2018

Validation of Virginia's Juvenile Risk Assessment Instrument

Jessica P. Schneider

Virginia Commonwealth University

Validation of Virginia's Juvenile Risk Assessment Instrument

A dissertation submitted in
partial fulfillment of the requirements for the
degree of Doctor of Philosophy at
Virginia Commonwealth University.

by
Jessica P. Schneider

Richmond, Virginia
2018

Dissertation Chair

Hayley M. Cleary, PhD
Assistant Professor
L. Douglas Wilder School of Government and Public Affairs
Virginia Commonwealth University

Dissertation Committee

Ashlee R. Barnes, PhD
Assistant Professor
L. Douglas Wilder School of Government and Public Affairs
Virginia Commonwealth University

Patrick V. Dattalo, PhD
Professor
School of Social Work
Virginia Commonwealth University

Patrick G. Lowery, PhD
Assistant Professor
L. Douglas Wilder School of Government and Public Affairs
Virginia Commonwealth University

The findings of this study are the responsibility of the researcher, and cooperation by the Virginia Department of Juvenile Justice in facilitating this research should not be construed as an endorsement of the conclusions drawn by the researcher.

Table of Contents

Acronyms and Abbreviations

AUC: Area Under the Receiver Operating Characteristic (ROC) Curve

BOT: Back on Track

CMAP: Case Management Assessment Protocol

DIFR: Dispersion Index for Risk

DJJ: Virginia Department of Juvenile Justice

FY: Fiscal Year (in Virginia)

IDA: Iowa Delinquency Assessment

LOS: Length of Stay

LSI-R: Level of Service Inventory – Revised

Orbis: Orbis Partners Inc.

PACT: Positive Achievement Change Tool

RNR: Risk-Needs-Responsivity

ROC: Receiver Operating Characteristic

SAVRY: Structured Assessment of Violence Risk in Youth

WSJCA: Washington State Juvenile Court Assessment

YASI: Youth Assessment and Screening Instrument

YLS/CMI: Youth Level of Service/Case Management Inventory

Tables and Figures

Abstract

Utilizing a validated risk assessment tool to predict future offending is recommended as best practices in corrections by a number of professional organizations (Latessa & Lovins, 2010). Guided by the risk-needs-responsivity model, risk assessment tools have evolved to help inform criminal justice practitioners by identifying offenders most in need of intervention or supervision, guiding the case plan to optimize outcomes (Bonta & Andrews, 2007). The Virginia Department of Juvenile Justice (DJJ) utilizes the Youth Assessment and Screening Instrument (YASI) at all stages of contact with youthful offenders, including intake, probation, commitment, and parole (DJJ, 2016). However, risk assessment instruments do not always generalize across populations (Schwalbe, 2007) and are not always used effectively for case planning decisions (Singh et al., 2014). This study focused on the accuracy, equity, and usage of YASI in the Virginia juvenile justice system. Findings suggested that YASI performed at the expected and adequate levels of predictive validity in comparison to existing research. The predictive validity of the overall and dynamic risk scores and levels was statistically equivalent for males and females, but the *Community/Peers* and *Family* domains had stronger predictive validity for males than females. The predictive validity was statistically equivalent for White and Black youth for overall risk levels and dynamic risk scores and levels; however, the predictive validity for the overall risk scores was higher for White youth than Black youth. Each domain had a positive correlation between risk and assignment as a case planning priority area with a wide variation in the strength of correlation. Future research should focus on instrumental validity, protective factors, inter-rater reliability, domain interactions and clusters, reoffense types and timing, additional group and geographical differences, weighting and scoring, service matching, recidivism reduction, and program evaluations. Policy recommendations regarding risk

assessment use in juvenile justice systems include a repeated cycle of determining purpose and function, conducting staff and stakeholder training, testing, and calibrating and modifying the tool.

Chapter 1: Introduction and Statement of the Problem

Assessing risk is a common practice across multiple disciplines: physicians assess risks for diseases or death; economists assess risks for market impacts; psychiatrists assess risks for suicide. These fields practice the prediction of negative future events based on known factors, with varying degrees of accuracy, to guide their decisions. The field of criminal justice, likewise, attempts to predict the likelihood of offenders repeating their criminogenic behavior. As in these examples, assessing risk of reoffending helps to inform criminal justice practitioners' actions by identifying offenders most in need of intervention or supervision. Additionally, risk assessment tools help to systemize the information collected and increase consistency and equity in the treatment of offenders through objective scoring (Schwalbe, Fraser, Day, & Arnold, 2004).

Utilizing a validated risk assessment tool to predict future offending is recommended as a best practice in corrections by a number of professional organizations (Latessa & Lovins, 2010). Risk assessments are used in pretrial, probation, correctional settings, and parole (Latessa & Lovins, 2010). They inform courts in deciding bonds and sentences, probation and parole officers in deciding supervision and program levels, and correctional systems in deciding classification levels and releases (Latessa & Lovins, 2010). They may be completed by forensic psychologists or mental health professionals (Olver, Stockdale, & Wormith, 2009), probation or parole officers, case managers, or other staff within the justice system (Baird et al., 2013). As early as the 1990s, the majority of adult probation and parole agencies reported using standardized and objective risk assessment instruments and viewed them as beneficial to both the offenders and staff (Jones, Johnson, Latessa, & Travis, 1999). Similarly, in juvenile correctional settings, the use of risk assessments has increased from a third of states in 1990 to 86% in 2003 (Schwalbe, 2007).

In order to portray the significant role of risk assessments in juvenile justice settings and the need for the current research, the history of juvenile justice will be briefly discussed, followed by a description of the risk-needs-responsivity (RNR) model that serves as the foundation of risk assessments, the evolution of risk assessments, and the debates among experts concerning best approaches to risk assessments. Finally, the current state of risk assessment in the Virginia juvenile justice system will be presented as the focus of the current research.

Brief History of Juvenile Justice

The first juvenile justice system was created in Illinois in 1899 to provide a rehabilitative intervention for criminal youth (Benekos & Merlo, 2005), and most other states quickly followed (McGowan et al., 2007). The new informal systems focused on therapeutic and rehabilitative services for youth rather than punitive measures in response to the offense; using the doctrine of *parens patriae*, judges worked with the individual child's best interest in mind (Benekos & Merlo, 2005; Snyder & Sickmund, 1999). The separate courts for juveniles were constructed in response to the acceptance of developmental and psychosocial differences between adults and youth. Juveniles were believed to be less aware of consequences, less responsible, and less culpable for their actions. Furthermore, youth were considered more malleable and therefore more amenable to rehabilitative treatments and services than adults (McGowan et al., 2007).

The due process constitutional rights protecting adults were not originally conveyed to youth because of the benevolent purpose of the juvenile justice systems; additional protections were viewed as unnecessary (Snyder & Sickmund, 1999). However, as youth received punishments without effective treatment or due process, concern arose in the 1950s and 1960s that youth were receiving the "worst of both worlds" (Snyder & Sickmund, 1999). In 1966, after a judge transferred a youth to adult court without clear rationale, the Supreme Court determined

in *Kent v. United States* that courts must provide youth some due process rights for transfers to adult court, including written reasons for the decision. Soon after, in response to a 15-year-old's commitment until age 21 for an obscene prank telephone call, the 1967 *In re Gault* decision established juveniles' rights to notice of charges, counsel, question witnesses, and protection against self-incrimination in cases that may result in confinement in an institution. *In re Winship* (1970) then established the reasonable doubt standard of evidence for juvenile cases rather than the lower civil standard of preponderance of evidence. These changes transfigured the informal juvenile process to a more formal system with similarities to the adult criminal courts.

Despite these expansions of due process rights in the fact-finding stages, juvenile courts were not completely aligned with the adult system. For example, *McKeiver v. Pennsylvania* (1971) distinguished the juvenile court process from the adult system by not extending the right to a jury trial to juvenile cases. Overall, the Supreme Court decisions of the 1960s and 1970s continued to support the developmental differences between youth and adults that served as the foundation of the juvenile court while acknowledging the need for protections in a system sometimes disseminating harsh punishments.

Beginning in the 1970s, with some due process rights in place, the rehabilitative approach of juvenile justice systems gradually shifted to a punitive model (Benekos & Merlo, 2005). The news media during the 1980s and 1990s perpetuated the public's support of harsher punishments for juvenile offenders by highlighting particularly violent criminal acts committed by youth (Brannen et al., 2006) and a rise in violent juvenile crime rates, reinforcing the system's reliance on punishment over treatment (McGowan et al., 2007). The term "superpredator" was coined in the 1990s to describe what was believed to be a new breed of young violent criminals (Scott & Steinberg, 2008). Furthermore, researchers questioned whether offenders were amenable to

treatment, propagating the message that "nothings works" (Martinson, 1974). In the face of these messages, the public and policy-makers perceived juvenile justice systems as too lenient.

In reaction, systems moved toward the removal of individualized considerations and the adultification of the process (Merlo and Benekos, 2010). By 1997, 45 states enacted laws to increase the transfer of juveniles to the adult system by modifying age or offense criteria, 31 states expanded courts' sentencing options, and 47 states reduced the confidentiality protections within the juvenile system (Snyder & Sickmund, 1999). Thus, the ideology of treatment and care of individual children shifted to policies and legislation emphasizing crime control and public safety (Hjalmarsson, 2009).

However, a trend away from these punitive practices in juvenile justice emerged in recent decades (Mears, Cochran, Greenman, Bhati, & Greenwald, 2011). The Supreme Court influenced this redirection in decisions citing research on development, culpability, and rehabilitation. In 2002, the Supreme Court decided in *Atkins v. Virginia* that the execution of a "mentally retarded" 18-year-old constituted cruel and unusual punishment. Though not directly related to juveniles, this decision indicated that developmental differences impact the appropriateness of punishments. In fact, the same conclusion regarding executions was applied to juveniles in *Roper v. Simmons* (2005). By 2012, the Supreme Court further extended the restrictions on harsh sentences for juveniles by determining life without parole for non-homicide offenses (*Graham v. Florida,* 2010) and mandatory life without parole for homicides (*Jackson v. Hobbs*, 2012; *Miller v. Alabama*, 2012) were cruel and unusual punishments. In these decisions, the Supreme Court clearly distinguished the differences between youth and adults and emphasized the importance of developmental considerations for determining culpability and sentencing.

Along with these formal changes, evidence-based interventions for juvenile offenders came to the forefront (Mears et al., 2011), including providing mental health treatment and expanding community-based services (Merlo & Benekos, 2010). This transition to less punitive policies was motivated by an emphasis on outcomes: programs and practices shown to be effective in reducing delinquency grew to be valued by juvenile justice systems (Mears et al., 2011). The "nothing works" doctrine was contradicted by subsequent research (e.g., Izzo & Ross, 1990) and was even retracted by the original author (Martinson, 1979). Some states increased the age of jurisdiction, eliminated juvenile life without parole, and reversed some of the automatic transfers to adult court that had been instituted in previous decades (Merlo & Benekos, 2010).

Since its inception, juvenile justice systems have experienced a series of transformations involving both philosophy and structural components. Beginning with a rehabilitative model, they gradually added systemized protections in the face of harsh punishments and lack of effective services. These more formal systems evolved into an increasingly punitive model focused on public safety. However, recent trends have acknowledged the developmental differences between youth and adults and the potential for effective interventions. Thus, the goal of juvenile justice systems has now largely returned to rehabilitation, though many of the laws and policies enacted over past decades shifted the structure and processes away from the original informal configuration. Although punishment remains a major component of juvenile courts, the techniques and services currently utilized are ultimately meant to reduce juvenile reoffending.

RNR Model

The rehabilitative approach of juvenile justice systems is guided by the RNR model. Andrews, Bonta, and Hoge (1990) first formalized the adult RNR model to classify offenders

and determine the most appropriate and effective intervention to reduce their odds of reoffending. Based on personality and cognitive social learning theories (Grieger & Hosser, 2014), the RNR model combines the balance of *risk*, *needs*, and *responsivity* characteristics of an offender to indicate the optimal approach for individualized rehabilitative treatment.

Risk factors are static (i.e., fixed) or dynamic (i.e., changeable) characteristics related to a higher likelihood of offending that can be assessed prior to service delivery (Andrews et al., 1990). Using the cumulative risk property (Fraser, 2004), these risk factors have an additive effect on an individual's overall risk level. The risk principle, dating to the 1940s (Andrews et al., 1990), states that increasing levels of risk of reoffending require proportionally increasing levels or intensities of intervention to optimize rehabilitative outcomes (Andrews, Bonta, & Wormith, 2006; Lipsey, 2009; Makarios, Sperber, & Latessa, 2014). While high risk individuals benefit most from higher dosages of services, these higher service dosages are not as effective and may be detrimental to low risk individuals (Lowenkamp & Latessa, 2005).

Adherence to the risk principle does not guarantee positive results; intensive services that are not appropriate for the individual may harm outcomes (Andrews et al., 1990). Thus, the *needs* principle states that interventions should focus on the criminogenic needs relating to the individual's dynamic risk factors (Bonta & Andrews, 2007). Criminogenic needs are a subset of risk factors directly related to criminality (Andrews et al., 1990; Bonta & Andrews, 2007). These factors are the characteristics that are changeable rather than static or historical, including substance abuse, peer delinquency, and school-related problems (Schwalbe, 2007). According to this principle, by targeting the individual's specific criminogenic needs, the odds of reoffending are likely to reduce (Andrews et al., 1990). Indeed, evidence suggests that case planning and

service delivery matching individuals' criminogenic needs reduces recidivism (Luong & Wormith, 2011; Vieira, Skilling, & Peterson-Badali, 2009).

Finally, the *responsivity* principle refers to the use of individualized treatment tailored to the strengths, protective factors, and other characteristics of the offender in order to promote learning (Bonta & Andrews, 2007). General responsivity is related to the best practice of cognitive social learning interventions relying on a positive relationship between the practitioner and client and structured toward prosocial change (Bonta & Andrews, 2007); specific responsivity is individualized adaptation to the particular client's personality, mental health, cognitive abilities, learning styles, and motivation (Andrews et al., 1990; Andrews et al., 2006; Bonta & Andrews, 2007; Vieira et al., 2009). For specific responsivity, the interpersonal style and method of the intervention delivery should be adjusted to fit the individual offender in order to maximize positive results (Andrew et al., 1990).

The three components of the RNR model — risk of reoffending, criminogenic needs, and responsivity in service delivery — can help guide the type and dosage of rehabilitative intervention for each individual (Andrews et al., 1990). Program evaluation studies have consistently demonstrated that interventions utilizing the risk principle and the RNR model are effective at reducing recidivism (Andrews et al., 2006; Koehler, Lösel, Akoensi, & Humphreys, 2012; Vieira et al., 2009). Guided by the RNR model, risk assessment tools were developed and used in criminal and juvenile justice systems to evaluate and categorize an individual's probability of reoffending. In fact, these tools have evolved over time to address not only risk but other individual-level factors to inform appropriate programs and services according to the RNR model, particularly in rehabilitation-focused juvenile justice systems. However, these

developments have resulted in a variety of types and purposes of risk assessment instruments that complicate the utilization of these tools.

Evolution of Risk Assessments

The history of justice system risk assessments demonstrates an impressive improvement in predictive validity (the ability of the instrument to accurately measure the likelihood of reoffending). The first generation consisted of professional assessments based on personal or clinical judgment without a structured scoring system or empirical support (Schwalbe, 2007). Thus, the ability of these types of assessments to accurately predict reoffending was inferior to the later tools developed via statistical analyses (Grove & Meehl, 1996; Upperton & Thompson, 2007).

Beginning in the 1970s (Bonta & Andrews, 2007), second generation assessments moved toward an evidence-based, actuarial approach with a reliance on statistical relationships between static factors and recidivism (Schwalbe, 2007). These quantitative assessments aimed to classify offenders according to their probability of reoffending; assessments used numerical scores for characteristics such as offense history or family criminality, creating a summed risk score as well as a risk category (e.g., low, moderate, high) based on raw score cut-off points (Schwalbe, 2007). Relying on data availability and statistical associations, second generation tools were often atheoretical (Bonta & Andrews, 2007). Youth risk assessments were also first developed during this time period (Baird et al., 2013).

Considerable evidence consistently demonstrated second generation assessments as superior to the first generation in accurately predicting reoffending (Andrews et al., 2006; Bonta & Andrews, 2007), making them the optimal choice for classifying risk groups and developing practices aligned with the risk principle. Based on the results of a risk assessment, higher risk

individuals could be identified for more intensive interventions. However, these tools were not intended for use in identifying the ideal type or method of intervention (Schwalbe, 2007); they were built without a focus on the theoretical relevance of items, meaning they failed to address the needs or responsivity principles (Bonta & Andrews, 2007).

Third generation assessments extended the scope of the tools to not only classify offenders' likelihood to reoffend based on static factors but also guide interventions by adding dynamic factors that could be targeted for improvements, such as substance abuse, peer delinquency, and school-related problems (Schwalbe, 2007). This incorporation of dynamic risk factors allowed for the assessment of risk changes over time and the development of more appropriate services. While second generation tools were empirically based with a focus exclusively on predicting reoffending, third generation tools extended their content to theoretically grounded risks and needs that could both predict reoffending and reduce risk by informing interventions (Bonta & Andrews, 2007). Thus, third generation instruments were able to identify the factors important to both the risk and the needs principles, thereby expanding the purpose and role of risk assessment instruments. These risk assessments generally utilize information from case files, interviews with juveniles and their families, and information from other sources (e.g., school, service providers) and result in risk categories of low, moderate, or high.

Finally, fourth generation risk assessment instruments expanded again by incorporating items relating to responsivity, or the individual's protective factors and readiness to change (Baird et al., 2013). As the third generation tools informed service provision based on needs, the addition of responsivity items informed service provision based on personality and strengths. With this expanded information, the length and complexity of tools grew; while early

instruments often had a dozen or fewer items, fourth generation instruments often have 42 to 150 items (Baird et al., 2013).

From the second generation of tools forward, the development of actuarial youth risk assessment instruments has followed one of two routes. Some are "home-grown" instruments, customized to data from a specific state or jurisdiction and avoiding the potential challenge of generalizing across locations (Schwalbe, 2007). For example, Arizona developed a five-item instrument with an index of predictive factors using an estimation sample of Arizona first-time offenders and tested the validity of the tool on a separate comparison sample (Krysik & LeCroy, 2002). These types of assessments tend to be brief and focused solely on risk classification (Schwalbe, 2007), often falling into the second generation category of instruments (Baird et al., 2013).

Others are commercially sold, generic instruments that have been validated in different settings (Schwalbe, 2007), often as adapted versions of adult tools (Olver et al., 2009). For example, the Youth Level of Service/Case Management Inventory (YLS/CMI) was developed from the adult Level of Service Inventory – Revised (LSI-R) and assesses risk based on 42 items in eight separate domains (Hoge & Andrews, 2001). These types of tools are generally more comprehensive and incorporate dynamic factors (Schwalbe, 2007), often falling into the third or fourth generation categories (Baird et al., 2013).

Interestingly, a family of commercially sold risk assessment tools originated from a locally developed tool created in 1998 (Taxman, 2016) by the Washington State Association of Public Policy and the Washington State Association of Juvenile Court Administrators, called the Washington State Juvenile Court Assessment (WSJCA; Baglivio & Jackowski, 2013). This tool was created specifically for Washington State, though it was based on prior research and existing

actuarial instruments rather than on the state's own data (Hamilton, van Wormer, & Barnoski, 2015). A case planning process, the Case Management Assessment Protocol (CMAP), was established in 2000 to work in conjunction with the assessment (Washington State Department of Social & Health Services, 2004). The assessment and case planning process was then adapted and rebranded by a proprietary vendor as the Positive Achievement Change Tool (PACT) in Florida (Baglivio & Jackowski, 2013) and similarly adapted by various distributors under different names, including Back on Track (BOT), Iowa Delinquency Assessment (IDA), and the Youth Assessment and Screening Instrument (YASI; Taxman, 2016). Even Washington State now uses the PACT-branded version as their risk assessment instrument (Hamilton et al., 2015). Although originating from a local effort to create a risk assessment tool, this family of instruments is more representative of the third and fourth generation commercially sold options.

The distinguishing feature between second generation tools and third or fourth generation tools was their focus and purpose relating to recidivism. Second generation instruments were meant for categorizations and classifications based on predicted reoffending. These tools were ideal for quicker, one-time evaluations for determining public safety needs at different stages of the justice process (e.g., diversion decisions, dispositions, security level classifications). They often could be completed using records rather than necessitating an interview with the offender and additional information gathering. Second generation tools lacked the ability to inform interventions because of their focus on static factors that, while higher in predictive power, were unchangeable through treatment. Later-generation instruments were meant to provide additional insights to inform targeted services and to track changes over time. Thus, the fourth generation tools moved the risk assessment process from an original focus on the risk principle to the complete RNR model (Bonta & Andrews, 2007). However, this trend from simple classification

toward comprehensively managing and reducing risk through case planning was viewed as ill-advised by some experts, as will be discussed in the next section.

Debates on Best Approaches

While the use of risk assessment instruments is established as a best practice in juvenile justice systems, jurisdictions' selection of the specific tool most appropriate for their system and population is not as straightforward due to the variety of options. The effectiveness of the instrument relies on the strategic implementation decisions of juvenile justice practitioners, but experts disagree on the best approach. Two interrelated decisions must be made by the jurisdiction when selecting a risk assessment option: 1) Should the tool simply predict offending or also inform case planning? 2) Should the tool be a jurisdiction-specific instrument developed using local data or be purchased "off-the-shelf" as a generic instrument?

