 17.1	0 / 45.7
24. Low interest/ commitment to school	40.0 / 28.6	25.7 / 5.7	34.3 / 22.9	0 / 42.9

Table 6
Percentage of Individual SAVRY Protective Factor Item Ratings

SAVRY Protective Factor	Expert / File-Only		
	Present	**Absent**	**Missing**
P1. Prosocial involvement	57.1 / 22.9	40.0 / 14.3	2.9 / 62.9
P2. Strong social support	80.0 / 45.7	14.3 / 8.6	5.7 / 45.7
P3. Strong attachments and bonds	97.1 / 42.9	0 / 5.7	2.9 / 51.4
P4. Positive attitude towards intervention and authority	74.3 / 31.4	20.0 / 17.1	5.7 / 51.4
P5. Strong commitment to school or work	40.0 / 25.7	57.1 / 37.1	2.9 / 37.1
P6. Resilient personality	48.6 / 20.0	51.4 / 17.1	0 / 62.9

Table 7
Agreement with Experts by Rater

File-Only Rater	Individual Items IRR	Summary Risk Rating IRR	Overall IRR
Rater 1	.206	.471[a]	.219
Rater 2	.466	.308[a]	.462
Rater 3	.414	1.00[b]	.415
Rater 4	.333	.714[c]	.345
Rater 5	.508	.500[a]	.505

[a] Based on 4 out of 7 Summary Risk Ratings. These raters coded their remaining 3 Summary Risk Ratings as "Unable to Rate."

[b] Based on 3 out of 7 Summary Risk Ratings. This rater coded the remaining 4 Summary Risk Ratings as "Unable to Rate."

[c] Based on 5 out of 7 Summary Risk Ratings. This rater coded the remaining 2 Summary Risk Ratings as "Unable to Rate."

Vita

Jeffrey D. Burl

Jeffrey Burl is a 5th-year doctoral student in the Clinical Psychology PhD program at Drexel University. After obtaining his undergraduate degree from Fordham University, Jeff volunteered abroad for 2 years at the Central Prison in Belize, Central America. He then earned a master's degree in Forensic Psychology at John Jay College of Criminal Justice in New York. At Drexel, Jeff has served as an interventionist on a NIDA-funded R21 study seeking to develop an intervention for low-risk drug court clients and has conducted research on variations in curricula across graduate forensic training programs. He has also completed practica at the Ferris School (a secure facility for adjudicated adolescent males), the Forensic Assessment Clinic, and the Student Counseling Center. He is currently on internship in the Clinical-Forensic track at the University of North Carolina at Chapel Hill School of Medicine. He defended his dissertation in May 2012 and expects to complete the requirements of the doctoral degree in August 2012.

Education and Training

University of North Carolina at Chapel Hill School of Medicine, Chapel Hill, NC
APA-Accredited Predoctoral Clinical Internship (2011-2012)

Drexel University, Philadelphia, PA
Ph.D. in Clinical Psychology (expected August 2012)

John Jay College of Criminal Justice (CUNY), New York, NY
M.A. in Forensic Psychology (2007)

Fordham University, Bronx, NY
B.A. in American Studies and Anthropology (2001)

Publications

Burl, J., Shah, S., Filone, S., Foster, E., & DeMatteo, D. (2012). A survey of graduate training programs and coursework in forensic psychology. *Teaching of Psychology*, *39*, 48-53.

DeMatteo, D., Marczyk, G., Krauss, D. A., & **Burl, J.** (2009). Educational and training models in forensic psychology. *Training and Education in Professional Psychology*, *3*, 184-191.

Reprinted as DeMatteo, D., Marczyk, G., Krauss, D. A., & **Burl, J.** (2011). Educational and training models in forensic psychology. In C. R. Bartol & A. M. Bartol (Eds.), *Current perspectives in forensic psychology and criminal behavior* (3rd ed.) (pp. 2-10). Thousand Oaks, CA: Sage Publications.

Reprinted as DeMatteo, D., Marczyk, G., Krauss, D. A., & **Burl, J.** (in press). Educational and training models in forensic psychology. In J. B. Helfgott (Ed.), *Criminal psychology, volume 1*: *Theory and research on criminal psychology*.

DeMatteo, D., Marlowe, D., Winters, K., Festinger, D., Arabia, P., Anumba, N., & **Burl, J.** (2010). A prevention intervention for low-risk and low-need adult drug court clients: An intervention manual. Philadelphia, PA: Drexel University and Treatment Research Institute.

Teaching Experience

Adjunct Faculty	
2008, 2009, 2010	Personality Psychology
Graduate Assistant	
2009	Practicum Fieldwork Seminar
2008	Physiology of Behavior, Cognitive Psychology
2007	Neuropsychology