The first decision revolves around the focus, purpose, and resulting complexity of the tool. Some applaud the advances of third and fourth generation risk assessment tools, believing the addition of needs- and responsivity-based factors contributes to instruments' ability to inform individualized case plans that optimize outcomes based on the RNR model (Andrews et al., 2006). With empirical evidence supporting the RNR model, risk assessment tools that aid compliance to these principles can theoretically improve service delivery and outcomes. As Viljoen, Cochrane, and Jonnson (2018) stated, "predicting if someone will reoffend, in and of itself, has little value if nothing is done to manage risk" (p. 182).

However, others believe that these types of tools are over-complicated without adding to the ability to predict reoffending (Baird, 2009; Baird et al. 2013). Instead, in the real world, juvenile justice practitioners benefit from the simplest instrument that can accurately differentiate offenders into groups by their probability of reoffending. By emphasizing brevity, a

tool remains easy to understand and act upon, without being bogged down by the addition of weakly associated dynamic factors (Baird, 2009; Baird et al., 2013). This perspective emphasizes risk assessment instruments' ability to identify and classify different groups of risk as the most important, if not sole, objective. This objective may be more appropriate for jurisdictions or specific decision points that do not engage in detailed service or intervention planning. Ultimately, the choice is between a "prediction of recidivism" or "reduction of recidivism" (Skeem, Barnoski, Latessa, Robinson, & Tjaden, 2013, p. 109).

However, even if a jurisdiction decides to utilize a third or fourth generation risk assessment tool for case planning purposes, staff may not elect to use the information from the assessment to select appropriate services. Indeed, Latessa and Lovins (2014) identified this phenomenon as one of their "obstacles to good practice" (p. 214). If practitioners successfully implement a risk assessment instrument with high predictive validity and inter-rater reliability but fail to act on the instrument's needs and responsivity factors by providing the same services regardless of results, then the added time, effort, and resources used for third and fourth generation instruments were wasted. Despite these obstacles, the evidence supporting the effectiveness of interventions following the RNR model (Andrews et al., 2006; Koehler et al., 2012; Luong & Wormith, 2011; Vieira et al., 2009) suggests these instruments have promising value if implemented as intended. Thus, a jurisdiction selecting a risk assessment method must determine their intended purpose of the tool (i.e., predicting reoffending or informing interventions) while understanding the implementation challenges of the case planning functions of third and fourth generation tools.

The second decision focuses on the origins and development of the tool. Some instruments were developed using the specific jurisdiction's data on youth characteristics and

reoffending outcomes, customizing an algorithm to produce a risk score (Schwalbe, 2007); these tools tend to be second generation tools that are simpler and statistically determined (Baird et al., 2013). Others are purchased generic tools adapted from other tools and/or validated in other locations (Schwalbe, 2007); these tools more often incorporate the needs and responsivity elements of third and fourth generation tools to aid in case planning (Baird et al., 2013). Many jurisdictions may not have the resources or expertise to create the complex third and fourth generation tools; even in the case of Washington State, the instrument became commercially sold due to the need for software development by a third party (Hamilton et al., 2015). Thus, these two debates on purpose and origin have historically been linked.

Across the country, juvenile justice systems currently use a mixture of generic and home-grown risk assessment instruments. Juvenile Justice Geography, Policy, Practice, & Statistics reported that out of the 45 states (including Washington, D.C.) using a youth risk assessment tool, 24 used one or more of the most popular off-the-shelf tools (i.e., YLS/CMI, YASI, Structured Assessment of Violence Risk in Youth [SAVRY], PACT), and at least eight states used a tool with a name suggesting it was self-created (i.e., the state name was in the instrument's title; Juvenile Justice Geography, Policy, Practice, & Statistics, 2017). Similarly, a study of ten states' juvenile risk assessment tools yielded seven commercially purchased instruments and three state-generated instruments (Baird et al., 2013).

Despite the prevalence of commercially available tools, risk assessment instruments may not always generalize across populations (Schwalbe, 2007), leading to the potential for an imprecise tool in a jurisdiction that purchased one of these tools. Research is mixed regarding the predictive validity of risk assessment instruments being applied to populations other than the one for which it was originally constructed (e.g., Baird et al., 2013; Jones, Harris, Fader, &

Grubstein, 2001; Miller & Lin, 2007). These inconsistent findings indicate there is no clearly superior method for developing or implementing a risk assessment tool. Skeem and Monahan (2011) concluded there is "little evidence that one validated instrument predicts violence significantly better than another" (p. 40) because they all contain the same common factors with variations in specificity, wording, or format. Although one tool or approach does not consistently outperform another, it is clear that some risk assessment tools fail to accurately predict reoffending for specific populations (e.g., Jones et al., 2001, Miller & Lin, 2007), perhaps due to the nuances of those variations applied to different populations in different locations in different times. Differences in staff training, system processes, data availability, and fidelity to the instrument may impact the predictive accuracy. Schwalbe (2007) emphasized that "few instruments have been validated in multiple samples" (p. 459), adding to the variability in validation results. Thus, jurisdictions should consider their resources and needs when deciding to purchase or create a risk assessment tool. Then, more importantly, they must validate and periodically revalidate the instrument for the specific jurisdiction in which it is used to ensure its appropriateness among the specific system and juvenile population served. It cannot be taken for granted that a generic tool will work for a specific population or even that a home-grown tool will remain accurate over time.

In conclusion, juvenile risk assessment tools vary in their purpose (predicting reoffending versus informing interventions) and origin (locality-created versus commercial or generic), resulting in differences in length, purpose, accuracy, and usability. Likewise, states and localities vary in their populations (e.g., low risk versus high risk), policies and procedures (e.g., staff training, programs and services, workloads) and goals (e.g., predicting recidivism, identifying services) related to risk assessments. A single instrument is not the "right" tool for every

jurisdiction, resulting in different jurisdictions utilizing different types of risk assessment tools based on their resources and needs. In the face of myriad risk assessment options, each jurisdiction must consider its own characteristics and goals from the risk assessment tool in order to select and implement the ideal instrument for its specific needs. In particular, the items included in third and fourth generation tools may be superfluous in a system that does not require case planning or does not properly utilize the tool for service matching. Additionally, the predictive validity and case planning application of off-the-shelf tools may not be transferrable across jurisdictions with different system structures and different populations. Thus, after implementation, it is imperative that jurisdictions adequately monitor and test their selected tool to assess whether it is meeting their goals. The next section will describe one such jurisdiction and its risk assessment instrument.

Juvenile Risk Assessment in Virginia

In Virginia, the Department of Juvenile Justice (DJJ) operates and/or certifies 34 court service units and one juvenile correctional center (DJJ, 2016; DJJ, 2017). The court service units oversee juvenile intakes; deciding whether to petition, detain, divert, or resolve a case; as well as probation and parole supervision. The juvenile correctional center houses juveniles who have been committed by the courts to the state and have been admitted to direct care; these juveniles may be committed indeterminately (i.e., DJJ decides the length of stay [LOS]) or determinately (i.e., the court decides the LOS for eligible offense severities), and some may have a blended sentence with a combined juvenile commitment and adult sentence (DJJ, 2016). Each of these agency responsibilities requires judgments and planning based on the juveniles' risks, needs, and strengths to provide appropriate services to promote successful outcomes (DJJ, 2016).

YASI, a fourth generation instrument, is used as a risk assessment and case planning tool throughout all stages of the juvenile justice system in Virginia (DJJ, 2016). As mentioned above, YASI evolved from the WSJCA and is sold by Orbis Partners Inc. (Orbis; Orbis, 2007; Taxman, 2016). It was originally modified for New York State's juvenile probation departments but has since been customized and used by jurisdictions in nine states (Orbis, 2007). The tool is tailored to each system and its terminology, and over time, items, weighting, scoring, and even the domains have shifted based on further analysis (Orbis, 2007).

YASI can be completed as a pre-screen or a full assessment (Orbis, 2007) and is used in multiple stages of the juvenile justice system in Virginia to support data-driven decisions and evidence-based practices (DJJ, 2017). First, YASI is sometimes used by the DJJ intake officer to inform the decision to petition or divert a case; often, the pre-screen is completed at this phase (R. Harris, personal communication, October 30, 2017). According to Orbis (2007), the pre-screen is meant to identify high risk cases in need of intervention and inform whether a full assessment is needed for further case planning. In Virginia, the YASI full assessment is completed as part of the Social History Report, which is compiled as either a court-ordered document to inform dispositions or after placement on probation or commitment to the state (DJJ, 2016).

Based on the results of the YASI, DJJ staff determine the level and intensity of supervision and develop a case plan for juveniles on probation and parole directly in the YASI Caseworks software application (DJJ, 2013). The case plan includes the YASI priority domains as well as targets, long-term goals, short-term goals, and action steps (DJJ, 2013). According to procedure, the case plan should focus on the dynamic risk factors related to the juvenile's offenses and should "prioritize criminogenic needs, reduce risk, address skill deficits, and build

competencies in order to break the cycle of offending and prevent recidivism" (DJJ, 2013, p. 2-3). Finally, for juveniles committed to the state indeterminately, a combination of the YASI overall risk level, dynamic risk level, and dynamic protective level along with the most serious committing offense are used to assign the LOS for their direct care stay (DJJ, 2016). Beginning in January 2013, juveniles in direct care or on probation or parole supervision were scheduled to be reassessed every 180 days (DJJ, 2016).

In fiscal year (FY) 2016, 5,848 initial YASIs (i.e., the first assessment completed for an individual youth, regardless of the point in the system) were completed by court service units in Virginia (DJJ, 2016). Of those, 48.6% were classified as none/low risk, 41.0% as moderate risk, and 10.4% as high risk (DJJ, 2016). Of the 3,647 probation placements in FY 2016, 23.8% were classified as none/low risk, 50.6% moderate risk, and 21.1% high risk, with 4.4% missing assessments (DJJ, 2016). Of the 319 admissions to direct care, 0.9% were classified as none/low risk, 23.8% moderate risk, and 74.0% high risk, with 1.3% missing assessments (DJJ, 2016). Finally, of the 283 parole placements, 2.5% were classified as none/low risk, 36.4% moderate risk, and 59.0% high risk, with 2.1% missing assessments (DJJ, 2016).

Supporting YASI's ability to assess relative risk of reoffending, DJJ's recidivism analysis demonstrated that high risk samples had the highest recidivism rates (DJJ, 2016). Using direct care releases as an example, 47.5% of high risk juveniles were reconvicted while 34.8% of moderate risk and 29.4% of none/low risk juveniles were reconvicted (see Table 1 for additional rates).

Table 1. 12-Month Rearrest Rates in Virginia by Risk Level, FY 2015 Cohort

	Probation Placements	Direct Care Releases	Parole Placements
High	54.7%	55.2%	61.0%
Moderate	36.1%	47.9%	51.8%
None/Low	16.0%	25.0%	30.0%

(DJJ, 2016)

Additionally, a higher percentage of juveniles on parole following commitment (i.e., youth in the deepest end of the system) had high risk levels (59.0%) compared to juveniles on probation (21.1%), and both of these groups had higher percentages of high risk levels compared to juveniles being assessed for the first time (10.4%; DJJ, 2016). However, rearrest rates have remained relatively stable since the full implementation of YASI-based case planning in 2013, suggesting the use of the tool has not impacted recidivism outcomes (see Table 2; DJJ, 2016).

Table 2. 12-Month Rearrest Rates in Virginia, FY 2011-2015 Cohorts

	FY 2011	FY 2012	FY 2013	FY 2014	FY 2015
Probation	35.7%	37.2%	34.2%	34.1%	33.9%
Direct Care	48.6%	50.0%	51.7%	49.5%	51.5%
Parole	54.4%	57.2%	61.5%	58.7%	58.0%

(DJJ, 2016)

Summary of the Problem

It is not a question of whether juvenile justice systems should use a risk assessment tool; however, the type of tool that is best for a particular jurisdiction is under debate and requires additional and individualized research. As described above, Virginia's juvenile justice system uses YASI, a commercially sold fourth generation risk assessment tool, extensively for the purposes of 1) accurately predicting the likelihood of reoffending, 2) standardizing decisions by using a consistent tool, and 3) informing interventions for effective rehabilitation. Given the questionable generalizability of off-the-shelf tools to various populations and the debate about the most appropriate purpose of risk assessment instruments, it is imperative to confirm the

tool's accuracy, equity, and usage in Virginia. The study of these factors will inform whether Virginia's selection and implementation of YASI is appropriately serving its purpose and suggest implications for public policy administrators in RNR-focused juvenile justice systems.

In order to provide background and support for the current study, the following chapter will discuss previous research concerning common risk factors for juvenile reoffending and risk assessment instruments' predictive validity. It will then narrow focus to the more specific aims of the current study by describing research focused specifically on YASI, possible sex and race differences related to risk assessments, and practitioners' application of risk assessment results to service delivery.

Chapter 2: Review of Previous Research

Risk Factors for Reoffending

Risk assessment instruments, regardless of generation, purpose, or origin, are based on common risk factors for offending. Throughout the multitude of studies of adolescent characteristics associated with delinquent or criminal behavior, several general domains consistently predict reoffending: offense history, psychological characteristics, family and social history, and peer associations. These factors appear across risk assessment tools, whether developed based on conceptual frameworks or actuarial findings.

One of the primary models for criminogenic risk factors was established by the same developers of the RNR model. Following the conceptual framework of the RNR model and findings from meta-analyses of criminal risk factors, Andrews and colleagues (2006; Andrews & Bonta, 2010) described the "Central Eight" risk factors: 1) history of antisocial behavior, 2) antisocial personality pattern, 3) antisocial cognition, 4) antisocial associates, 5) family/marital circumstances, 6) school/work, 7) leisure/recreation, and 8) substance abuse (see Table 3). The first four represent the "Big Four," with strong predictive relationships with reoffending while the latter four represent the "Moderate Four," with additional but weaker predictive power (Andrews et al., 2006; Andrews & Bonta, 2010). Of the Big Four, three are dynamic risk factors that could be addressed in treatment: antisocial personality pattern, antisocial cognition, and antisocial associates (McGrath & Thompson, 2012). Conversely, the Moderate Four are dynamic factors that can increase opportunities for criminal behavior (Andrews & Bonta, 2010). Thus, seven of the eight factors are dynamic risk factors, and these criminogenic needs may be targeted through interventions to reduce the likelihood of reoffending (Bonta & Andrews, 2007). In fact, Andrews and colleagues (2006) list dynamic need targets for each risk factor (e.g., "build

problem-solving skills, self-management skills, anger management and coping skills" [p. 11] for antisocial personality pattern).

Table 3. Central Eight Risk Factors from Andrews et al. (2006)

Risk Factor	Static or Dynamic
Big Four	
1- History of antisocial behavior	Static
2- Antisocial personality pattern	Dynamic
3- Antisocial cognition	Dynamic
4- Antisocial associates	Dynamic
Moderate Four	
5- Family/marital circumstances	Dynamic
6- School/work	Dynamic
7- Leisure/recreation	Dynamic
8- Substance abuse	Dynamic

The LSI-R (Andrews & Bonta, 1995) is an adult risk assessment tool developed to reflect the Central Eight risk and needs factors. In multiple studies, the LSI-R was found to have higher predictive validity than competing instruments for adults (Gendreau, Goggin, & Smith, 2002; Gendreau, Little, & Goggin, 1996). The YLS/CMI applied this tool to adolescent offenders (Hoge, & Andrews, 2001) and accurately predicted reoffending in a number of validation studies, though with variations in overall effect sizes (Schwalbe, 2007; Olver, Stockdale, & Wong, 2012; Olver et al., 2009) and predictive incremental validity of factors (Grieger & Hosser, 2014; Andrews & Bonta, 2010).

As Skeem and Monahan (2011) stated, most risk assessment instruments contain the same common factors. Given the close connection of the Central Eight to the development of the RNR model that risk assessment tools are intended to address, it is not surprising that the domains of YASI (*Legal History*, *Family*, *School*, *Community/Peers*, *Alcohol/Drugs*, *Mental*

Health, *Aggression*, *Attitudes*, *Skills*, *Employment/Free Time*; see Appendix A for an outline of the domains' items) closely align with the Central Eight factors (see Table 4).

Table 4. Domains of YASI, Mapped onto Central Eight Risk Factors

Domain	Related Central Eight Factor(s)
Legal History	1- History of antisocial behavior
Family	5- Family/marital circumstances
School	6- School/work
Community/Peers	4- Antisocial associates
Alcohol/Drugs	8- Substance abuse
Mental Health*	N/A
Aggression	2- Antisocial personality pattern
Attitudes	3- Antisocial cognition
Skills	2- Antisocial personality pattern
Employment/Free Time	6- School/work; 7- Leisure/recreation

Note: The *Mental Health* YASI domain results in a flag to indicate a need for further assessment and/or intervention without indicating an increased risk.

In addition to the theoretically based models, similar risk factors for reoffending were identified across studies. Approaching predictive risk factors from a purely actuarial approach rather than a conceptual framework, a meta-analysis of 23 studies of delinquent youth identified 30 predictor variables (Cottle, Lee, & Heilbrun, 2001). These factors, statistically associated with recidivism, merged with the Central Eight concepts. (See Table 5 for the variables independently positively associated with reoffending. See Table 6 for the meta-analysis domains mapped onto Central Eight factors.) Some exceptions, including the demographic characteristics found in the meta-analysis and the antisocial cognitions from the Central Eight, were not included in their counterparts, suggesting slight variations in predictors. Additionally, Cottle and colleagues (2001) found that a composite risk score from any formal risk assessment instrument was positively associated with recidivism, further supporting common risk factors across measures.

Table 5. Predictive Factors from Cottle et al.'s (2001) Meta-Analysis of Juvenile Recidivism

Domain	Factors	Static or Dynamic
Demographic	Male	Static
	Minority race (not significantly related after socioeconomic background accounted for)	Static
	Low socioeconomic background	Static
Offense History	Earlier age of first contact with the law	Static
	Earlier age at first commitment	Static
	More prior arrests	Static
	More previous commitments	Static
	Longer incarcerations	Static
	Committed more serious crimes	Static
Family and Social	History of physical or sexual abuse	Static
	Raised in a single-parent home	Static
	Greater number of out-of-home placements	Static
	Significant family problems	Dynamic
	Ineffective use of leisure time	Dynamic
	Delinquent peers	Dynamic
Educational Standardized Test Scores	History of special education	Static
	Lower standardized achievement score	Static/Dynamic
	Lower full-scale IQ score	Static
	Lower verbal IQ score	Static
Substance Use History	Substance abuse	Dynamic
Clinical	History of conduct problems	Dynamic
	History of non-severe pathology	Dynamic

Table 6. Domains from Cottle et al. (2001), Mapped onto Central Eight Risk Factors

Domain	Related Central Eight Factor
Demographic	None
Offense History	1- History of antisocial behavior
Family and Social	4- Antisocial associates 5- Family/marital circumstances 7- Leisure/recreation
Educational Standardized Test Scores	6- School/work
Substance Use History	8- Substance abuse
Clinical	2- Antisocial personality pattern

Predictive Validity

As risk assessments moved away from the first generation, clinically or professionally based judgments toward actuarial tools, predictive validity became a more important measure of a useful instrument. Without accuracy in practice, the development and use of an assessment instrument is without real-world value. Studies of youth risk assessment tools' predictive validity are discussed below. Due to the large number of instruments available, general findings are discussed rather than focusing on specific tools.

In a meta-analysis of 28 studies of second and third generation juvenile risk assessment tools' predictive validity, Schwalbe (2007) found that these instruments as a whole predicted recidivism almost as well as adult risk assessment tools, with the more comprehensive third generation tools tending to perform better than second generation instruments. Other studies and meta-analyses examining specific tools generally found similar results of acceptable predictive validity, with youth assessment instruments performing comparably to adult measures (see Olver et al., 2009). Even tools with varying purposes had similar items and predictive performances;

for example, the YLS/CMI was developed as a measure for predicting general recidivism, the Psychopathy Checklist – Youth Version (PCL-YV) was developed as a measure for diagnosing psychopathy, and the SAVRY was developed for predicting violence, yet their content overlapped considerably, and their results predicted both general and violent reoffending (Olver et al., 2009).

In contrast to Schwalbe's (2007) findings that longer third generation tools performed better than second generation tools, a study of ten states' instruments found that simplified, shorter tools built from actuarial analysis of the existing assessments performed similarly to their lengthier counterparts (Baird et al., 2013), indicating they might be an optimal option for systems low in resources or focused on only the predictive classifications of risk. Coid and colleagues (2011) also found that several individual items on adult risk assessment tools did not predict violent recidivism, and the predictive validity of the tool as a whole did not diminish with their removal. These findings indicate that some items on risk assessment instruments could be superfluous, or, at the very least, provide diminishing returns to the predictive ability of the tool. Thought must be given, though, to whether these items are included for the primary purpose of risk prediction or are utilized more as guidance for interventions for recidivism reduction, in which their lack of predictive ability is not as problematic.

Evidence is mixed for superior predictive validity in youth risk assessment instruments created for specific jurisdictions versus generic versions sold commercially. While Baird and colleagues (2013) found commercial instruments among the tools with the highest predictive validity in their comparison of ten jurisdictions, other studies (e.g., Jones et al., 2001; Miller & Lin, 2007) found a generic instrument's application to a specific jurisdiction resulted in failure to predict reoffending accurately. For example, Jones and colleagues (2001) tested the risk factors

of a juvenile risk assessment for chronic offending developed empirically using data from Orange County, California, to a population of young offenders in Philadelphia. They found that few of Orange County's constructs were statistically associated with chronic offending in the Philadelphia sample. Of 29 items, only two (instrumental communication and prior dependency referral to Department of Human Services) predicted chronic offending at the significance level of 0.05 (Jones et al., 2001). Similarly, Miller and Lin (2007) found that even a generic juvenile risk assessment tool designed to be used across jurisdictions lacked the ability to predict recidivism when applied to a New York City population, performing worse than both a locally developed tool and probation officers' clinical judgment. The generic tool used risk factors repeatedly validated in 13 jurisdictions, and the tool itself had been validated in two published studies; however, the authors speculated that differences in population demographics, specificity and availability of data, and system differences affected the predictive validity of the generic tool (Miller & Lin, 2007).

Thus, juvenile risk assessment instruments often demonstrate acceptable levels of predictive validity without wide variation between tools, with some exceptions for specific applications. The literature provides inconsistent findings regarding the superiority of a tool based on its length and complexity as well as instruments' generalizability to various populations. Therefore, the lack of reliable conclusions regarding predictive validity necessitates the individualized study of a tool for a particular jurisdiction.

YASI-Specific Research

According to Orbis, their collection of assessment products (including YASI and others) has been validated "in a number of large jurisdictions in both community and residential settings and shows excellent predictive accuracy and reliability…the instrument is effective for both girls

and boys [and] youth from different ethnic backgrounds" (Orbis, 2017b), including New York, California, Illinois, and Alberta, Canada (Orbis, 2017a). Orbis (2007) acknowledges customization in content and scoring for a jurisdiction requires the tool to be revalidated in the new setting.

However, scarce information regarding YASI exists in peer-reviewed publications beyond studies by Orbis (2007) and Canadian Master's theses (Geck, 2012; Harris, 2013; Jones, 2011). This limitation may be a result of the origins of the tool; as discussed, it was created in 1998 for Washington State but was adapted and rebranded by a proprietary vendor as the PACT in Florida (Baglivio & Jackowski, 2013) and also similarly adapted by various distributors under different names, including BOT, IDA, and YASI (Taxman, 2016). Therefore, peer-reviewed research focused specifically on YASI rather than the more commonly studied PACT instrument is rare.

Thus, only four studies were located that focused on predictive accuracy of YASI, utilizing populations from New York (Orbis, 2007; Jones, 2011), Canada (Jones, Brown, Robinson, & Frey, 2016), and Virginia (Baird et al., 2013). They used various timeframes, population breakdowns (e.g., sex, race/ethnicity, program type), and definitions of recidivism. Overall, the area under the receiver operating characteristic (ROC) curve (AUC) values, a measure of predictive validity ranging from zero to one (discussed in more detail in Chapter 3), ranged from 0.64 to 0.79 for the pre-screen overall risk and 0.62 to 0.63 for the full assessment of dynamic risk. Using guidelines from Rice & Harris (2005), these AUCs varied from small to strong; all but one test of the pre-screen risk score (Jones et al., 2016) fell short of the 0.70 cut-off for acceptable suggested by van der Put and colleagues (2011). However, the AUCs reported for YASI (Orbis, 2007; Jones, 2011; Baird et al., 2013) were comparable to other risk assessment

tools examined in a meta-analysis, ranging from 0.53 to 0.78 (Schwalbe, 2007), and the authors generally found the results to be acceptable. These limited studies on YASI predictive accuracy are summarized in Table 7, and specifics are described below.

Table 7. YASI AUC Results from Previous Studies

	Orbis, 2007	Jones, 2011	Jones et al., 2016	Baird et al., 2013
Location	New York	New York	Canada	Virginia
Population	Probation	Probation	Community Supervision	Probation
Recidivism Measure	Negative Outcome	Reconviction	Rearrest	Reconviction
Follow-up	24 months	24 months	18 months	12 months
Pre-Screen	0.65	0.64	0.79	0.68
Female	0.61	0.60	0.68	0.67
Male	0.68	0.64	0.82	0.71
Black				0.66
White				0.68
Dynamic Risk	0.62	0.63		
Female	0.59	0.62		
Male	0.64	0.63		
Dynamic Protective	0.60			
Female	0.59	0.59		
Male	0.60	0.59		
Domain Dynamic Risk	0.55-0.63		0.54-0.73	
Female		0.50-0.60		
Male		0.50-0.62		
Domain Dynamic Protective	0.50-0.59		0.61-0.68	
Female		0.52-0.60		
Male		0.50-0.58		

Note: Studies may have reported additional AUCs not displayed in the summary table above. Orbis (2007) and Jones (2011) both studied New York populations, resulting in similar AUCs. In the Orbis (2007) study, "negative outcome" was defined as a new referral/arrest, violation of probation, or adjudication/conviction; AUCs for the pre-screen after item weight and cut-off point revisions are displayed.

Orbis (2007) completed a validation study of their instrument in New York State using a sample of juveniles on probation. In this study using several methods of measuring recidivism, the authors found a "sufficient level of predictive accuracy" (p. 6-1) and stated the results were "very promising [and] most encouraging" (p. 6-2) for the pre-screen risk score and full assessment's overall dynamic risk score. The overall dynamic protective scores performed slightly worse. Of the domain dynamic risk scores, *Community and Peers* and *Attitudes* performed best, and *Alcohol and Other Drugs*, *Skills*, and *Free Time* performed worst (Orbis, 2007). Of the domain dynamic protective scores, *Employment* performed worst, resulting in plans to combine with the *Free Time* domain. Based on these findings, Orbis (2007) adjusted the pre-screen scoring method from the matrix-based scoring used in the Washington model to a simple additive scoring and modified the weighting of items, resulting in an assessment more similar to Virginia's (DJJ, 2016). Jones (2011) completed a master's thesis on the same or similar population in New York, resulting in similar findings.

The recent and only known peer-reviewed article, authored in part by Orbis staff, studied the validity of YASI for juveniles on community supervision in Alberta, Canada (Jones et al., 2016). This study resulted in the highest AUCs, with the male AUC exceeding the maximum found in a previous meta-analysis (Schwalbe, 2007).

The use of YASI in Virginia was studied near the beginning of its implementation as part of a multi-state study (Baird et al., 2013). Using data from the first year of piloting, they found that YASI performed relatively well regarding inter-rater reliability and validity. However, due to the stage of implementation within the state, the study was limited to a subsample of localities and to juveniles on probation. Now, several years into Virginia's YASI implementation, the tool is used throughout all localities in the state and throughout all stages of the system from intake to

parole. Thus, it is appropriate to reexamine the use of YASI in the Virginia juvenile justice system.

Group Differences

A discussion of youth risk assessment instruments must include note of their application to sex[1] and racial subgroups. Most tools were created and tested on predominantly male populations (Anderson et al., 2016); however, females and minority groups often experience different rates of comorbid risk factors (e.g., trauma, victimization, education) than White males (Shepherd, Luebbers, & Dolan, 2013; Thompson & McGrath, 2012). Additionally, the system itself treats females and minorities differently from the majority population in terms of arrests, diversions, sanctions, etc., thereby impacting the offense history risk factors evaluated in the risk assessment tools as well as future criminal trajectories (Shepherd et al., 2013; Thompson & McGrath, 2012). Schwalbe and colleagues (2004) found that a risk assessment tool predictive for the entire population sample was no longer a consistently significant model when the population was divided by race/ethnicity and sex.

Thus, it is important to study the validity of youth risk assessment instruments for sex and racial subgroups to assure accuracy and equity across the population. Any inequities in the predictive accuracy of the tool by either sex or race could impact the decision-making regarding those cases, resulting in potentially unfair consequences for youth in these subgroups. These consequences may include an overly severe punishment not warranted by the true risk profile or a lack of services for youth with higher needs. Given the existing disparities in juvenile justice

[1] Many studies use "gender" instead of "sex" when discussing differences between females and males; however, "sex" is used throughout this study to be consistent with the term used by DJJ. DJJ collects data for "sex assigned at birth" (sex) rather than "gender identity" (gender). None of the reviewed literature studied non-binary sex or gender groups.

systems, it is imperative that any evidence-based tool used to inform decisions must not contribute and compound these disparities. Differences between groups in predictive validity for overall scores and domains may indicate inequities in the tool, requiring adjustments to ensure the instrument is equitable and free from discrimination.

Sex

Studies on risk assessments by sex have generally demonstrated comparable predictive validity between males and females (Jones, 2011; Olver et al., 2009; Orbis, 2007; Pusch & Holtfreter, 2018; Schwalbe, 2008; Shepherd et al., 2013; Smith, Cullen, & Latessa, 2009), though cut-off points may be adjusted to account for different base rates of reoffending (Baird et al., 2013; Orbis, 2007). YASI was developed as a gender-neutral instrument (Orbis, 2007), but in a breakdown by predictive validity by sex on an earlier version of the tool, AUCs for females were generally slightly lower than for males, with females over-classified as high risk (Orbis, 2007). Based on these findings, Orbis (2007) modified the overall and domain scoring cut-off points to differentiate between sexes. However, Jones and colleagues (2016) found a disparity in predictive accuracy between males and females even after these modifications, with the predictive validity for males exceeding that for females (AUCs of 0.82 and 0.68, respectively).

Some experts, though, argue for a completely separate risk assessment tool, customized for females and their specific risks and needs rather than simply changing cut-off points (Emeka & Sorenson, 2009). Risk factors of trauma, sexual and domestic abuse, substance abuse, and economic disadvantage are more prevalent among females (Reisig, Holtfreter, & Morash, 2006). Daly (1992, 1994) presented various pathways to criminal behavior among women based on abuse history, substance addiction, family relationships, and economic status rather than the male criminological theories. In addition to the pathways and risk factors, the actual offense

characteristics also differ between sexes, with boys exhibiting more chronic, serious, and violent offending than girls (Chesney-Lind, Morash, & Stevens, 2008). Finally, females may have some gender-specific criminogenic needs (Hollin & Palmer, 2006), but gender-responsive programming for girls is lacking in both availability and evidence of effectiveness (Chesney-Lind et al., 2008). Thus, feminist criminologists argue that pathways to criminal behavior, presentation of that criminal behavior, and effective rehabilitative services are unique for females (Reisig et al., 2006), and the risk assessments built around male criminological theory and tested on largely male populations are inappropriate for females.

Indeed, some studies have found disparate predictive validity of risk factors and instruments between sexes (Anderson et al., 2016; Jones et al., 2016; Reisig et al., 2006). Risk domains related to interpersonal relationships, particularly families, tend to be more predictive of offending for females than males. Family (Anderson et al., 2016; Gavazzi, Yarcheck, & Chesney-Lind, 2006; Hubbard & Pratt, 2002; Schmidt, Campbell, & Houlding, 2011; van der Put et al., 2014), trauma or abuse (Gavazzi et al., 2006; Hubbard & Pratt, 2002), and peer relationship factors (Gavazzi et al., 2006) were found to be strong predictors for girls. Using Daly's (1992, 1994) female criminogenic pathways, Reisig and colleagues (2006) found that a risk assessment instrument was accurate for only the subset of economically disadvantaged females; the instrument did not predict recidivism among the women categorized under a gendered pathway (i.e., "street women," "drug-connected," "harmed and harming," or "battered"). Given the differences in both the predictive risk domains between sexes and the accuracy among female pathway subgroups, there is some support for the feminist critiques of gender-neutral risk assessment tools.

In Virginia, recidivism rates are generally lower for females than males (DJJ, 2016). For example, most recent one-year rearrest rates for females were 24.5% for probation placements and 44.4% for direct care releases while the rates for males were 36.8% for probation placements and 52.2% for direct care releases (DJJ, 2016). Therefore, YASI in Virginia may result in similar over-classifications as were found in other jurisdictions if cut-off points are not appropriate (Orbis, 2007), with further modifications necessary if domains differ in their predictive abilities.

Race

Minorities are overrepresented at every stage of the juvenile justice system nationwide, and the effect is cumulative as youth move through the decision points (Kakar, 2006). As evidence of the severity of this problem, the Juvenile Justice and Delinquency Prevention Act of 1974 was amended in 1989 to include earmarked funding for states to tackle the issue (Kakar, 2006). The subsequent reauthorization of the act in 1992 further emphasized the problem by making the issue one of four core requirements, meaning a failure to comply would mean a loss of funding for the state (Kakar, 2006). The focus on disproportionate minority *confinement* then shifted to disproportionate minority *contact* in 2002 to indicate the issue was present across the system (Kempf-Leonard, 2007).

The higher rate of minorities involved in the justice system may be the result of several factors. Many risk factors are more prevalent among minority communities, including poverty, exposure to violence, school failure, mental disorders, and other socio-cultural factors (Chapman, Desai, Falzer, & Borum, 2006), potentially making minorities more likely to become involved in the juvenile justice system and to score higher on a risk assessment tool. Disproportionality may also be a product of systematic biases; for example, targeted police surveillance in poor, minority neighborhoods may result in more frequent Black arrests (Kempf-

Leonard, 2007). Discrimination is another factor in disproportionality; a meta-analysis of studies on race and arrests indicated that Blacks were 30% more likely to be arrested even after controlling for other relevant factors (Kochel, Wilson, & Mastrofski, 2011). While this short overview of possible contributors to racial disproportionality is far from comprehensive, it is clearly a widespread and complex problem.

Given the disproportionate representation of minorities in the justice system, a risk assessment instrument that informs placements and services has the ability to exacerbate the disproportionality (Schwalbe, Fraser, Day, & Cooley, 2006). The higher base rate for offending among minority groups may then compound the disproportionality problem by increasing the *Legal History* domain risk score on an assessment used for decision-making, resulting in deeper penetration into the system. As Thompson and McGrath (2012) note, "The direct and indirect effects of such factors may result in…'cumulative disadvantage' within the juvenile justice system" (p. 346). However, studies on the predictive validity of risk assessments by race have been even less consistent than those on sex, with some finding statistically significant differences between groups and others not (Thompson & McGrath, 2012). Orbis claims YASI is appropriate for all races (Orbis, 2017b) and found minimal over-classification of a racial group (Orbis, 2007).

Similar to sex, recidivism rates by race differ in Virginia. For example, one-year rearrest rates for White youth were 28.5% for probation placements and 44.2% for direct care releases while the rates for Black youth were 40.2% for probation placements and 55.1% for direct care releases (DJJ, 2016).

Usage in Case Planning

In addition to a risk assessment instrument's overall predictive accuracy and equity regarding sex and race subgroups, its effective usage in case planning is essential for any third or fourth generation tool. The longer, more complex assessments are intended to inform interventions and reduce the risk of reoffending; however, this purpose is complicated by the apparent lack of juvenile case management decision-making that occurs in response to third and fourth generation risk assessment instruments.

Despite evidence supporting the effectiveness of interventions following the RNR model (Andrews et al., 2006; Koehler et al., 2012; Vieira et al., 2009) and the additional factors included in third and fourth generation instruments, Singh and colleagues (2014) found that in a sample of assessed youth, only half were provided services targeting their identified vulnerabilities, and a quarter were provided services targeting their identified strengths. Similarly, in Vieira and colleagues' (2009) sample of evaluated youth, an average of only 35% of juveniles' identified criminogenic needs and 26% of juveniles' responsivity factors were matched to their treatment services.

Staff understanding and commitment to the tool may be a cause of this failure. In a systematic review of youth and adult risk assessments in correctional and psychiatric settings, practitioners were mixed in their buy-in and utilization of the tools for risk management purposes, including some who responded they did not use the tool even when required by their organization (Viljoen et al., 2018). Service matching according to identified needs was rated as mixed to low, and evidence that the implementation of a risk assessment reduced offending was insufficient – an unsurprising finding given the lack of compliance and utilization (Viljoen et al., 2018). The results of these studies indicated the more complex tools were not being utilized as

intended, and data collection was occurring that did not contribute to the decision-making of cases.

Thus, an accurate risk assessment instrument selected with the intention of reducing reoffending through RNR-based service matching may not be implemented properly. Without the appropriate application to the intervention stage, these types of tools waste the time and resources required to complete the additional needs and responsivity elements. Therefore, the success of the instrument's impact on reoffending relies on its usage in case planning and service delivery.

Chapter 3: Method

Restatement of the Problem

DJJ uses YASI, a fourth generation instrument, as a risk assessment and case planning tool throughout all stages of the juvenile justice system (DJJ, 2016). The only study of its predictive validity in Virginia used data from the first pilot year (Baird et al., 2013). Thus, additional research is necessary to determine if YASI in Virginia is serving its purpose of 1) accurately predicting the likelihood of reoffending, 2) standardizing decisions by using a consistent tool, and 3) informing interventions for effective rehabilitation. Three research questions were developed to evaluate these purposes and inform DJJ's future policy decisions regarding their use of risk assessments.

Research Questions

The current study focused on three important concepts relating to the implementation of a youth risk assessment instrument in a juvenile justice setting: accuracy in predicting reoffending, equity in predictive validity for different sex and racial groups, and usage as a case planning tool.

Accuracy: What is the ability of YASI risk levels to predict reoffending in DJJ's populations?

The YASI risk levels are used by staff to make decisions about dispositions, intensity of supervision, and case planning. Therefore, the YASI's overall risk level and overall dynamic risk level should accurately predict the likelihood of recidivism. It is important to note that the overall risk level is computed from the pre-screen while the overall dynamic risk level is computed from the full assessment; therefore, the predictive accuracy of both overall levels was calculated. Based on previous studies of YASI predictive validity (see Table 7), it was hypothesized that the overall and dynamic risk levels would result in AUCs in the mid-0.60s (Hypothesis #1).

Equity: Do sex or race group differences exist in predictive validity of overall and dynamic risk levels and domain dynamic risk levels?

The YASI level cut-off points are different for males and females (identified in Virginia as the sex assigned at birth rather than gender identity), but otherwise, the instrument has uniform scoring across groups. However, due to the implications of the decision-making based on the YASI levels (e.g., case planning, supervision levels), equity between groups on the tool is imperative to ensure it is not contributing to disproportional treatment. Additionally, different groups may experience the influence of risk and protective factors differently, impacting the predictive validity of the scoring algorithm. Based on previous findings of a lower predictive ability of YASI for females (Orbis, 2007; Jones et al., 2016), it was hypothesized that YASI in Virginia would better predict reoffending among males than females (Hypothesis #2a). Based on the research on sex differences in the importance of risk factors (Anderson et al., 2016; Gavazzi et al., 2006; Hubbard & Pratt, 2002; Schmidt et al., 2011; van der Put et al., 2014), it was hypothesized that the YASI domains of *Family* and *Community/Peers* would be stronger predictors of recidivism for females than males (Hypothesis #2b). Without consistent findings concerning risk assessment differences between racial groups, it was hypothesized that YASI would perform comparably for White and Black youth (Hypothesis #2c).

Usage: Are assigned case planning priority areas congruent with higher dynamic risk level domains?

Up to three domains are designated by the case planner as high priority areas for the juvenile. These priorities are meant to guide the service delivery with the aim of improving outcomes. Thus, the changeable criminogenic needs (i.e., dynamic risks) of the individual should inform the priority areas. Based on previous findings supporting the effectiveness of RNR-based

interventions (Andrews et al., 2006; Koehler et al., 2012; Vieira et al., 2009) and the lack of improvements in Virginia's recidivism rates since the implementation of YASI (see Table 2), it was hypothesized that discrepancies would exist between the higher risk domains and the assigned priority areas in the case plan (Hypothesis #3).

Setting

In Virginia, DJJ oversees various stages of the juvenile justice system, including intake, probation, commitment, and parole (DJJ, 2016; DJJ, 2017). The implementation of YASI in Virginia began in 2008 in pilot localities (DJJ, 2008). Prior to YASI implementation, DJJ utilized a 12-item home-grown risk assessment tool (DJJ, 2008), which was completely phased out by July 2010 (DJJ, 2011). YASI is now used as the risk assessment and case planning tool throughout all stages of the juvenile justice system in Virginia (DJJ, 2016).

Initial YASI training was conducted by Orbis, the owner and developer of YASI (DJJ, 2017). Although limited training has been ongoing since 2008, there was no systematic method for training new employees or providing continuing education for trained staff. In some cases, staff were expected to complete YASIs after peer-to-peer instruction without completing formal training. A renewed effort to provide training to all staff conducting and supervising YASIs, as well as to certify DJJ staff as YASI instructors, was conducted in 2016 and 2017 with plans for ongoing training (DJJ, 2017). Furthermore, quality assurance efforts have been introduced to monitor and improve fidelity to the instrument (R. Hurt, personal communication, October 30, 2017).

YASI Details

Virginia's YASI contains ten domains (*Legal History*, *Family*, *School*, *Community/Peers*, *Alcohol/Drugs*, *Mental Health*, *Aggression*, *Attitudes*, *Skills*, *Employment/Free Time*) with 87

items and additional sub-items (see Appendix A for an outline of the domains' items). It weights item responses to produce various numerical summed scores and uses cut-off points specific to males and females to produce levels of risk (see Appendix B). The YASI pre-screen includes select items from each domain, with the *Legal History* items constituting almost half. The pre-screen results in the overall risk level of *Low*, *Moderate*, or *High* to represent the likelihood of reoffending. Rather than readdressing the prediction of reoffending, the full assessment provides the more detailed dynamic and static risk – overall and for each domain – thereby focusing on case planning based on the need and responsivity principles (Orbis, 2007). The full assessment overall dynamic risk levels include *Low*, *Low-Moderate*, *Moderate*, *Moderate-High*, *High*, and *Very High* while the domain dynamic risk levels include *None, Low*, *Moderate*, and *High*. (Note: The *Mental Health* domain results in a flag to indicate a need for further assessment and/or intervention without indicating a risk score or level.)

Staff collect information for the assessment from a variety of sources, including the youth, family, and educational service provider records, with open-ended interviews as the primary technique (Orbis, 2007). As mentioned previously, juveniles on probation or parole supervision are scheduled to be reassessed every 180 days (DJJ, 2016) to determine the level and intensity of supervision and develop a case plan, including up to three domain priority areas (DJJ, 2013).

Sample

A total of 11,888 youth placed on probation or parole with DJJ from FY 2014 to FY 2016 (i.e., July 1, 2013, through June 30, 2016) with a completed YASI full assessment close to the placement date (i.e., within the timeframe of 90 days prior to the placement date to 180 days after the placement date) were included in the study. This sample balanced the completeness of

YASI data in Virginia and the follow-up time required to track recidivism. (Virginia's gradual implementation of YASI began in 2008, but assessments were required for probation, direct care, and parole populations beginning in January 2013. DJJ uses FYs to segment their data, making FY 2014 the first complete year after full implementation.)

According to DJJ (2016), 13,650 youth were placed on probation or parole between FY 2014 and FY 2016, meaning 1,762 (12.9%) of the probation and parole population did not have a completed YASI close to the placement date. DJJ did not provide data on these youth with missing assessments. The vast majority (94.6%) of the sample were placed on probation, with 5.4% placed on parole. The majority (76.9%) were male; 47.0% were White, 46.5% were Black, 1.1% were Asian, and 5.5% were other or unknown races. Ages ranged from 7.8 to 21.0 years (M = 16.13, SD = 1.54).

Design

The current study was approved by both Virginia Commonwealth University's Institutional Review Board and DJJ's Human Research Review Committee and Director. De-identified, case-specific administrative data was obtained from DJJ, including demographics (age, sex, race), supervision type (probation or parole), overall and domain YASI risk scores and levels, case plan priority areas, and rearrest status within one year. The YASI assessment selected was the closest assessment up to 90 days prior to or 180 days after the placement date. Rearrest data were selected, as opposed to reconviction data, because DJJ institutes an extra one-year time lag in the reporting of reconviction data due to cases still pending (DJJ, 2016). By using rearrest data, an additional year could be utilized in the sample compared to using reconviction data. The use of rearrest or reconviction as the outcome measure varies among relevant studies, with both options appearing throughout the literature.

Accuracy: Predictive Validity

The predictive validity of risk assessments was initially reported in the literature primarily as the effect sizes measured by Cohen's *d* and the Pearson's correlation coefficient, *r* (Rice & Harris, 2005). Pearson's r^2, measuring the amount of variance accounted for, was also used to provide an index for predictive accuracy (Rice & Harris, 2005). Unfortunately, Cohen's *d* was meant for continuous, normally distributed scores (Rice & Harris, 2005), which the majority of risk assessment tools do not provide. Correlation coefficients as a measure of effect size may fluctuate if the base reoffense rate is not 50% (Rice & Harris, 2005). Thus, these measures can disguise the importance of the predictive accuracy findings for risk assessment tools (Rice & Harris, 2005).

Contemporary studies of predictive validity of risk assessment tools more often rely on the AUC, which "equals the probability that a score (on an ordinal or continuous measure such as a risk assessment instrument) drawn at random from one sample or population (e.g., recidivists' scores) is higher than that drawn at random from a second sample or population (e.g., nonrecidivists' scores)" (Rice & Harris, 2005, p. 618). AUCs represent the sensitivity and specificity of the instrument. Sensitivity values indicate the true positives: the percentage of positive outcomes (i.e., reoffended) correctly classified (i.e., reoffense was predicted). False positives, the percentage of negative outcomes (i.e., did not reoffend) incorrectly classified (i.e., reoffense was predicted), are indicated by 1 - Specificity values. It is a stable measure regardless of the base rate of the given population; therefore, it is more easily compared across populations and studies.

Although van der Put and colleagues (2011) suggested AUC values greater than 0.70 are acceptable, there is no clear consensus. Rice and Harris (2005) recommended the following categorizations:

- Less than 0.55: Negligible or Weak
- 0.56 - 0.63: Small
- 0.64 - 0.71: Moderate
- 0.72 or Greater: Strong

Conversely, Baird et al. (2013), suggested a more strenuous rubric:

- Less than 0.60: Fail
- 0.60 - 0.69: Poor
- 0.70 - 0.79: Fair
- 0.80 - 0.89: Good
- Greater than 0.90: Excellent

Due to the lack of agreement on acceptable levels, some have argued that the common statistical method for determining predictive validity, the AUC, has little practical usage (Baird et al., 2013). Furthermore, a high AUC is possible even with minimal differentiation between groups (Baird et al., 2013). For instance, if the base rate of reoffending is very low, a substantial proportion of the population would be expected to fall in the low risk category of the assessment, and vice versa (Baird et al., 2013), providing limited value to a juvenile justice organization in making case decisions. If a variable had a small range of values within the population being assessed, it would have diminished discriminative power to predict reoffending (Cottle et al., 2001). Instead, some argue that agencies require a tool that will help differentiate levels of risk across their specific populations in order to assign appropriate services and interventions (Baird

et al., 2013). Thus, the base rate of reoffending for the population targeted by an organization must be considered when measuring and interpreting predictive validity findings.

Baird and colleagues (2013) recommend the Dispersion Index for Risk (DIFR) to measure predictive validity, which measures the separation and proportionality between classification groups within a sample. While the AUC determines the sensitivity and specificity of the instrument, DIFR determines the discrimination of the instrument by comparing the recidivism rate of each classification group to the base rate of the total sample while weighting the groups by their size.

Therefore, based on the literature concerning the most appropriate methods of reporting predictive validity of risk assessment tools, both the AUC and the DIFR were calculated in the current study (more details described below). The overall risk scores (i.e., numerical values) and levels (i.e., *Low*, *Moderate*, and *High*) were used as the independent variables. The levels were important to examine in addition to the more precise numerical scores in order to reflect the real-life use of the instrument by practitioners and to account for the differential algorithms built into the instrument for males and females. Rearrest within one year was used as the dependent variable. Dynamic risk scores and levels and domain risk scores and levels were also used as independent variables to complete parallel analyses. From this point forward, references to domain scores and levels reflect dynamic risk, with the exception of the *Legal History* domain which only results in a static risk score and level.

The AUC was calculated for the scores and levels. Though a universal standard for measuring the strength of AUC values does not exist (Baird et al., 2013), the guidelines suggested by Rice and Harris (2005) were used for evaluating the AUC values:

- Less than 0.55: Negligible or Weak

- 0.56 - 0.63: Small
- 0.64 - 0.71: Moderate
- 0.72 or Greater: Strong

Additionally, the DIFR was calculated for the overall risk level to provide a direct comparison to the previous study of YASI in Virginia (Baird et al., 2013) and to measure the separation and proportionality between classification groups. Because Virginia uses YASI at all stages of the system, it was important that the distinction between classification groups is substantial enough to be meaningful and usable. Although this calculation is not widely used in risk assessment validation studies, it provides another perspective on the predictive accuracy of the instrument as a classification tool. According to Baird and colleagues (2013), the formula for DIFR is as follows:

$$DIFR = \sqrt{\sum_{i=1}^{k} (1n\left(\frac{P}{1-P}\right) - 1n\left(\frac{P_i}{1-P_i}\right))^2 * \frac{n_i}{N}}$$

where *k* is the number of subgroups in the risk classification model, *P* is the total sample base rate of the outcome, *N* is the total sample size, p_i represents the base rate of each of the *k* subgroups, and n_i is the size of each *k* subgroup (p. 19-20).

However, according to replications of the calculations in the published article, the two occurrences of "1n" in the formula were typographical errors that should be natural logs (ln).

Equity: Group Differences

DJJ only records sex assigned at birth (not gender identity), without the option for blank or missing values. In addition to Black and White, DJJ also records Asian and Other/Unknown racial categories, but these groups represent small percentages of the populations (0.3% and 4.4%, respectively, of direct care admissions in FY 2016; DJJ, 2016) and were excluded from the

racial analyses. DJJ does not record ethnicity consistently (47.0% of direct care admissions in FY 2016 were missing ethnicity information; DJJ, 2016), so ethnicity was not included in the analyses. To investigate group differences in predictive validity, the AUC and DIFR procedures described above were repeated for each of the four subgroups (i.e., females, males, Black youth, and White youth).

Usage: Priority Areas

For each domain, 1) the number and percentage of youth with the domain listed as a priority area and 2) the breakdown by domain dynamic risk level of youth with and without each priority area were calculated. For each domain, a one-tailed bivariate Spearman correlation between the dynamic risk level (*None*, *Low*, *Moderate*, or *High*) and the priority area status (*yes* or *no*) was completed.

Chapter 4: Results

The results that follow were guided by the study research questions and hypotheses, as summarized in Table 8.

Table 8. Research Questions and Hypotheses

Research Questions	Hypotheses
Accuracy: What is the ability of YASI risk levels to predict reoffending in DJJ's populations?	1. The overall and dynamic risk levels would result in AUCs in the mid-0.60s.
Equity: Do sex or race group differences exist in predictive validity of overall and dynamic risk levels and domain dynamic risk levels?	2a. YASI would better predict reoffending among males than females. 2b. YASI domains of *Family* and *Community/Peers* would be stronger predictors of recidivism for females than males. 2c. YASI would perform comparably for White and Black youth.
Usage: Are assigned case planning priority areas congruent with higher dynamic risk level domains?	3. Discrepancies would exist between the higher risk domains and the assigned priority areas in the case plan.

Descriptive Statistics

Recidivism

Overall, 34.4% of the sample were rearrested within 12 months. Independent samples t-tests were completed to determine group differences in rearrest rates between females and males and between Black and White youth. There was a statistically significant difference in 12-month rearrest rates between females (M = .26, SD = 0.44) and males (M =.37, SD = 0.48); X^2 (1, N = 11,888) = 113.95, p < .001, with a higher percentage of males rearrested than females. There was also a statistically significant difference in 12-month rearrest rates between Black youth (M =.41, SD = 0.49) and White youth (M =.28, SD = 0.45); X^2 (1, N = 11,110) = 198.94, p < .001, with a higher percentage of Black youth rearrested than White youth. Descriptive statistics by sex and race are displayed in Table 9.

Table 9. Descriptive Statistics by Sex and Race

	n	Percentage of Sample	12-Month Rearrest Rate
Sex			
Female	2,749	23.1%	25.9%
Male	9,139	76.9%	36.9%
Race			
Black	5,525	46.5%	41.1%
White	5,585	47.0%	28.4%
Other	650	5.5%	31.2%
Unknown/Missing	128	1.1%	19.9%
Total	11,888	100.0%	34.4%

YASI Assessments

The YASI assessment results in a numeric composite score for the overall and dynamic risk as well as the nine domains. Risk scores for the full sample are displayed in Table 10. Cronbach's alpha, a coefficient of an instrument's reliability, for the nine domain risk scores was .82.

Table 10. Risk Scores

	Minimum	Maximum	*M*	SD
Overall	0	91	28.6	15.9
Dynamic	0	192	70.5	40.6
Domains				
Legal History	0	39	7.7	6.6
Family	0	44	8.6	7.8
School	0	30	8.8	7.4
Community/Peers	0	28	9.3	7.4
Alcohol/Drugs	0	20	6.1	6.4
Aggression	0	17	7.4	5.6
Attitudes	0	35	12.4	10.0
Skills	0	35	15.9	10.7
Employ./Free Time	0	7	1.9	1.6

The domains of *Skills* and *Attitudes* had the highest average scores of the domains. Females had statistically significantly higher scores in the *Family*, *School*, *Aggression*, and *Employment/Free Time* domains whereas males scored higher in the *Legal History*, *Community/Peers*, and *Alcohol/Drugs* domains. There were no statistically significant differences between sexes in the overall or dynamic scores or the *Attitudes* and *Skills* domains. With the exception of the *Family* and *Alcohol/Drugs* domains, Black youth scored statistically significantly higher in the overall, dynamic, and domain scores. White youth scored higher in the *Family* and *Alcohol/Drugs* domains. Risk score values for sex and race subgroups are displayed in Table 11.

Table 11. Risk Scores by Subgroup

Risk Score	Female *M* (SD)	Male *M* (SD)	Black *M* (SD)	White *M* (SD)
Overall	28.95 (16.27)	28.44 (15.72)	29.99 (15.70)*	27.35 (15.84)*
Dynamic	71.38 (41.17)	70.27 (40.49)	73.89 (40.04)*	67.37 (40.91)*
Domains				
Legal History	6.63 (6.23)*	8.08 (6.64)*	9.08 (6.92)*	6.58 (6.02)*
Family	10.03 (8.43)*	8.23 (7.60)*	8.37 (7.56)*	8.87 (8.04)*
School	9.16 (7.43)*	8.69 (7.39)*	9.21 (7.47)*	8.39 (7.30)*
Community/Peers	8.67 (7.30)*	9.51 (7.41)*	10.19 (7.51)*	8.46 (7.15)*
Alcohol/Drugs	5.41 (6.24)*	6.29 (6.43)*	5.72 (6.17)*	6.43 (6.60)*
Aggression	8.00 (5.64)*	7.18 (5.60)*	8.00 (5.56)*	6.83 (5.61)*
Attitudes	12.26 (10.08)	12.42 (9.98)	13.40 (9.98)*	11.46 (9.93)*
Skills	15.67 (10.84)	15.97 (10.64)	16.92 (10.42)*	14.96 (10.86)*
Employ./Free Time	2.17 (1.61)*	1.88 (1.55)*	1.99 (1.61)*	1.90 (1.51)*

Note: Significant mean differences between sex or racial groups ($p < .05$), determined by independent samples t-tests, are indicated by an asterisk with the larger mean shaded.

The YASI tool then converts these numeric scores to risk levels, using different cut-off points for females and males. Of the full sample's overall risk levels, 28.1% were low, 50.9% were moderate, and 21.0% were high. Of the full sample's dynamic risk levels, 30.3% were low,

19.3% were low-moderate, 24.2% were moderate, 16.2% were moderate-high, 6.3% were high, and 3.8% were very high. Unlike the scores, females had statistically significantly lower levels than males with the exception of no significant difference in the *Employment/Free Time* domain. Similar to the scores, Black youth had statistically significantly higher levels than White youth with the exception of the Alcohol/Drugs domain, which followed the opposite pattern, and the *Family* domain, which had no significant difference. Descriptive statistics for the full sample and subgroups are displayed for the overall and dynamic risk levels in Table 12 and for the domains in Table 13. Cronbach's alpha for the nine domain risk levels was .80.

Table 12. Risk Levels by Subgroup

Risk	Level	Full Sample	Female	Male	Black	White
Overall	Low	28.1%	44.7%	23.2%	23.8%	31.6%
	Moderate	50.9%	47.0%	52.1%	52.6%	49.5%
	High	21.0%	8.4%	24.8%	23.6%	18.9%
Dynamic	Low	30.3%	57.3%	22.1%	26.4%	33.7%
	Low-Mod.	19.3%	17.9%	19.7%	18.8%	19.9%
	Moderate	24.2%	13.9%	27.3%	25.4%	23.0%
	Mod.-High	16.2%	8.7%	18.4%	18.4%	14.3%
	High	6.3%	1.8%	7.6%	6.9%	5.7%
	Very High	3.8%	0.4%	4.8%	4.0%	3.4%

Note: All levels were significantly different between sex and racial groups ($p < .05$), determined by chi-squared tests. Females had statistically significantly lower levels than males. Black youth had statistically significantly higher levels than White youth.

Table 13. Domain Levels by Subgroup

Domain	Level	Full Sample	Female	Male	Black	White
Legal History	Low	22.0%	46.7%	14.5%	15.7%	27.6%
	Moderate	59.4%	45.9%	63.4%	59.7%	58.9%
	High	18.7%	7.4%	22.1%	24.7%	13.5%
Family	None	9.5%	8.0%	9.9%	8.6%	10.3%
	Low	35.1%	69.7%	24.7%	35.5%	34.6%
	Moderate	43.4%	18.8%	50.8%	44.5%	42.6%
	High	12.0%	3.5%	14.6%	11.4%	12.4%
School	None	14.5%	12.8%	15.1%	12.3%	16.6%
	Low	16.5%	39.2%	9.7%	16.1%	16.7%
	Moderate	58.1%	42.2%	62.9%	59.7%	57.1%
	High	10.8%	5.7%	12.3%	11.9%	9.6%
Community/ Peers	None	20.5%	23.5%	19.6%	17.4%	23.3%
	Low	26.4%	35.6%	23.6%	24.4%	28.6%
	Moderate	38.3%	38.9%	38.1%	40.1%	36.6%
	High	14.8%	2.0%	18.7%	18.0%	11.5%
Alcohol/ Drugs	None	40.6%	45.6%	39.1%	41.6%	39.7%
	Low	17.6%	10.0%	19.8%	19.1%	16.3%
	Moderate	23.1%	42.5%	17.3%	22.0%	24.0%
	High	18.7%	1.9%	23.8%	17.2%	20.0%
Aggression	None	22.2%	19.0%	23.2%	18.1%	25.7%
	Low	21.2%	44.1%	14.4%	20.0%	22.3%
	Moderate	43.7%	32.4%	47.1%	47.6%	40.2%
	High	12.9%	4.5%	15.4%	14.3%	11.8%
Attitudes	None	7.3%	8.2%	7.0%	5.5%	8.9%
	Low	32.7%	40.6%	30.4%	29.0%	36.1%
	Moderate	48.1%	39.1%	50.9%	51.9%	44.7%
	High	11.9%	12.1%	11.8%	13.6%	10.2%
Skills	None	6.8%	7.8%	6.5%	5.4%	8.1%
	Low	25.9%	39.2%	21.9%	22.9%	28.5%
	Moderate	45.8%	39.1%	47.8%	47.7%	44.2%
	High	21.5%	13.9%	23.8%	24.0%	19.1%
Employment/ Free Time	None	26.0%	20.5%	27.7%	26.1%	26.1%
	Low	58.2%	66.2%	55.8%	56.5%	60.1%
	Moderate	14.1%	11.2%	15.0%	15.6%	12.6%
	High	1.6%	2.1%	1.5%	1.8%	1.3%

Note: All domain levels were significantly different between sex and racial groups ($p < .05$), determined by chi-squared tests, with the exception of the Family domain for race and the Employment/Free Time domain for sex. Females had statistically significantly lower levels than

males with the exception of no significant difference in the Employment/Free Time domain. Black youth had statistically significantly higher levels than White youth with the exception of the Alcohol/Drugs domain, which followed the opposite pattern, and the Family domain, which had no significant difference.

A summary of the statistically significant differences in scores and levels by sex and race is displayed in Table 14, with the higher subgroup listed.

Table 14. Summary of Statistically Significant Assessment Differences by Subgroup

	Sex		Race	
	Score	Level	Score	Level
Overall	n.s.	Male	Black	Black
Dynamic	n.s.	Male	Black	Black
Domains				
Legal History	Male	Male	Black	Black
Family	Female	Male	White	n.s.
School	Female	Male	Black	Black
Community/Peers	Male	Male	Black	Black
Alcohol/Drugs	Male	Male	White	White
Aggression	Female	Male	Black	Black
Attitudes	n.s.	Male	Black	Black
Skills	n.s.	Male	Black	Black
Employment/Free Time	Female	n.s.	Black	Black

Note: The subgroup with the higher score or level is listed.

Accuracy: Predictive Validity

As risk scores increased, the actual 12-month rearrest rates generally increased for overall and dynamic risk. Likewise, as risk levels increased, the actual 12-month rearrest rates consistently increased for overall and dynamic risk for all subgroups. Similarly, as domain risk levels increased, the actual 12-month rearrest rates generally increased for all subgroups (exceptions: *School*, *Alcohol/Drugs*, and *Employment/Free Time* domains for Black youth and *Employment/Free Time* domain for females). The risk scores by 12-month rearrest rate are displayed for overall and dynamic risk in Figure 1 and Figure 2, respectively. The risk levels by

12-month rearrest rate are displayed in Table 15 for overall and dynamic risk and in Table 16 for domains. The analyses that follow further investigate these relationships.

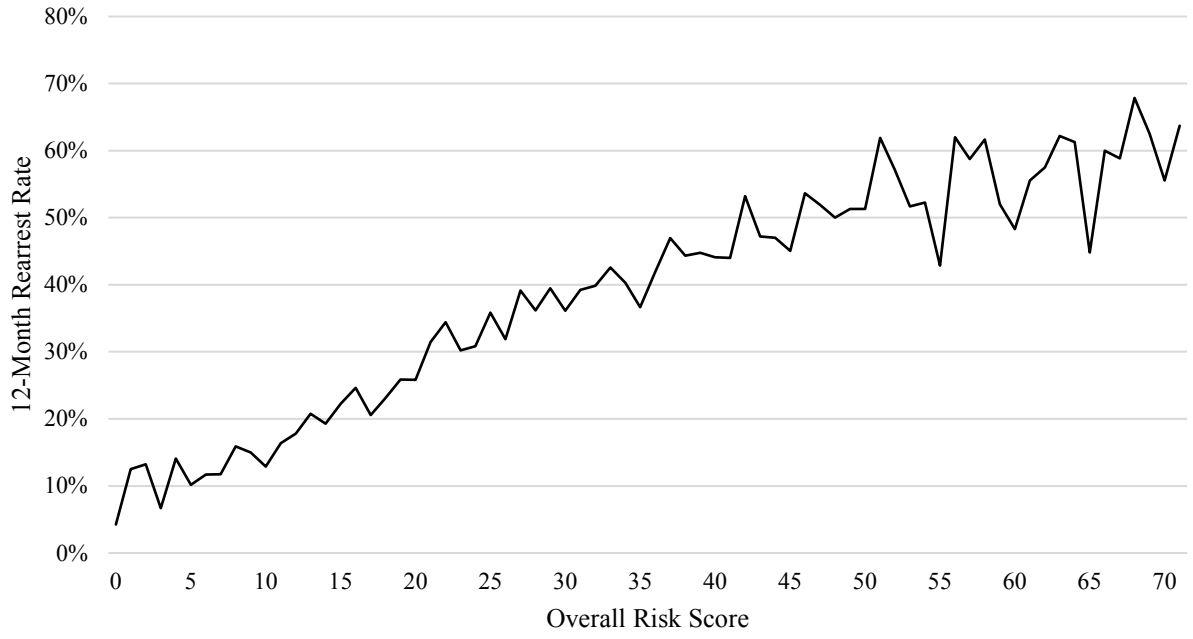


Figure 1. Overall Risk Score by 12-Month Rearrest Rate

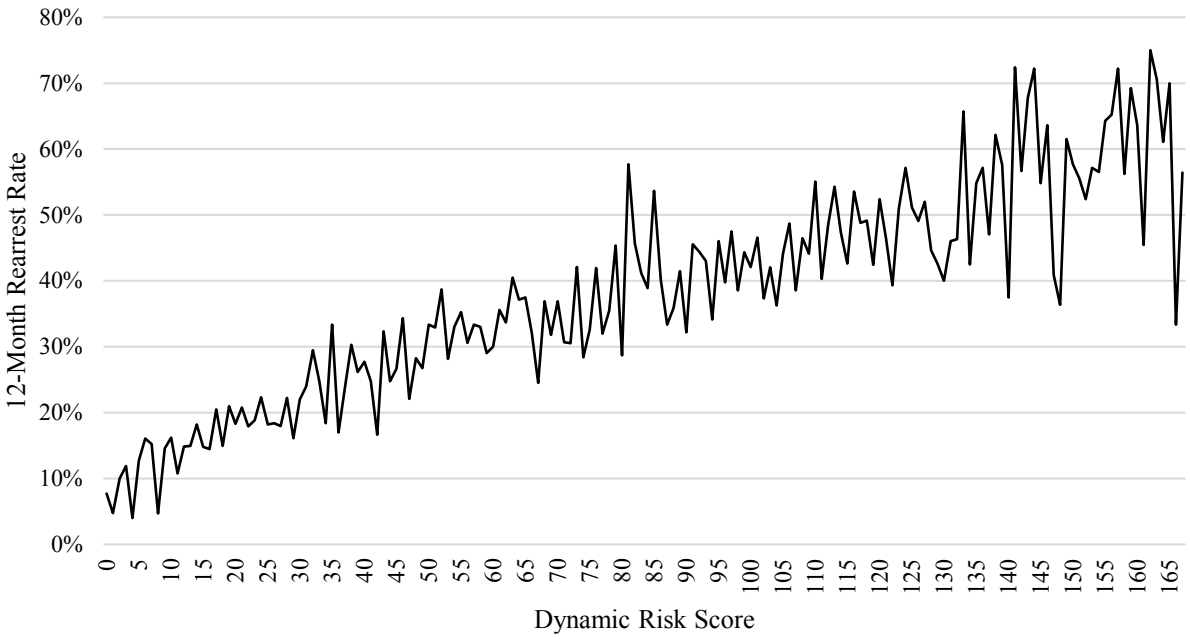


Figure 2. Dynamic Risk Score by 12-Month Rearrest Rate

Table 15. 12-Month Rearrest Rates by Risk Levels

Risk	Level	Full Sample	Female	Male	Black	White
Overall	Low	16.4%	15.8%	16.8%	21.7%	13.0%
	Moderate	35.9%	31.7%	37.0%	41.9%	30.0%
	High	54.9%	47.4%	55.6%	59.2%	50.1%
Dynamic	Low	19.1%	19.1%	19.2%	23.7%	15.9%
	Low-Mod.	29.8%	29.1%	30.0%	36.2%	25.2%
	Moderate	39.1%	36.7%	39.4%	45.7%	32.6%
	Mod.-High	47.8%	40.8%	48.8%	52.9%	41.6%
	High	51.6%	48.0%	51.9%	56.3%	46.4%
	Very High	63.5%	60.0%	63.6%	70.4%	56.6%

Table 16. 12-Month Rearrest Rates by Domain Levels

Domain	Level	Full Sample	Female	Male	Black	White
Legal History	Low	20.8%	18.2%	23.4%	25.1%	19.1%
	Moderate	33.7%	30.6%	34.3%	39.4%	28.4%
	High	52.6%	45.3%	53.3%	55.6%	47.3%
Family	None	21.2%	14.5%	22.8%	27.5%	17.0%
	Low	26.0%	24.9%	26.9%	31.5%	21.2%
	Moderate	39.8%	32.4%	40.6%	47.7%	32.2%
	High	49.9%	37.5%	50.8%	56.0%	44.7%
School	None	25.2%	18.2%	27.0%	34.0%	19.2%
	Low	25.2%	23.0%	27.9%	30.4%	21.5%
	Moderate	36.7%	29.8%	38.1%	43.1%	30.6%
	High	48.4%	34.2%	50.4%	53.5%	43.1%
Community/ Peers	None	20.1%	16.9%	21.3%	26.7%	16.3%
	Low	27.2%	23.3%	28.9%	33.4%	22.8%
	Moderate	39.1%	32.9%	41.0%	44.3%	33.7%
	High	54.8%	41.1%	55.2%	58.5%	50.1%
Alcohol/ Drugs	None	24.1%	20.1%	25.5%	31.4%	17.1%
	Low	37.5%	22.5%	39.8%	45.9%	28.6%
	Moderate	38.6%	31.6%	43.8%	44.5%	34.2%
	High	48.6%	54.7%	48.4%	55.2%	43.7%
Aggression	None	22.5%	15.7%	24.2%	28.6%	18.6%
	Low	26.9%	22.8%	30.7%	31.1%	24.0%
	Moderate	39.3%	34.0%	40.4%	44.7%	33.3%
	High	50.3%	40.8%	51.1%	59.2%	41.5%
Attitudes	None	19.6%	16.0%	20.8%	24.8%	16.9%
	Low	25.5%	20.1%	27.7%	31.5%	21.6%
	Moderate	38.9%	29.6%	41.0%	44.5%	32.9%
	High	49.6%	40.1%	52.6%	55.6%	42.7%
Skills	None	18.4%	14.5%	19.8%	25.2%	14.1%
	Low	24.8%	19.9%	27.5%	28.3%	22.7%
	Moderate	37.1%	30.1%	38.9%	42.9%	31.8%
	High	45.0%	37.3%	46.4%	53.4%	35.0%
Employ./ Free Time	None	26.4%	17.6%	28.4%	33.6%	20.2%
	Low	35.4%	27.9%	38.0%	41.9%	29.9%
	Moderate	43.9%	29.3%	47.1%	50.2%	36.7%
	High	44.3%	26.3%	51.9%	48.5%	44.4%

The AUCs for the overall and dynamic risk scores and levels ranged from 0.65 to 0.68. The confidence intervals ranged from a low of 0.64 to a high of 0.69, with most overlapping. (See Table 17 and Figure 3.) Each of these values fell into Rice and Harris' (2015) moderate range.

Table 17. AUC Values

Risk	AUC	Std. Error	*p* value	95% CI
Overall Score	0.68	0.01	< 0.001	0.67 - 0.69
Overall Level	0.66	0.01	< 0.001	0.65 - 0.67
Dynamic Score	0.65	0.01	< 0.001	0.64 - 0.66
Dynamic Level	0.66	0.01	< 0.001	0.65 - 0.67

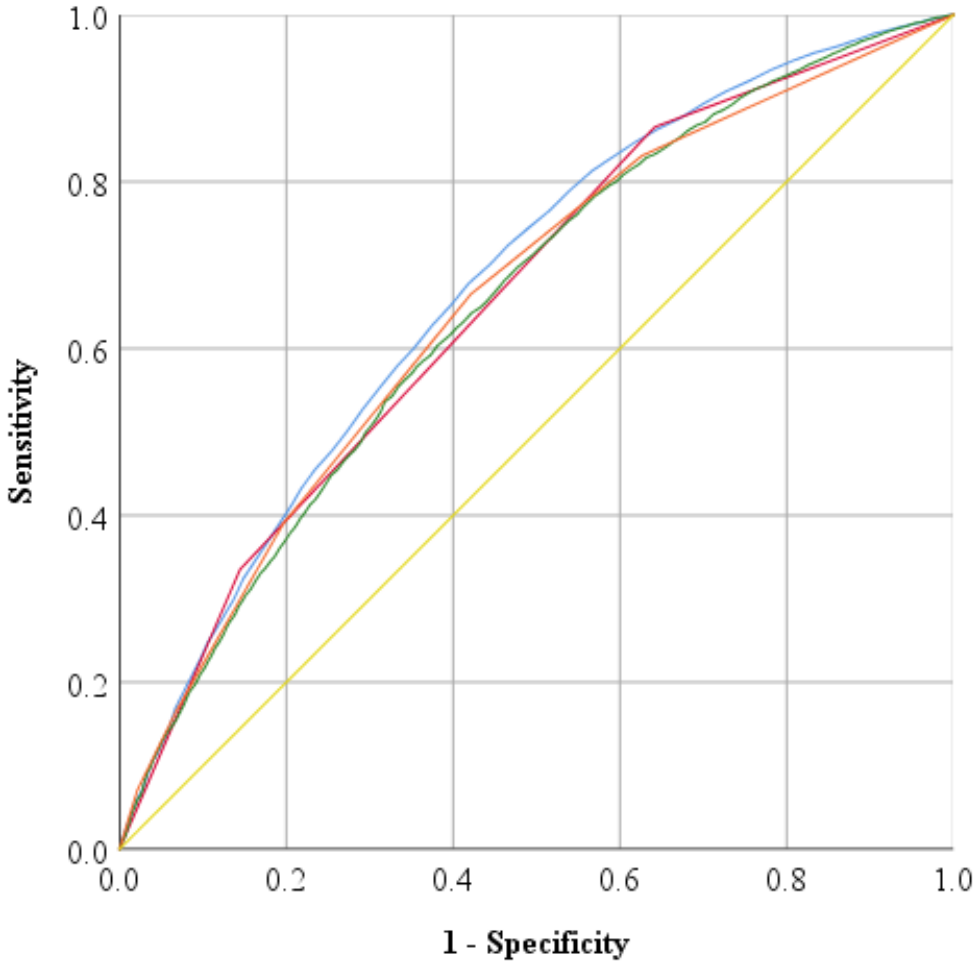


Figure 3. ROC Curves for Overall and Dynamic Risk Scores and Levels

The AUCs for both the domain scores and levels ranged from 0.56 to 0.64. (See Table 18 and Table 19.) The AUC values fell mostly into Rice and Harris' (2015) small range, with *Legal History* (score only) and *Community/Peers* (score and level) designated as moderate. Although many of the confidence intervals overlapped, the *Legal History* and *Community/Peers* domains had the highest predictive ability, and the *School* and *Employment/Free Time* domains had the lowest predictive ability. Based on the confidence intervals, there was no significant difference between the scores and the levels with the exception of the *Legal History* domain, where the score was slightly more predictive than the level.

Table 18. AUC Values for Domain Scores

Domain	AUC	Std. Error	*p* value	95% CI
Legal History	0.64	0.01	< 0.001	0.63 - 0.65
Family	0.60	0.01	< 0.001	0.59 - 0.61
School	0.59	0.01	< 0.001	0.58 - 0.60
Community/Peers	0.64	0.01	< 0.001	0.63 - 0.65
Alcohol/Drugs	0.61	0.01	< 0.001	0.60 - 0.62
Aggression	0.61	0.01	< 0.001	0.60 - 0.62
Attitudes	0.62	0.01	< 0.001	0.61 - 0.63
Skills	0.60	0.01	< 0.001	0.59 - 0.61
Employ./Free Time	0.56	0.01	< 0.001	0.55 - 0.57

Table 19. AUC Values for Domain Levels

Domain	AUC	Std. Error	*p* value	95% CI
Legal History	0.61	0.01	< 0.001	0.60 - 0.62
Family	0.61	0.01	< 0.001	0.60 - 0.62
School	0.58	0.01	< 0.001	0.57 - 0.59
Community/Peers	0.64	0.01	< 0.001	0.62 - 0.65
Alcohol/Drugs	0.61	0.01	< 0.001	0.60 - 0.62
Aggression	0.61	0.01	< 0.001	0.60 - 0.62
Attitudes	0.60	0.01	< 0.001	0.59 - 0.62
Skills	0.60	0.01	< 0.001	0.59 - 0.61
Employ./Free Time	0.56	0.01	< 0.001	0.55 - 0.57

The DIFR was 0.65 for the overall risk level.

Equity: Group Differences

Sex

The AUCs for the overall and dynamic risk scores and levels ranged from 0.62 to 0.67 for females and from 0.65 to 0.68 for males, with the confidence intervals largely overlapping. (See Table 20 and Table 21.) For females, the scores fell into Rice and Harris' (2015) moderate range, and the levels fell into the small range. Both scores and levels fell into the moderate range for males. Although the male AUC values were consistently higher than females', each of the confidence intervals between females and males overlapped, indicating that any differences in AUC values were not significantly different. The DIFR for the overall risk level was higher for males than females (0.64 and 0.53, respectively).

Table 20. AUC Values for Female Overall and Dynamic Risk

Risk	AUC	Std. Error	*p* value	95% CI
Overall Score	0.67	0.01	< 0.001	0.64 - 0.69
Overall Level	0.63	0.01	< 0.001	0.61 - 0.66
Dynamic Score	0.64	0.01	< 0.001	0.62 - 0.67
Dynamic Level	0.62	0.01	< 0.001	0.59 - 0.64

Table 21. AUC Values for Males Overall and Dynamic Risk

Risk	AUC	Std. Error	*p* value	95% CI
Overall Score	0.68	0.01	< 0.001	0.67 - 0.69
Overall Level	0.65	0.01	< 0.001	0.64 - 0.66
Dynamic Score	0.66	0.01	< 0.001	0.65 - 0.67
Dynamic Level	0.65	0.01	< 0.001	0.64 - 0.66

The AUCs for the domain scores ranged from 0.55 to 0.63 for females and from 0.57 to 0.65 for males. The AUC values for the domain scores fell mostly in the small range (exceptions: *Employment/Free Time* was weak for females; *Legal History* and *Community/Peers* were moderate for males). For females, the *Legal History* domain had the highest AUC value, with the

confidence interval exceeding the *School* and *Employment/Free Time* domains. For males, the *Community/Peers* domain had the highest AUC value, with the confidence interval exceeding the *Family*, *School*, *Alcohol/Drugs*, *Aggression*, *Attitudes*, *Skills*, and *Employment/Free Time* domains. The *Legal History* domain also had a confidence interval exceeding the *School*, *Skills*, and *Employment/Free Time* domains for males. The only domain scores with a non-overlapping confidence intervals between females and males was *Community/Peers*, with higher AUC values for males. (See Table 22 and Table 23.)

Table 22. AUC Values for Female Domain Scores

Domain	AUC	Std. Error	*p* value	95% CI
Legal History	0.63	0.01	< 0.001	0.61 - 0.66
Family	0.60	0.01	< 0.001	0.57 - 0.62
School	0.58	0.01	< 0.001	0.56 - 0.61
Community/Peers	0.61	0.01	< 0.001	0.59 - 0.64
Alcohol/Drugs	0.60	0.01	< 0.001	0.57 - 0.62
Aggression	0.62	0.01	< 0.001	0.59 - 0.64
Attitudes	0.61	0.01	< 0.001	0.59 - 0.64
Skills	0.61	0.01	< 0.001	0.59 - 0.64
Employ./Free Time	0.55	0.01	< 0.001	0.53 - 0.57

Table 23. AUC Values for Male Domain Scores

Domain	AUC	Std. Error	*p* value	95% CI
Legal History	0.64	0.01	< 0.001	0.63 - 0.65
Family	0.61	0.01	< 0.001	0.60 - 0.63
School	0.60	0.01	< 0.001	0.58 - 0.61
Community/Peers	0.65	0.01	< 0.001	0.64 - 0.66
Alcohol/Drugs	0.61	0.01	< 0.001	0.60 - 0.63
Aggression	0.61	0.01	< 0.001	0.60 - 0.63
Attitudes	0.62	0.01	< 0.001	0.61 - 0.63
Skills	0.60	0.01	< 0.001	0.59 - 0.61
Employ./Free Time	0.57	0.01	< 0.001	0.56 - 0.58

The AUCs for the domain levels ranged from 0.55 to 0.61 for females and from 0.57 to 0.64 for males. The AUC values for the domain levels fell mostly in the small range (exceptions: *Employment/Free Time* was weak for females; *Community/Peers* was moderate for males). For females, the *Legal History* domain had the highest AUC value, with the confidence interval exceeding the *Employment/Free Time* domain. For males, the *Community/Peers* domain had the highest AUC value, with the confidence interval exceeding all other domains. The *Alcohol/Drugs* domain also had a confidence interval exceeding the *School* and *Employment/Free Time* domains for males. The only domain levels with a non-overlapping confidence intervals between females and males were *Family* and *Community/Peers*, with higher AUC values for males. (See Table 24 and Table 25.)

Table 24. AUC Values for Female Domain Levels

Domain	AUC	Std. Error	*p* value	95% CI
Legal History	0.61	0.01	< 0.001	0.58 - 0.63
Family	0.56	0.01	< 0.001	0.53 - 0.58
School	0.56	0.01	< 0.001	0.54 - 0.59
Community/Peers	0.59	0.01	< 0.001	0.57 - 0.62
Alcohol/Drugs	0.59	0.01	< 0.001	0.56 - 0.61
Aggression	0.60	0.01	< 0.001	0.58 - 0.63
Attitudes	0.60	0.01	< 0.001	0.57 - 0.62
Skills	0.60	0.01	< 0.001	0.57 - 0.62
Employ./Free Time	0.55	0.01	< 0.001	0.52 - 0.57

Table 25. AUC Values for Male Domain Levels

Domain	AUC	Std. Error	*p* value	95% CI
Legal History	0.60	0.01	< 0.001	0.59 - 0.61
Family	0.60	0.01	< 0.001	0.59 - 0.61
School	0.57	0.01	< 0.001	0.56 - 0.58
Community/Peers	0.64	0.01	< 0.001	0.62 - 0.65
Alcohol/Drugs	0.61	0.01	< 0.001	0.60 - 0.62
Aggression	0.60	0.01	< 0.001	0.59 - 0.62
Attitudes	0.60	0.01	< 0.001	0.59 - 0.61
Skills	0.59	0.01	< 0.001	0.58 - 0.60
Employ./Free Time	0.57	0.01	< 0.001	0.56 - 0.58

Race

The AUCs for the overall and dynamic risk scores and levels ranged from 0.64 to 0.65 for Black youth and from 0.65 to 0.69 for White youth, with the confidence intervals largely overlapping. (See Table 26 and Table 27.) All scores and levels fell into the moderate range according to Rice and Harris (2015). The AUC values for White youth were consistently higher than those for Black youth, but only the confidence intervals for the overall risk scores between Black and White youth did not overlap. The DIFR for the overall risk level was higher for White youth than Black youth (0.68 and 0.58, respectively).

Table 26. AUC Values for Black Overall and Dynamic Risk

Risk	AUC	Std. Error	*p* value	95% CI
Overall Score	0.65	0.01	< 0.001	0.64 - 0.67
Overall Level	0.64	0.01	< 0.001	0.63 - 0.65
Dynamic Score	0.64	0.01	< 0.001	0.63 - 0.66
Dynamic Level	0.65	0.01	< 0.001	0.63 - 0.66

Table 27. AUC Values for White Overall and Dynamic Risk

Risk	AUC	Std. Error	*p* value	95% CI
Overall Score	0.69	0.01	< 0.001	0.67 - 0.70
Overall Level	0.67	0.01	< 0.001	0.65 - 0.68
Dynamic Score	0.66	0.01	< 0.001	0.64 - 0.67
Dynamic Level	0.65	0.01	< 0.001	0.64 - 0.67

The AUCs for the domain scores ranged from 0.56 to 0.63 for Black youth and from 0.57 to 0.65 for White youth. The AUC values for the domain scores fell mostly in the small range (exceptions: *Community/Peers* and *Alcohol/Drugs* were moderate for White youth). For Black youth, the *Community/Peers* and *Legal History* domains had the highest AUC values, with confidence intervals exceeding the *School* and *Employment/Free Time* domains. For White youth, the *Community/Peers* and *Alcohol/Drugs* domains had the highest AUC values, with the confidence intervals exceeding the *Family*, *School*, *Aggression*, *Attitudes*, *Skills*, and *Employment/Free Time* domains. The only domain score with non-overlapping confidence intervals between Black youth and White youth was *Alcohol/Drugs*, with higher AUC values for White youth. (See Table 28 and Table 29.)

Table 28. AUC Values for Black Domain Scores

Domain	AUC	Std. Error	*p* value	95% CI
Legal History	0.62	0.01	< 0.001	0.61 - 0.64
Family	0.60	0.01	< 0.001	0.58 - 0.61
School	0.57	0.01	< 0.001	0.56 - 0.59
Community/Peers	0.63	0.01	< 0.001	0.61 - 0.64
Alcohol/Drugs	0.60	0.01	< 0.001	0.58 - 0.61
Aggression	0.60	0.01	< 0.001	0.59 - 0.62
Attitudes	0.61	0.01	< 0.001	0.60 - 0.63
Skills	0.61	0.01	< 0.001	0.59 - 0.62
Employ./Free Time	0.56	0.01	< 0.001	0.54 - 0.57

Table 29. AUC Values for White Domain Scores

Domain	AUC	Std. Error	*p* value	95% CI
Legal History	0.63	0.01	< 0.001	0.61 - 0.65
Family	0.61	0.01	< 0.001	0.60 - 0.63
School	0.60	0.01	< 0.001	0.58 - 0.62
Community/Peers	0.65	0.01	< 0.001	0.63 - 0.66
Alcohol/Drugs	0.64	0.01	< 0.001	0.63 - 0.66
Aggression	0.60	0.01	< 0.001	0.59 - 0.62
Attitudes	0.61	0.01	< 0.001	0.59 - 0.63
Skills	0.59	0.01	< 0.001	0.57 - 0.61
Employ./Free Time	0.57	0.01	< 0.001	0.55 - 0.58

The AUCs for the domain levels ranged from 0.56 to 0.62 for Black youth and from 0.57 to 0.64 for White youth. The AUC values for the domain levels fell mostly in the small range (exceptions: *Community/Peers* and *Alcohol/Drugs* were moderate for White youth). For Black youth, the *Community/Peers* and *Aggression* domains had the highest AUC values, with confidence intervals exceeding the *School* and *Employment/Free Time* domains. For White youth, the *Community/Peers* and *Alcohol/Drugs* domains had the highest AUC values, with the confidence intervals exceeding the *Legal History*, *School*, *Aggression*, *Attitudes*, *Skills*, and *Employment/Free Time* domains. The only domain levels with a non-overlapping confidence intervals between Black youth and White youth was *Alcohol/Drugs*, with higher AUC values for White youth. (See Table 30 and Table 31.)

Table 30. AUC Values for Black Domain Levels

Domain	AUC	Std. Error	*p* value	95% CI
Legal History	0.60	0.01	< 0.001	0.59 - 0.62
Family	0.61	0.01	< 0.001	0.59 - 0.62
School	0.57	0.01	< 0.001	0.55 - 0.58
Community/Peers	0.62	0.01	< 0.001	0.60 - 0.63
Alcohol/Drugs	0.60	0.01	< 0.001	0.58 - 0.61
Aggression	0.61	0.01	< 0.001	0.60 - 0.63
Attitudes	0.59	0.01	< 0.001	0.58 - 0.61
Skills	0.60	0.01	< 0.001	0.59 - 0.62
Employ./Free Time	0.56	0.01	< 0.001	0.54 - 0.57

Table 31. AUC Values for White Domain Levels

Domain	AUC	Std. Error	*p* value	95% CI
Legal History	0.60	0.01	< 0.001	0.58 - 0.62
Family	0.61	0.01	< 0.001	0.59 - 0.63
School	0.58	0.01	< 0.001	0.57 - 0.60
Community/Peers	0.64	0.01	< 0.001	0.62 - 0.65
Alcohol/Drugs	0.64	0.01	< 0.001	0.62 - 0.65
Aggression	0.60	0.01	< 0.001	0.59 - 0.62
Attitudes	0.60	0.01	< 0.001	0.58 - 0.62
Skills	0.58	0.01	< 0.001	0.56 - 0.60
Employ./Free Time	0.57	0.01	< 0.001	0.55 - 0.58

A summary of the statistically significant differences in AUC values by sex and race is displayed in Table 32, with the higher subgroup listed.

Table 32. Summary of Statistically Significant AUC Differences by Subgroup

	Sex		Race	
	Score	Level	Score	Level
Overall	n.s.	n.s.	White	n.s.
Dynamic	n.s.	n.s.	n.s.	n.s.
Domains				
Legal History	n.s.	n.s.	n.s.	n.s.
Family	n.s.	Male	n.s.	n.s.
School	n.s.	n.s.	n.s.	n.s.
Community/Peers	Male	Male	n.s.	n.s.
Alcohol/Drugs	n.s.	n.s.	White	White
Aggression	n.s.	n.s.	n.s.	n.s.
Attitudes	n.s.	n.s.	n.s.	n.s.
Skills	n.s.	n.s.	n.s.	n.s.
Employment/Free Time	n.s.	n.s.	n.s.	n.s.

Note: The subgroup with the higher AUC value is listed.

Usage: Priority Areas

The majority (62.2%) of cases had three priority areas identified in the case plan, and almost all (96.0%) had at least one priority area identified in the case plan. (See Table 33.)

Table 33. Number of Assigned Priority Areas

Priority Areas	*n*	Percentage of Sample
None	477	4.0%
1	1,708	14.4%
2	2,306	19.4%
3	7,397	62.2%
Total	11,888	100.0%

The domains with the highest percentage of cases with moderate to high risk levels were *School* (68.9%), *Skills* (67.3%), and *Attitudes* (60.0%). The domains with the highest percentage of cases with assigned priority areas were *Skills* (45.3%), *Aggression* (37.7%), and *Attitudes* (36.0%). The *Employment/Free Time* domain had both the lowest percentage of cases with

moderate to high domain risk (15.7%) and the lowest percentage of cases with the domain assigned as a priority area (6.0%). The domains varied in the percentage of youth with moderate to high risk levels with the domain assigned as a priority area; 63.9% of youth assessed as moderate to high risk in *Alcohol/Drugs*, 55.0% assessed as moderate to high risk in *Skills*, and 54.1% assessed as moderate to high risk in *Aggression* were assigned the respective priority area. Conversely, 12.6% of youth assessed as moderate to high in *Employment/Free Time*, 22.3% assessed as moderate to high risk in *Family*, and 38.1% assessed as moderate to high risk in *School* were assigned the respective priority area. (See Table 34.)

Table 34. Domain and Priority Area Prevalence

Domain	% of Sample with Mod.-High Domain Risk	% of Sample with Domain Assigned Priority	% of Mod.-High Domain Risk with Assigned Priority	% of Assigned Priority with Mod-High Domain Risk
Family	55.4%	15.8%	22.3%	78.0%
School	68.9%	30.2%	38.1%	87.0%
Community/Peers	53.1%	28.7%	43.0%	79.6%
Alcohol/Drugs	41.8%	33.1%	63.9%	80.6%
Aggression	56.6%	37.7%	54.1%	81.3%
Attitudes	60.0%	36.0%	48.2%	80.3%
Skills	67.3%	45.3%	55.0%	81.6%
Employ./Free Time	15.7%	6.0%	12.6%	33.2%

With the exception of the *Employment/Free Time* domain, the priority areas were assigned to domains scoring moderate to high 78.0% (*Family*) to 87.0% (*School*) of the time. However, 66.8% of the cases with *Employment/Free Time* assigned as a priority area were scored as none or low risk in that domain. For all domains except *Employment/Free Time*, the percentage with the assigned priority area but no risk level was less than 7% (12.4% of cases assigned an *Employment/Free Time* priority area had no risk level for the domain). For all

domains, the percentage without the assigned priority area but high risk level was 10% or less. (See Table 35.)

Table 35. Dynamic Risk Levels by Assigned Priority Areas

		Dynamic Risk Level				
Domain	Priority	None	Low	Moderate	High	Total
Family	Yes	2.0%	20.0%	52.2%	25.8%	1,883
	No	10.9%	38.0%	41.7%	9.4%	10,005
School	Yes	3.5%	9.5%	70.3%	16.7%	3,594
	No	19.3%	19.6%	52.8%	8.3%	8,294
Community/Peers	Yes	4.7%	15.7%	52.7%	26.9%	3,409
	No	26.9%	30.7%	32.5%	10.0%	8,479
Alcohol/Drugs	Yes	5.9%	13.5%	37.3%	43.3%	3,937
	No	57.8%	19.6%	16.0%	6.5%	7,951
Aggression	Yes	6.2%	12.5%	56.5%	24.9%	4,476
	No	31.9%	26.5%	35.9%	5.7%	7,412
Attitudes	Yes	2.0%	17.7%	59.7%	20.6%	4,280
	No	10.2%	41.2%	41.6%	7.0%	7,608
Skills	Yes	1.9%	16.5%	50.5%	31.1%	5,391
	No	10.8%	33.7%	42.0%	13.5%	6,497
Employment/Free Time	Yes	12.4%	54.4%	25.5%	7.7%	710
	No	26.9%	58.5%	13.4%	1.2%	11,178

Note: All domain risk levels were statistically significantly different by corresponding priority area ($p < .05$).The *Mental Health* domain is not displayed above because it does not have an associated risk level; however, 7.0% of cases were assigned *Mental Health* as a priority area.

All domains had a statistically significant positive one-tailed Spearman correlation between the domain risk level (*None*, *Low*, *Moderate*, or *High*) and priority area assignment (*yes* or *no*). The *Alcohol/Drugs* domain had the strongest correlation ($r = 0.59$), followed by

Aggression ($r = 0.43$). The *Employment/Free Time* domain had the weakest correlation ($r = 0.15$). (See Table 36.)

Table 36. Spearman Correlations between Domain Level and Priority Area Assignment

Domain	*r*	*p*
Family	0.23	< 0.001
School	0.27	< 0.001
Community/Peers	0.36	< 0.001
Alcohol/Drugs	0.59	< 0.001
Aggression	0.42	< 0.001
Attitudes	0.33	< 0.001
Skills	0.31	< 0.001
Employ./Free Time	0.12	< 0.001

Chapter 5: Discussion

Accuracy: Predictive Validity

Regarding the research question of the ability of YASI risk levels to predict reoffending in DJJ's population, the hypothesis that the overall and dynamic risk levels would result in AUCs in the mid-0.60s was supported (Hypothesis #1). The incremental relationship between risk and rearrest suggested that YASI had predictive validity, and each of the AUC values for overall and dynamic risk scores and levels (0.65 - 0.68) fell into Rice and Harris's (2015) moderate range. The shorter pre-screen performed as well or better than the full assessment's dynamic risk, indicating the additional items in the full assessment do not improve predictive accuracy.

For individual domains, the AUC values (0.56 – 0.64) fell mostly into Rice and Harris's (2015) small range. The *Legal History* and *Community/Peers* domains had the highest predictive ability, and the *School* and *Employment/Free Time* domains had the lowest predictive ability.

These AUC values were comparable to most previous studies of YASI for different populations (see Table 37). Likewise, the DIFR of 0.65 for the overall risk level was similar to the previously reported DIFR of 0.68 in Virginia (Baird et al., 2013). While the values did not indicate a strong predictive ability, YASI in Virginia performed at the expected and adequate level with respect to existing research on youth risk assessment instruments.

Table 37. YASI AUC Results of Current Study Compared to Previous Studies

	Orbis, 2007	Jones, 2011	Jones et al., 2016	Baird et al., 2013	Current - Scores	Current - Levels
Pre-Screen	0.65	0.64	0.79	0.68	0.68	0.66
Female	0.61	0.60	0.68	0.67	0.67	0.63
Male	0.68	0.64	0.82	0.71	0.68	0.65
Black				0.66	0.65	0.64
White				0.68	0.69	0.67
Dynamic Risk	0.62	0.63			0.65	0.66
Female	0.59	0.62			0.64	0.62
Male	0.64	0.63			0.66	0.65
Domain Dynamic Risk	0.55-0.63		0.54-0.73		0.56-0.64	0.56-0.64
Female		0.50-0.60			0.55-0.63	0.55-0.61
Male		0.50-0.62			0.57-0.65	0.57-0.64

Note: Studies varied in population, follow-up timeframes, and recidivism measures and may have reported additional AUCs not displayed in the summary table above. Orbis (2007) and Jones (2011) both studied New York populations, resulting in similar AUCs. In the Orbis (2007) study, "negative outcome" was defined as a new referral/arrest, violation of probation, or adjudication/conviction; AUCs for the pre-screen after item weight and cut-off point revisions are displayed.

These findings represent the typical results of juvenile risk assessment validation studies in which few tools achieve a strong level of predictive accuracy (Schwalbe, 2007). Despite these consistently unexceptional findings, risk assessment instruments perform better than relying solely on professional judgment (Andrews et al., 2006; Bonta & Andrews, 2007), and their utilization remains a recommended best practice (Latessa & Lovins, 2010). YASI in Virginia continues this trend by demonstrating similarly adequate levels of predictive accuracy given the limitations of the field.

The choice to purchase a commercial product validated in other jurisdictions did not appear to greatly handicap the tool in Virginia, though it is unknown how a locally created instrument would have performed. This conclusion is counter to some previous research suggesting that off-the-shelf tools may not be generalizable to other jurisdictions (Jones et al.,

2001; Miller & Lin, 2007; Schwalbe, 2007). Instead, Skeem and Monahan's (2011) view may be more realistic in that most validated risk instruments have so many common elements that there are few significant distinctions in predictive validity, regardless of location. The congruence of the factors measured by risk tools to the Central Eight concepts (Andrews et al., 2006; Andrews & Bonta, 2010) and the commonalities found among predictive studies (Cottle et al., 2001) further support this position. As Yang, Wong, and Coid (2010) stated regarding nine adult risk assessment tools, the instruments were "essentially interchangeable" (p. 759) in predictive validity and only differed in their additional features and functions.

Instead, variations in predictive validity findings across instruments and jurisdictions may be more due to differences in staff training, system processes, data availability, or fidelity to the instrument rather than the actual items of the instrument. Therefore, given the myriad factors that can impact its implementation in the real world, it is important to revalidate an instrument for the specific jurisdiction periodically to ensure it remains accurate over time.

Equity: Group Differences

Sex

A higher percentage of males were rearrested than females. Females and males had similar overall and dynamic risk scores; however, females were more likely than males to be identified as lower risk in both overall and dynamic risk levels due to YASI's distinct cut-off points by sex. Almost half (44.7%) of females were assessed as low overall risk whereas only 23.2% of males were assessed as low overall risk. Given the lower baseline of rearrests for females as compared to males, this difference in risk level distribution was warranted. For instance, females and males with overall low risk levels had 15.8% and 16.8% 12-month rearrest rates, respectively, indicating that "overall low risk" had similar meanings between sexes. The

gap between rearrest rates widened for moderate (females: 31.7%; males: 37.0%) and high (females: 47.4%; males: 55.6%) overall risk levels, though, suggesting a slight shift in cut-off points may be warranted.

Additionally, there were significant differences in the prevalence of risk by domain. Females had statistically significantly higher scores in the *Family*, *School*, *Aggression*, and *Employment/Free Time* domains whereas males scored higher in the *Legal History*, *Community/Peers*, and *Alcohol/Drugs* domains. However, the cut-off points to convert scores to levels resulted in lower risk levels for females than males in all domains with the exception of no significant difference in the *Employment/Free Time* domain. Similar to the overall and dynamic risk, the different cut-off points to convert scores to levels in domains were an important function to account for the lower baseline rearrest rates for females.

Differences in rearrest distributions and some of the risk scores were expected given the sex variations in comorbid risk factors and interactions with the justice system (Reisig et al., 2006; Shepherd, et al., 2013; Thompson & McGrath, 2012). For instance, *Family* scores were higher for females than males, aligning with the literature citing poor family relationships and history of abuse as more prevalent and important factors for pathways to criminal behavior among female offenders (Daly, 1992; Daly, 1994; Reisig et al., 2006; Shepherd et al., 2013). However, *Alcohol/Drugs* is also a common risk area attributed to female offenders in the literature (Daly, 1992; Daly, 1994; Reisig et al., 2006), but the current study found higher scores among males. Interestingly, females scored higher in *Aggression*, yet males tend to exhibit more violent offending (Chesney-Lind et al., 2008), suggesting aggression in females may not present as violence. These mixed findings suggest the need for additional research focused not only on

the predictive validity of risk instruments by sex but also in the different characteristics, needs, and pathways among female offenders.

Regarding the research question of sex differences in predictive validity of overall and dynamic risk levels and domain risk levels, the hypothesis that YASI in Virginia would better predict reoffending among males than females was not supported (Hypothesis #2a). The predictive validity of the overall and dynamic risk scores and levels was statistically equivalent for males and females using AUCs. This result was congruent with many previous studies on other risk tools demonstrating comparable predictive validity between males and females (Jones, 2011; Olver et al., 2009; Orbis, 2007; Pusch & Holtfreter, 2018; Schwalbe, 2008; Shepherd et al., 2013; Smith et al., 2009) but not others (Anderson et al., 2016; Reisig et al., 2006), including a previous study specifically focused on YASI that found higher predictive accuracy for males than females (Jones et al., 2016).

The DIFR for the overall risk level was higher for males than females (0.64 and 0.53, respectively), though the significance of this difference is unknown. The DIFR measures separation and proportionality of groupings, and the lower DIFR for females was likely the result of the large proportion of females with low (44.7%) or moderate risk (47.0%) compared to high risk (8.4%). With so few females assessed as high risk, the DIFR value was lower for females than males. While this type of unbalanced risk level distribution could be problematic for classifying a system's overall population, it is appropriate for the female subgroup given the comparatively low baseline rate of female rearrests.

The hypothesis that the YASI domains of *Family* and *Community/Peers* would be stronger predictors of recidivism for females than males was not supported (Hypothesis #2b); these two domains were the only domains with statistically significant sex differences in AUC

values, but they were higher for males than females (*Family*: levels only; *Community/Peers*: scores and levels). Thus, the *Community/Peers* and *Family* domains may have important sex differences but in the opposite direction as expected. Males tended to have higher risk scores and levels in the *Community/Peers* domain, and the predictive validity was stronger for males. The *Family* domain levels also had stronger predictive validity for males, though females had higher scores and lower levels. This result is in opposition to previous research that suggested interpersonal family and peer relationships tended to be more predictive of offending for females than males (Anderson et al., 2016; Gavazzi et al., 2006; Hubbard & Pratt, 2002; Schmidt et al., 2011; van der Put et al., 2014). Thus, the possibility of sex differences in pathways to crime (Daly, 1992; Daly 1994) and criminogenic needs (Hollin & Palmer, 2006), and how those differences should be incorporated into the construction of risk assessment instruments, remains unclear.

Race

A higher percentage of Black youth were rearrested than White youth. Likewise, Black youth had statistically significantly higher overall and dynamic risk scores and levels than White youth. For instance, 23.6% of Black youth and 18.9% of White youth were assessed as high overall risk. Black youth also had higher domain scores and levels than White youth with the exception of the *Family* and *Alcohol/Drugs* domains. For the *Family* domain, White youth had higher scores but not significantly different levels; for the *Alcohol/Drugs* domain, White youth had higher scores and levels. These generally higher risk characteristics and higher rearrest rates among Black youth were aligned with previous research indicating Black youth experienced higher levels of risk factors (Chapman et al., 2006) and contact with the juvenile justice system (Kakar, 2006).

There were also some important differences within risk levels across racial groups. Compared to the differences by sex discussed previously, the rearrest rates within risk levels were more substantially different between races. For youth assessed as low overall risk, 21.7% of Black youth were rearrested while 13.0% of White youth were rearrested. Likewise, for moderate overall risk, 41.9% of Black youth and 30.0% of White youth were rearrested; for high overall risk, 59.2% of Black youth and 50.1% of White youth were rearrested. Similar gaps existed in the dynamic risk levels, with independent sample t-tests indicating statistically significant differences in rearrest rates by race within each overall and dynamic risk level. These differences indicate that *Low*, *Moderate*, or *High* risk labels represent higher rearrest rates for a Black youth than for a White youth, and, unlike for females and males, there are no distinct cut-off points for racial groups to adjust for these differences.

These differences within risk levels suggest the risk assessment is not capturing a factor that increases the likelihood Black youth are rearrested. One possible explanation may be that Black youth's risks and likelihood to reoffend are underrated by staff or under-scored by the tool compared to White youth. For example, the tool may not capture certain risk factors more often experienced by Black than White youth. This interpretation assumes that rearrests are an accurate proxy for actual delinquent behavior, and the disparity exists in the assessment of risk. However, since not every delinquent act results in an arrest, rearrests may not be a true representation of youth's actual delinquent behavior.

Instead, system bias or discrimination may be the cause of the rate differences, as suggested in previous research (Kempf-Leonard, 2007; Kochel et al., 2011). With the same assessed risk characteristics, a Black youth was more likely to be rearrested than a White youth, similar to Kochel and colleagues' (2011) meta-analysis findings that Blacks were 30% more

likely to be arrested even after controlling for other relevant factors. This interpretation assumes that risk characteristics are a better proxy for actual delinquent behavior (i.e., youth with similar risk characteristics reoffend at similar rates), and the racial disparity exists in the rearrest events.

By better understanding risk characteristics and their relationship to subsequent system involvement, future studies may be able to isolate the sources of disparities and help advance research on the prevalence and potential causes of disproportionality in the juvenile justice system. Until then, it is important that YASI did not over-score Black youth in respect to their likelihood to be rearrested as compared to White youth, indicating that the tool may help to reduce disproportionality in the decisions it informs by focusing on known risk characteristics.

Regarding the research question of racial group differences in predictive validity of overall and dynamic risk levels and domain dynamic risk levels, the hypothesis that YASI would perform comparably for White and Black youth was partially supported (Hypothesis #2c). The predictive validity was statistically equivalent for White and Black youth for overall risk levels and dynamic risk scores and levels; however, the AUCs for the overall risk scores were statistically significantly higher for White youth than Black youth. Given the importance of the overall risk level in the decisions affecting youth beyond case planning and service matching (e.g., assigning LOS for a commitment), a difference in scores' predictive validity could be impactful if it becomes large enough to impact the risk levels; predictive validity of both scores and levels should continue to be monitored by DJJ. Furthermore, there was a difference in both prevalence and predictive validity between White and Black youth regarding the *Alcohol/Drugs* domain, with Black youth scoring lower on those items and the domain scores and levels having statistically significantly lower AUC values.

The DIFR for the overall risk level was also higher for White youth than Black youth (0.68 and 0.58, respectively). Unlike the DIFR differences by sex, proportionality was not the likely explanation for the lower DIFR for Black youth. Instead, the separation of rearrest rates by risk level was smaller for Black youth than White youth, indicating less of a distinction between risk levels in predicting reoffending. For example, the rearrest rate for Black youth assessed as high risk (59.2%) was 2.7 times higher than for those assessed as low risk (21.7%). Conversely, the rearrest rate for White youth assessed as high risk (50.1%) was 3.9 times higher than for those assessed as low risk (13.0%). Again, this difference may be a result of additional risk factors (including systemic bias) experienced by Black youth that were not captured in the assessment, making it more difficult for the tool to cleanly predict rearrests and separate the risk classification levels.

Overall, the instrument's predictive validity was comparable between Black and White youth according to the AUCs. The instrument did not exhibit racial bias that would inaccurately exaggerate assessed risk for Black youth to be rearrested and may even help to ameliorate disproportionality in the system. However, additional risk factors experienced by Black youth, including those relating to systemic bias, may not be fully captured, as indicated by the rearrest differences by level and DIFR values.

Unfortunately, the racial disproportionality in the juvenile justice system cannot be fully removed from the context of the risk assessment. As Thompson and McGrath (2012) cautioned, legal history items in a risk assessment instrument may be based on a biased system, and the prediction of a future rearrest reflects an event within the same biased system; even if the risk factors and predictive ability are accurate, it is a projection modeled on those biases. Therefore, in order to improve an instrument's predictive validity for Black youth and its potential for

reducing disparities in the system, it is important to continually examine potential racial disparities in individual items and modify or remove those related more to race than reoffending. Risk assessments have the potential to help inform and reduce racial disproportionality in the juvenile justice system through the continued study of offender characteristics in relation to racial differences in system involvement and by providing the foundation for more objective decision-making aligned with individualized risk and needs.

Usage: Priority Areas

Almost all (96.0%) of cases had at least one case planning priority area assigned. The consistent significant correlations between domain risk level and assigned priority areas indicated that staff used YASI results for case planning, with the strongest correlation for the *Alcohol/Drugs* domain ($r = 0.59$), followed by *Aggression* ($r = 0.42$). However, some of these correlations were weak. The *Employment/Free Time* domain in particular demonstrated the least congruence. Of all the domains, it had the lowest prevalence of moderate to high risk (15.7%), lowest occurrence of assigned priority area (6.0%), and the lowest correlation between domain level and priority area assignment ($r = 0.15$). The correlations of the eight domains (*Legal History* is excluded due to its static nature; *Mental Health* is excluded because it does not have an associated risk level) were ordered from strongest to weakest as follows:

- Alcohol/Drugs ($r = 0.59$)
- Aggression ($r = 0.42$)
- Community/Peers ($r = 0.36$)
- Attitudes ($r = 0.33$)
- Skills ($r = 0.31$)
- School ($r = 0.27$)

- Family ($r = 0.23$)
- Employ./Free Time ($r = 0.12$)

Regarding the research question of the congruence between assigned case planning priority areas and higher dynamic risk level domains, the hypothesis that discrepancies would exist between the higher risk domains and the assigned priority areas in the case plan was partially supported (Hypothesis #3). Initial findings suggested that YASI results were being used to inform case planning. Each domain had a positive correlation between risk and priority area assignment, though some were stronger correlations than others. With statistically significant correlations ranging from 0.15 to 0.59, there was a large variation in how domains were considered in the case plan, indicating that staff may emphasize some risk areas over others or value YASI results differently.

This variation may be appropriate if some domains were more predictive of rearrests, suggesting that interventions targeting those areas may have the most impact on rehabilitation. However, *Alcohol/Drugs* and *Aggression* were the strongest correlations between risk and priority area assignment, but the *Community/Peers* domain had the highest predictive ability (other than *Legal History*). Thus, staff may be partially using YASI results to inform their case planning while also relying on their professional opinions to value certain domains (e.g., *Alcohol/Drugs* and *Aggression*) over others that may be more important (e.g., *Community/Peers*). The *Employment/Free Time* domain exhibited low prevalence, low predictive validity, and low congruence with priority areas, suggesting that staff may not rate or prioritize this domain in a meaningful way for case planning purposes.

The case planning component of YASI as a fourth generation tool is critical to the choice in risk instrument, and it is important to note that this analysis serves as an exploratory first step

in looking into this feature of YASI. DJJ has incorporated YASI into its case planning policies and procedures, signifying its intention to use the instrument for the RNR model and not simply to predict reoffending. The relationships between domain risk, reoffending, and case planning require additional investigation given the promising yet inconsistent findings regarding priority area assignment.

As discussed previously, interventions utilizing the risk principle and the RNR model are effective at reducing recidivism (Andrews et al., 2006; Koehler et al., 2012; Luong & Wormith, 2011; Vieira et al., 2009); however, risk instruments are not always effectively utilized to inform these services (Singh et al., 2014; Vieira et al., 2009; Viljoen et al., 2018). Data collection that does not contribute to the decision-making of cases is wasted time and effort, and the findings suggest a mixed application of assessment results toward case planning, warranting further study and additional staff training.

Limitations

It is important to note several limitations to the current study. First, the study did not focus on all elements of YASI. More specifically, it did not account for the individual item data from the assessments. The items should relate to their associated domain and be distinct from the other domains so that the domains represent separate concepts that can effectively inform interventions. Without this information, the current study relied on the structure of the YASI tool to weight the items and load onto the domains, scores, and levels. Changes to weighting and scoring formulas could impact the influence specific items have on the overall scores and alter predictive validity of the tool. The study also focused on risk rather than protective factors, which constitute a major portion of the YASI results. Protective factors represent the responsivity elements of the instrument, which may be used for case planning and identifying

appropriate services. Thus, a priority area could be assigned based on a protective factor rather than risk. Without considering protective factors, the usage research question examining the congruence between domain risk and priority areas is limited in its interpretation.

Second, the study did not focus on inter-rater reliability. The data were collected in a real-world setting rather than in a controlled study environment. Staff completing assessments in a uniform way across the population is key for a risk assessment instrument to function as intended. If an assessment has low inter-rater reliability, findings regarding predictive validity might be compromised because any failures may be due to poor ratings rather than the structure of the tool itself.

Third, the data analyses across research questions represented preliminary investigations. Alternative methodologies would provide additional information and insights into the various elements, outputs, and uses of the assessment tool. Each domain's predictive validity was tested independently using the AUC; the interaction of domains in their ability to predict reoffending was not examined. If the domains were not completely distinct concepts or had interacting patterns, their contributions to the predictive ability of the instrument may differ from the independent AUC results. Furthermore, the priority area analysis does not account for the entire RNR profile of each case. For instance, if a youth scored high risk in every domain, some of those domains would not be assigned as priority areas due to the realistic capacity of an individual case plan.

Finally, the study was limited by its reliance on 12-month rearrest rates. Alternative timeframes and other definitions of reoffending (e.g., reconviction, self-report) might provide alternative results. For racial differences, in particular, reconviction could provide important insights in relation to the increasingly compounding disparity that exists at subsequent decision

points in the system. Additionally, rearrests indicate a charge for a new offense but do not necessarily represent that the individual actually committed the offense; incorporating self-report data to reflect delinquent behavior could contribute to the interpretations.

Future Research

There are several potential areas for future research on the use of YASI in Virginia, including instrumental validity, protective factors, inter-rater reliability, domain interactions and clusters, reoffense types and timing, additional group and geographical differences, weighting and scoring, service matching, recidivism reduction, and program evaluations. Ultimately, researchers should focus on studying ways to improve and enhance risk assessment tools' accuracy, equity, and usage in the real-world environment of juvenile justice systems.

In order to address the limitations described above, a study of instrumental validity is needed to determine if the individual items load onto their designated domains and if the domains are separate and distinct concepts. For example, factor analysis may determine that the items in the *Aggression* and *Skills* domains are a single concept rather than distinct domains. Additional analyses focused on protective factors and their relationship to the risk factors would also be beneficial.

An inter-rater reliability study would also be beneficial. Inter-rater reliability was satisfactory in early stages of Virginia's YASI implementation (Baird et al., 2013), but follow-up is necessary to determine if continued training efforts have maintained consistency in assessments. Poor inter-rater reliability can negatively impact the predictive validity of the tool, so measuring this aspect of implementation is important.

In order to investigate the interactions between domains, several analyses could be planned. First, cluster analysis could determine potential patterns in how the items or domains

present, identifying potential typologies of offenders with shared risk profiles. As indicated by the findings of domain differences by sex, cluster analysis may help identify distinct characteristics of female and male offenders. Clusters may also assist in case planning by aiding in the selection of effective service options that address common combinations of criminogenic needs. Over time, the tracking of successes or failures of youth in particular clusters and receiving different services may further facilitate case planning by providing staff with a suggested optimal service for a specific risk profile.

Second, a binary logistic regression would show the ability of each domain to predict reoffending as a comprehensive model rather than individual variables. Wald and significance values from preliminary binary logistic regression analyses suggested that the *Legal History* domain far outperformed the other domains in predicting 12-month rearrests. *Alcohol/Drugs* and *Community/Peers* domains followed in importance. Finally, *Aggression*, *School*, and *Attitudes* contributed slightly to the predictive ability, and *Family*, *Skills*, and *Employment/Free Time* were not statistically significant when all other domains were equal. These findings differ from the more uniform results of the AUC values and suggest further investigation is needed to determine the interaction between domains.

Differences by sex and race could also be examined through binary logistic regressions. Preliminary models for each of the four subgroups indicated additional differences in patterns that deserve further research to determine possible group differences in pathways to crime and criminogenic needs. For example, the *Community/Peers* was not statistically significant for females but was the second most important variable in the model for males. Similarly, the *School* domain was statistically significant for White youth but not Black youth, and the *Alcohol/Drugs*

domain was statistically significant for both Black and White youth but substantially more so for White youth.

In addition to addressing the limitations of the study, future research may expand the knowledge surrounding youth risk assessment instruments and YASI in Virginia by investigating reoffense severities and multiple reoffenses over time. This approach could examine reoffending as more than a binary outcome but rather a complex series of events that may indicate trajectories of increasing severity of criminal behavior or a desistance in offending. It could also inform risk assessment tools' ability to identify more specific risk such as violent reoffending or sexual reoffending.

Further study of group differences is also needed regarding both prevalence and predictive validity. As described above, some findings regarding sex and race differences in risk prevalence and rearrest rates require additional focus. In addition to race and sex differences, other populations may have varied risk assessment results. For example, differences may exist by age group that could impact the best approach for case planning, particularly in systems that serve a wide range of ages (e.g., Virginia serves youth from age 8 or younger to their 21st birthday [DJJ, 2016]). Socioeconomic status may also impact assessed risk, particularly in the *Family* domain items regarding the amount of adult supervision and family supports. Youth from single-parent households or with parents working multiple jobs and long hours may be disproportionately assessed as higher risk due to lack of supervision. Similarly, ethnicity or cultural groups may have distinct views and attitudes toward education, supervision, authority figures, or community involvement, potentially impacting how these factors are assessed and how predictive they are of reoffending. Furthermore, multilevel models are needed to study locality differences in jurisdictions such as Virginia that serve a large area with diverse

demographics and socioeconomic statuses by region. There may be variation in risk prevalence and predictive validity between localities that are rural versus urban or low versus high income. Sex, race, and other group risk differences across these settings may also change, particularly with the potential impact of resource availability and local policies and practices.

Based on these proposed analyses, studies on potential modifications to the tool should be completed. Further investigation is required regarding the weighting and scoring of the assessment in relation to the predictive power of items and domains, both overall and for specific subgroups. If there are items or factors included in the tool that do not substantively contribute to either reoffense prediction or effective case planning, they should be removed; likewise, factors not currently in the tool may need to be tested or added. The item weights should be representative of the relative predictive power without unnecessarily contributing to disparities between groups. For example, characteristics common among certain racial, socioeconomic, or ethnic or cultural groups that are perceived as risk factors (e.g., level of adult supervision in single-parent households) should be included only if they significantly predict reoffending after controlling for the group differences. Otherwise, the instrument may disproportionately assess a vulnerable or minority group as higher risk due to their circumstances rather than evaluate individual criminogenic risks and needs, resulting in additional disadvantages during the risk-informed decisions in the system.

The use and impact of the tool in the real-world setting is also important to study. Inter-rater reliability and predictive validity evaluate the efficacy of risk assessment tools to accomplish the primary goal of identifying the likelihood of recidivism; however, third and fourth generation tools are meant to a) guide intervention strategies, and b) reduce recidivism. The tool itself is not an intervention, so it is ineffective if not used to inform case planning

decisions. Therefore, future research should examine whether these risk-reduction assessment tools are actually used by practitioners to identify and deliver services and dosages that match to the RNR profile of youth.

This study indicated a limited level of congruence between risk levels and case planning, but the impact of matched service delivery on outcomes should be examined. Because some factors that are included in risk instruments are not highly related to future reoffending (Baird et al., 2013), it is not clear that targeting them for services would effectively reduce recidivism. Thus, a study on the impact of the use of risk assessment tools on reoffending would be beneficial. Ideally, a study including random assignment to either a second generation tool or a third or fourth generation tool to determine if there were differences in subsequent recidivism would help to determine if these additional factors were helpful to assess; however, this design is unlikely in the real-world setting and alternative methods may be necessary.

Finally, risk assessment data could be used for program evaluations by studying changes in dynamic risk over time. These tools provide an opportunity to assess a program not just with typical recidivism outcomes, but with the more sensitive milestones of decreased needs or increased strengths of the program participants. Thus, third or fourth generation tools provide important measures at both the individual-level and program- and system-wide levels for evaluating progress.

Policy Recommendations

There are several policy recommendations stemming from this study. The following recommendations can be applied to both DJJ specifically and to juvenile justice systems in general: determine purpose and function, conduct staff and stakeholder training, test, calibrate and modify, and repeat. A summary of the findings, future research topics, and policy

recommendations is included in a report to DJJ as Appendix C using their required one-page template, which will be provided to the agency along with the full dissertation.

Determine Purpose and Function

Juvenile justice agencies should determine the desired purpose and function of risk assessments within their systems before selecting a tool to ensure that they are using the most appropriate instrument for each decision point. They should consider the tool's generation, origin, and norms in relation to the agency's scope and population.

In the decision between generations, jurisdictions must weigh the benefits and downfalls of collecting large amounts of data, taking into consideration staff training and workloads. Depending on the specific characteristics of the organization and its policies, a fourth generation tool may not be feasible to implement in the desired settings due to the larger number of items that require skilled interviewing rather than simple collection of data from written records. Therefore, characteristics of specific measures must be considered from the perspective of the purpose fulfilled for the agency. Importantly, regardless of generation, agencies should avoid adding complexities that do not add to either predictive validity or quality of case planning.

For example, DJJ currently utilizes YASI not only to predict reoffending but also to target interventions and improve risk factors for youth on community supervision or in secure settings. The shorter pre-screen performed as well or better than the full assessment's dynamic risk, indicating the additional items in the full assessment do not improve predictive accuracy. However, given the various decision points in which DJJ uses risk assessment results, a tool with both a classification and service matching function is needed. Data indicated that staff are somewhat utilizing the results of the assessment to set youths' priority areas for services, which justifies the inclusion of some additional items beyond those that strongly predict recidivism. It

is recommended that DJJ continue using the YASI pre-screen as a classification tool and the full assessment as a fourth generation tool to inform case planning. If an agency decided to use an instrument solely as a classification tool at a different stage of the system (e.g., identifying diversion-appropriate cases), a second generation tool would be preferred for maximizing predictive validity while minimizing workload. For DJJ, the YASI pre-screen, providing the overall risk level, could likely serve this simpler function.

Based on available research, the benefits of selecting either a locality-specific tool or a generic commercial tool are mixed and, again, depend on the specific circumstances of the organization. Instruments created specifically for a jurisdiction are more easily modified and adjusted as reliability and validity findings inform improvements, but they require existing data and analysis expertise. Generic, commercial tools, on the other hand, can be implemented quickly but require validation for the specific population in the locality or state. Regardless, systems should consider customizing the tool, whether created for the jurisdiction or for a different population, to create risk levels that make sense for their youth and their priorities to protect the public and provide services.

Jurisdictions must consider the norms of the instrument and how those will be reflected in the distribution of their population. Most importantly, an organization should select a tool that can differentiate between the risk levels of the organization's specific population relative to the base rate of reoffending. Systems focused on the front-end of the system (e.g., identifying diversions) may need a different tool or different norms than an organization focused on the deep-end of the system (e.g., deciding releases from secure confinement). If they used the same tool, the former organization might identify the majority of their population as low risk while the latter might identify the majority of their population as high risk. Without consideration for their

unique population, the instrument lacks the differentiation in categorization that makes the tool useful.

Organizations serving all stages of the system may consider different risk tools or different cut-off points for different populations (e.g., diversions versus commitments) in order to achieve differentiation between risk levels. If the instrument is to be used among only juveniles in the deepest end of the system, primarily high risk, then the usefulness of an instrument that classifies all youth as high risk is minimal. Conversely, a system like DJJ, with oversight from intake through parole, may instead aim to use a single risk assessment tool throughout all stages from the front- to the deep-end of the system to help make uniform decisions in service provision; these agencies may be less persuaded by category differentiation and more interested in the needs and responsivity information provided by a third or fourth generation tool.

This first step toward utilizing a risk assessment instrument must be undertaken conscientiously as a starting point for purchasing or creating the "right" tool. If completed carelessly and an unsuitable instrument selected, the costs and effort required to launch the assessment practices and execute the remaining recommendations could be for naught.

Conduct Staff and Stakeholder Training

Sufficient training of staff on the instrument is essential. Viljoen and colleagues (2018) found that risk management training and guidelines for staff may improve adherence to the RNR model in case planning. Staff must be trained on how to accurately and consistently capture the responses to the instrument's items, including strategies for discovering sensitive risk factors (e.g., drug use). In addition, racial biases must be openly discussed in training along with tactics

for overcoming these tendencies. As discussed above, inter-rater reliability is a key for a tool to be used effectively across a population.

Training must also include steps to be taken after the assessment is completed, including identifying and providing the services that match the needs identified in the evaluation. As discussed earlier, a comprehensive risk assessment is wasteful if not acted upon. Staff should also learn to monitor individual youth progress via periodic reassessments using a tool that includes dynamic factors that can change over time, adjusting the delivery of services accordingly across a continuum of interventions and dosages. The concept that a risk assessment tool as part of an RNR-focused system could reduce recidivism can only be tested if staff are following through with appropriate service matching in the case plan.

Additionally, ongoing training is necessary to sustain fidelity. The similar results in predictive validity (i.e., AUC and DIFR) compared to the study conducted using FY 2009 data (Baird et al., 2013) suggested that Virginia successfully sustained training through several years of implementation. Part of this training must include the development of buy-in so that the tool becomes a useful piece of their professional decision-making process rather than a burden that may be ignored. Also, agencies should continue valuing and relying on the expertise of the professionals in the organization. Risk assessment tools are not perfectly accurate, and the success of the intervention is partly due to the aptitude and abilities of the person delivering the services. Andrews and colleagues (1990) discussed the professional override as an important component of the case planning process in which the professional considers the risk, needs, and responsivity of the offender in conjunction with the specific situations and conditions to decide on the most appropriate intervention.

In addition to training the staff directly responsible for conducting the assessments and developing the case planning, it is important to invest in the training of stakeholders. These stakeholders include the agency's administration and support staff; external partners within the justice system such as judges, prosecutors, and defense attorneys; and other youth-serving agencies and organizations (e.g., Department of Social Services, Department of Behavioral Health and Developmental Services). This training should be focused on the basic RNR model concepts, the structure and utility of the selected risk assessment tool, and the limitations of the instrument. The individuals responsible for the decision-making of policies, procedures, and court case processing should be aware of these topics, and those working with court-involved youth should understand how to interpret and utilize the results of a risk assessment for the individuals they serve. As stakeholders across various capacities better understand the RNR model and risk assessment tools, the multi-discipline systems that serve youth and their communities may better optimize outcomes.

As the literature and the current study discussed, youth risk assessments generally perform at a moderate level of predictive validity. They are limited in their ability to accurately predict reoffending and inform case planning and services, but they are the best practice available for system-involved youth. This transparency regarding the benefits and limitations of the tool will help construct a common language and understanding throughout the system and improve buy-in from all parties.

Test

Agencies should incorporate the use of the risk assessment instrument in the organization's policies and procedures, including periodic reassessments for offenders in systems using the tool to guide programming. Once implemented, agencies should partner with

researchers to test and retest the use of the tool for its population. This research should focus on predictive validity, inter-rater reliability, use of the tool for case planning and service delivery, and impacts on recidivism.

Predictive validity and inter-rater reliability are well established components for studying an instrument and require periodic retesting to ensure consistency over time as staff, youth, and systems may change. Special attention to any differences by race or sex is needed to monitor the equity of the tool. Agencies and researchers should also work to determine if systems that use these instruments actually apply the results to the case management decisions of service delivery and obtain positive outcomes. Although research has demonstrated that both adult and juvenile correctional programs adhering to the risk principle (Lipsey, 2009; Lowenkamp, Latessa, & Holsinger, 2006) as well as the RNR model (Andrews et al., 2006; Koehler et al., 2012; Vieira et al., 2009) resulted in larger reductions of recidivism, Singh and colleagues (2014) found there was limited application of the strengths and needs identified in the assessment to the types of services provided. Future studies should expand this research to investigate the connection of third and fourth generation risk assessment instruments to the implementation of RNR-matched interventions and dosages, examining both individual-level and system-level recidivism results.

Calibrate and Modify

As each system is different with varying populations, it may be necessary to customize the tool once sufficient data is collected. Ideal cut-off points to maximize predictive validity and differentiation between risk levels may not be identical across all systems using an off-the-shelf tool. As baseline rates differ, so too should cut-off points. Calibrating these values to optimize the tool for the system should be a priority for the agency after implementing a risk tool. However, systems should be careful to avoid using different cut-off points to disguise inequities

between subgroups. Females and White youth had lower baseline rearrest rates than males and Black youth, respectively, yet cut-off points are different only between sexes. These cut-off points are arguably justified by sex differences in risk or delinquency experiences, but the same method would be inappropriate and discriminatory for racial groups, resulting in more intensive interventions (including sanctions) for Black youth with equivalent risk characteristics.

There also may be item or domain selection and weighting changes required. For example, given the low prevalence, low predictive validity, and low congruence with priority areas, DJJ may reconsider whether *Employment/Free Time* should remain as a domain of the assessment. The recommended additional research regarding instrumental validity and item and domain interactions may inform additional modifications. Any modifications should also incorporate the considerations regarding potential group disparities as outlined above in the discussion on future research.

Risk assessment tools should not be viewed as an unalterable constant. While changing the tool may be challenging for staff and stakeholders, the ongoing training should emphasize that this instability is an anticipated and beneficial part of the process in order to maintain buy-in and understanding of the modifications. Furthermore, if advances to risk assessment instruments are achieved (e.g., cluster analysis with service recommendations), adding these features to the tool may improve its usability and value.

Repeat

The key to these recommendations is that they are not static, one-time decisions and actions. The recommendations above represent a cycle that should be repeated and reassessed periodically in order to maintain a risk assessment practice that works for the changing needs and populations of each system.

The philosophy of juvenile justice systems has changed over time, and, as a result, the populations and services have also evolved. These developments are no doubt going to continue to progress in the future and impact the optimal practice regarding risk assessment tools. Thus, the purpose and function of a risk assessment tool are not constants but, rather, might adjust along with the mission and needs of an agency.

Staff and stakeholder training is never complete. People must be informed of any changes regarding both practice and the tool itself, and turnover is always an added challenge for consistent implementation. Similarly, the recommended research for monitoring and altering the tool must be conducted recurrently to ensure the validity, reliability, and application to service planning are maintained throughout shifting populations, staff, and overall practices over time.

Conclusions

This study aimed to examine YASI's accuracy, equity, and usage in order to inform whether Virginia's selection and implementation of the tool is appropriately serving its purpose of 1) accurately predicting the likelihood of reoffending, 2) standardizing decisions by using a consistent tool, and 3) informing interventions for effective rehabilitation. Overall, the findings indicated adequate levels of overall predictive validity in comparison to the field of risk assessment research, general equity in predictive validity between sexes and races with areas of further study needed, and positive relationships of varied strengths between identified risks and case planning priority areas. These findings suggest that YASI is an appropriate tool for Virginia's juvenile justice system, but additional research and training is needed to improve its implementation and optimize its utilization.

With the help of researchers, systems should continue to push risk assessment tools from merely predicting recidivism to aiding in the prevention of recidivism through the utilization of

their measures in developing case management plans and service delivery. Those who argue against these types of instruments due to their complexities and lack of added predictive value fail to consider this second purpose of guiding intervention case planning and ultimately reducing recidivism, a mission of most, if not all, juvenile justice systems. The more expansive instruments provide a more comprehensive profile for the various stakeholders in the public safety and human services sectors (e.g., attorneys, judges, probation managers, social service case workers) to make informed decisions with the goal of improving outcomes. The field of youth risk assessments is still growing and evolving, and it is imperative that practitioners and researchers partner to continue progressing RNR-focused practices and improving outcomes for youth.

References

Anderson, V. R., Davidson, W. S., II, Barnes, A. R., Campbell, C. A., Peterson, J. L., & Onifade, E. (2016). The differential predictive validity of the Youth Level of Service/Case Management Inventory: the role of gender. *Psychology, Crime, & Law, 22*, 666-677.

Andrews, D. A., & Bonta, J. (1995). *Level of Service Inventory–Revised*. Ontario, Canada. Multi-Health Systems.

Andrews, D. A., & Bonta, J. (2010). *The psychology of criminal conduct (5th ed.).* New Providence, NJ: LexisNexis.

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, S. J. (2006). The recent past and near future of risk and/or need assessment. *Crime and Delinquency*, *52*, 7-27.

Atkins v. Virginia, 536 U.S. 304 (2002)

Baglivio, M. T., & Jackowski, K. (2013). Examining the validity of a juvenile assessment instrument across gender and race/ethnicity. *Youth Violence and Juvenile Justice, 11*, 26-43.

Baird, C. (2009). *A question of evidence: A critique of risk assessment models used in the justice system.* Madison, WI: National Council on Crime and Delinquency.

Baird, C., Healy, T., Johnson, K., Bogie, A., Dankert, E. W., & Scharenbroch, C. (2013). *A comparison of risk assessment instruments in juvenile justice.* Washington, DC: National Council on Crime and Delinquency.

Benekos, P. J., & Merlo, A. V. (2005). Reaffirming juvenile justice: Strategies for the future. In R. Muraskin & A. R. Roberts (Eds.), *Visions for Change: Crime and Justice in the Twenty-First Century* (4th ed.; pp. 17-42), Upper Sadler River: Pearson.

Bonta, J., & Andrews, D. A. (2007). *Risk-need-responsivity model for offender assessment and rehabilitation*. Ottawa, Ontario: Public Safety Canada.

Brannen, D. N., Salekin, R. T., Zapf, P. A., Salekin, K. L., Kubak, F. A., & DeCoster, J. (2006). Transfer to adult court: A national study of how juvenile court judges weigh pertinent *Kent* criteria. *Psychology, Public Policy, and Law, 12*, 332-355.

Chapman, J. F., Desai, R. A., Falzer, P. R., & Borum, R. (2006). Violence risk and race in a sample of youth in juvenile detention: The potential to reduce disproportionate minority contact. *Youth Violence and Juvenile Justice, 4*, 170-184.

Chesney-Lind, M., Morash, M., & Stevens, T. (2008). Girls' troubles, girls' delinquency, and gender responsive programming: A review. *The Australian and New Zealand Journal of Criminology, 41*, 162-189.

Coid, J. W., Yang, M., Ullrich, S., Zhang, T., Sizmur, S., Farrington, D., & Rogers, R. (2011). Most items in structured risk assessment instruments do not predict violence. *Journal of Forensic Psychiatry and Psychology, 22*, 3-21.

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.

Daly, K. (1992). Women's pathways to felony court: Feminist theories of lawbreaking and problems of representation. *Southern California Review of Law and Women's Studies, 2*, 11-52.

Daly, K. (1994). *Gender, crime, and punishment*. New Haven: CT: Yale University Press.

Emeka, T. Q., & Sorenson, J. R. (2009). Female juvenile risk: Is there a need for gendered assessment instruments? *Youth Violence and Juvenile Justice, 7*, 313-330.

Fraser, M. W. (Ed.). (2004). *Risk and resilience in childhood: An ecological perspective* (2nd ed.). Washington, DC: NASW Press.

Gavazzi, S. M., Yarcheck, C. M., & Chesney-Lind, M. (2006). Global risk indicators and the role of gender in a juvenile detention sample. *Criminal Justice and Behavior, 33*, 597–612.

Geck, C. M. R. (2012). The Youth Assessment Screening Instrument: A psychometric evaluation with Canadian male youthful offenders. (Unpublished master's thesis). Carleton University, Ottawa, Canada.

Gendreau, P., Goggin, C., & Smith, P. (2002). Is the PCL-R really the "unparalleled" measure of offender risk?: A lesson in knowledge cumulation. *Criminal Justice and Behavior, 29*, 397–426.

Gendreau, P., Little, T., & Goggin, C. (1996). A meta-analysis of the predictors of adult offender recidivism: What works. *Criminology, 34*, 557–607.

Graham v. Florida, 130 S. Ct. 2011 (2010)

Grieger, L., & Hosser, D. (2014). Which risk factors are really predictive? An analysis of Andrews and Bonta's "central eight" risk factors for recidivism in German youth correctional facility inmates. *Criminal Justice and Behavior, 41*, 613-634.

Grove, W. M., & Meehl, P. E. (1996). Comparative efficiency of informal (subjective, impressionistic) and formal (mechanical, algorithmic) prediction procedures: The clinical-statistical controversy. *Psychology, Public Policy, and Law, 2*, 293–323.

Hamilton, Z., van Wormer, J., & Barnoski, R. (2015). *PACT validation and weighting results: Technical report.* Pullman, WA: Washington State University.

Harris, M. (2013). An exploration of strengths, promotive, and protective factors among justice involved youth. (Unpublished master's thesis). Carleton University, Ottawa, Canada.

Hjalmarsson, R. (2009). Juvenile jails: A path to the straight and narrow or to hardened criminality? *Journal of Law and Economics, 52*, 779-809.

Hoge, R. D., & Andrews, D. A. (2001). *The youth level of service/case management inventory (YLS/CMI): Intake manual and item scoring key*. Carleton University.

Hollin, C. R., & Palmer, E. J. (2006). Criminogenic need and women offenders: A critique of the literature. *Legal and Criminological Psychology, 11*, 179-195.

Hubbard, D. J., & Pratt, T. C. (2002). A meta-analysis of the predictors of delinquency among girls. *Journal of Offender Rehabilitation, 34*, 1-13.

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

In re Winship, 397 U.S. 358 (1970)

Izzo, R. L., & Ross, R. R. (1990). Meta-analysis of rehabilitation programs for juvenile delinquents: A brief report. *Criminal Justice and Behavior, 17*, 134-142.

Jackson v. Hobbs, 132 S. Ct. 1733 (2012)

Jones, D. A., Johnson, S., Latessa, E. J., & Travis, L. F. (1999). Case classification in community corrections: Preliminary findings from a national survey. In *Topics in community corrections: Classification and risk assessment* (pp. 4–10). Washington, DC: Department of Justice.

Jones, P. R., Harris, P. W., Fader, J., & Grubstein, L. (2001). Identifying chronic juvenile offenders. *Justice Quarterly, 18*, 479–507.

Jones, N. J. (2011). Merging theoretical frameworks to inform risk assessment for the young female offender. (Unpublished master's thesis). Carleton University, Ottawa, Canada.

Jones, N. J., Brown, S. L., Robinson, D., & Frey, D. (2016). Validity of the Youth Assessment and Screening Instrument: A juvenile justice tool incorporating risks, needs, and strengths. *Law and Human Behavior, 40*, 182-194.

Juvenile Justice Geography, Policy, Practice, & Statistics. (2017). *Evidence-based practices: Risk instruments.* Retrieved from http:// www.jjgps.org/juvenile-justice-services#evidence-based-practices?tabId=2&view=risk-instruments

Kakar, S. (2006). Understanding the causes of disproportionate minority contact: Results of focus group discussions. *Journal of Criminal Justice, 34*, 369-381.

Kempf-Leonard, K. (2007). Minority youths and juvenile justice: Disproportionate minority contact after nearly 20 years of reform efforts. *Youth Violence and Juvenile Justice, 5*, 71-87.

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

Kochel, T. R., Wilson, D. B., & Mastrofski, S. D. (2011). Effect of suspect race on officers' arrest decision. *Criminology, 49*, 473-512.

Koehler, J. A., Lösel, F. A., Akoensi, T. D., & Humphreys, D. K. (2012). A systematic review and meta-analysis on the effects of young offender treatment programs. *Journal of Experimental Criminology, 9*, 19–43.

Krysik, J., & LeCroy, C.W. (2002). The empirical validation of an instrument to predict risk of recidivism among juvenile offenders. *Research on Social Work Practice, 12*, 71–81.

Latessa, E. J., & Lovins, B. (2010). The role of offender risk assessment: A policy maker guide. *Victims & Offenders, 5*, 203-219.

Lipsey, M. W. (2009). Primary factors that characterize effective interventions with juvenile offenders: A meta-analytic overview. *Victims & Offenders, 2*, 124–147.

Lowenkamp, C. T., & Latessa, E. J. (2005). Increasing the effectiveness of correctional programming through the risk principle: Identifying offenders for residential placement. *Criminology & Public Policy, 4*, 262-290.

Lowenkamp, C. T., Latessa, E. J., & Holsinger, A. (2006). The risk principle in action: What we have learned from 13,676 offenders and 97 correctional programs. *Crime and Delinquency, 52*, 1–17.

Luong, D., & Wormith, J. C. (2011). Applying risk/need assessment to probation practice and its impact on the recidivism of young offenders. *Criminal Justice and Behavior, 38*, 1177-1199.

Makarios, M., Sperber, K. G., & Latessa, E. J. (2014). Treatment dosage and the risk principle: A refinement and extension. Journal of Offender Rehabilitation, 53, 334-350.

Martinson, R. (1974). What works? Questions and answers about prison reform. *The Public Interest, 35*, 22.

Martinson, R., (1979). New findings, new views: A note of caution regarding sentencing reform. *Hofstra Law Review, 7*, 243-258.

McGowan, A., Hahn, R., Liberman, A., Crosby, A., Fillilove, M., Johnson, R., et al. (2007). Effects on violence of laws and policies facilitating the transfer of juveniles from the juvenile justice system to the adult justice system. *American Journal of Preventive Medicine, 32*, S7-S24.

McGrath, A., & Thompson, A. P. (2012). The relative predictive validity of the static and dynamic domain scores in risk-need assessment of juvenile offenders. *Criminal Justice and Behavior, 39*, 250-263.

McKeiver v. Pennsylvania, 403 U.S. 528 (1971)

Mears, D., Cochran, J., Greenman, S., Bhati, A., & Greenwald, M. (2011). Evidence of effectiveness of juvenile court sanctions. *Journal of Criminal Justice, 39*, 509-520.

Merlo, A. V., & Benekos, P. J. (2010). Is punitive juvenile justice policy declining in the United States? A critique of emergent initiatives. *Youth Justice, 10*, 3-24.

Miller, J., & Lin, J. (2007). Applying a generic juvenile risk assessment instrument to a local context. *Crime & Delinquency, 53*, 552-580.

Miller v. Alabama, 132 S. Ct. 2455 (2012)

Olver, M. E., Stockdale, K. C., & Wong, S. C. P. (2012). Short and long-term prediction of recidivism using the Youth Level of Service/Case Management Inventory in a sample of serious young offenders. *Law and Human Behavior, 36*, 331-344.

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.

Orbis Partners Inc. (2007). *Long-term validation of the Youth Assessment and Screening Instrument (YASI) in New York State juvenile probation.* Ottawa, Ontario: Author.

Orbis Partners Inc. (2017a). *Assessment.* Retrieved from http://orbispartners.com/assessment/

Orbis Partners Inc. (2017b). The Youth Assessment and Screening Instrument (YASI) from Orbis Partners [Video]. Retrieved from http://orbispartners.com/assessment/youth-assessment-yasi/

Pusch, N., & Holtfreter, K. (2018). Gender and risk assessment in juvenile offenders: A meta-analysis. *Criminal Justice and Behavior, 45*, 56-81.

Reisig, M. D., Holtfreter, K., & Morash, M. (2006). Assessing recidivism risk across female pathways to crime. *Justice Quarterly, 23*, 384-405.

Rice, M. E., & Harris, G. T. (2005). Comparing effect sizes in follow-up studies: ROC area, Cohen's d, and r. *Law and Human Behavior, 29*, 615-620.

Roper v. Simmons, 543 U.S. 551 (2005)

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.

Schwalbe, C. S. (2007). Risk assessment for juvenile justice: A meta-analysis. *Law and Human Behavior, 31*, 449-462.

Schwalbe, C. S. (2008). A meta-analysis of juvenile justice risk assessment instruments: Predictive validity by gender. *Criminal Justice and Behavior, 35*, 1367-1381.

Schwalbe, C. S., Fraser, M. W., Day, S. H., & Arnold, E. M. (2004). North Carolina Assessment of Risk (NCAR): Reliability and Predictive Validity with Juvenile Offenders. *Journal of Offender Rehabilitation, 40*, 1-22.

Schwalbe, C. S., Fraser, M. W., Day, S. H., & Cooley, V. (2006). Classifying juvenile offenders according to risk of recidivism: Predictive validity, race/ethnicity, and gender. *Criminal Justice & Behavior, 33*, 305-324.

Scott, E. S., & Steinberg, L. (2008). Adolescent development and the regulation of youth crime. *The Future of Children, 18*, 15-33.

Shepherd, S. M., Luebbers, S., & Dolan, M. (2013). Gender and ethnicity in juvenile risk assessment. *Criminal Justice and Behavior, 40*, 388-408.

Simourd, L., & Andrews, D. (1994). Correlates of delinquency: A look at gender differences. *Forum on Corrections Research, 6*, 26-31.

Singh, J. P., Desmarais, S. L., Sellers, B. G., Hylton, T., Tirotti, M., & Van Dorn, R. A. (2014). From risk assessment to risk management: Matching interventions to adolescent offenders' strengths and vulnerabilities. *Children and Youth Services Review, 47*, 1-9.

Skeem, J., Barnoski, R., Latessa, E., Robinson, D., & Tjaden, C. (2013). Youth risk assessment approaches: Lessons learned and questions raised by Baird et al.'s study (2013). In C. Baird, T. Healy, K. Johnson, A. Bogie, E. W. Dankert, & C. Scharenbroch, *A comparison of risk assessment instruments in juvenile justice* (pp. 107-121). Washington, DC: National Council on Crime and Delinquency.

Skeem, J. L., & Monahan, J. (2011). Current directions in violence risk assessment. *Current Directions in Psychological Science, 20*, 38-42.

Smith, P., Cullen, F. T., & Latessa, E. J. (2009). Can 14,737 women be wrong? A meta-analysis of the LSI-R and recidivism for female offenders. *Criminology, 8*, 183-208.

Snyder, H. N., & Sickmund, M. (1999). *Juvenile justice: A century of change*. Washington, DC: U.S. Department of Justice, Office of Justice Programs, Office of Juvenile Justice and Delinquency Prevention.

Taxman, F. S. (2016). *Handbook on Risk and Need Assessment: Theory and Practice*. New York, NY: Routledge.

Thompson, A. P., & McGrath, A. (2012). Subgroup differences and implications for contemporary risk-need assessment with juvenile offenders. *Law and Human Behavior, 36*, 345-355.

Upperton, R. A., & Thompson, A. P. (2007). Predicting juvenile offender recidivism: Risk–need assessment and juvenile justice officers. *Psychiatry, Psychology and Law, 14*, 138-146.

van der Put, C. E., Dekovic, M., Geert, J. J., Stams, M., van der Laan, P. H., Hoeve, M., & Van Amelsfort, L. V. (2011). Changes in risk factors during adolescence: Implications for risk assessment. *Criminal Justice and Behavior, 38*, 248–262.

van der Put, C. E., Dekovic, M., Hoeve, M., Stams, G. J. J., van der Laan, P. H., & Langewouters, F. E. (2014). Risk assessment of girls are there any sex differences in risk factors for re-offending and in risk profiles? *Crime & Delinquency, 60*, 1033–1056.

Vieira, T., Skilling, T., & Peterson-Badali, M. (2009). Matching court-ordered services with treatment needs: Predicting treatment success with young offenders. *Criminal Justice & Behavior, 36*, 385–401.

Viljoen, J. L., Cochrane, D. M., & Jonnson, M. R. (2018). Do risk assessment tools help manage and reduce risk of violence and reoffending? A systematic review. *Law and Human Behavior, 42*, 181-214.

Virginia Department of Juvenile Justice. (2008). *Data resource guide: Fiscal year 2008.* Richmond, VA: Author.

Virginia Department of Juvenile Justice. (2011). *Data resource guide: Fiscal year 2011.* Richmond, VA: Author.

Virginia Department of Juvenile Justice. (2013). *Agency Policies and Procedures Manual: 9324. Probation Supervision Plan.* Richmond, VA: Author.

Virginia Department of Juvenile Justice. (2016). *Data resource guide: Fiscal year 2016.* Richmond, VA: Author.

Virginia Department of Juvenile Justice. (2017). *Transformation Plan 2017 Update.* Richmond, VA: Author.

Washington State Department of Social & Health Services. (2014). *Report to the legislature: Juvenile court block grant report.* Olympia, WA: Author.

Yang, M., Wong, S. C. P., & Coid, J. (2010). The efficacy of violence prediction: A meta-analytic comparison of nine risk assessment tools. *Psychological Bulletin, 136*, 740-767.

Appendices

Appendix A. YASI Full Assessment Outline

Legal History

1. Previous intake contacts for offenses
2. Age at first intake contact
3. Intake contacts for offenses
4. Felony-level offenses
5. Weapons offenses
6. Offenses against another person
7. Felony-level offenses against another person
8. Placements
9. Juvenile detention
10. DJJ custody
11. Escapes
12. Failure-to-appear in court
13. Violations of probation/parole/diversion

Family

1. Runaways/lock-outs
2. History of child neglect
3. Compliance with parental rules
4. Circumstances of family members living at home
5. Historic problems of family members at home
6. Youth's current living arrangements
7. Parental supervision
8. Appropriate consequences

9. Appropriate rewards
10. Parental attitudes
11. Family support network
12. Family member(s) the youth feels close to
13. Family provides opportunities for participation
14. Family provides opportunities for learning success
15. Parental love, caring, and support
16. Family conflict

School

1. Current enrollment status
2. Attendance
3. Conduct in past year
4. Academic performance in past year
5. Current conduct
6. Current academic performance
7. Special education student
8. Youth believes in the value of education
9. Encouraging school environment
10. Expulsion and suspensions
11. Age at first expulsion
12. Involvement in school activities
13. Teachers/staff/coaches youth likes

Community/Peers

1. Associates the youth spends time with
2. Attachment to positively influencing peer(s)
3. Admiration/emulation of tougher delinquent peers
4. Months associating with delinquent friends/gang
5. Free time spent with delinquent peers
6. Strength of delinquent peer influence
7. Number of positive adult relationships in community
8. Pro-social community ties

Alcohol/Drugs

1. Alcohol and drug use
2. Receptive to substance use treatment
3. Previous substance use treatment

Mental Health

1. Mental health problems
2. Homicidal ideation
3. Suicidal ideation
4. Sexual aggression
5. Physical/sexual abuse
6. Victimization

Aggression

1. Violence
2. Hostile interpretation – actions/intentions of others

3. Tolerance for frustration
4. Belief in use of physical aggression to resolve a disagreement or conflict
5. Belief in use of verbal aggression to resolve a disagreement or conflict

Attitudes

1. Responsibility for delinquent/criminal behavior
2. Understanding impact of behavior on others
3. Willingness to make amends
4. Optimism
5. Attitude during delinquent/criminal acts
6. Law-abiding attitudes
7. Respect for authority figures
8. Readiness to change

Skills

1. Consequential thinking skills
2. Social perspective-taking skills
3. Problem-solving skills
4. Impulse-control skills to avoid getting in trouble
5. Loss of control over delinquent/criminal behavior
6. Interpersonal skills
7. Goal-setting skills

Employment/Free Time

1. History of employment
2. Number of times employed

3. Longest period of employment
4. Positive relationship with employers
5. Structured recreational activities
6. Unstructured recreational activities
7. Challenging/exciting hobbies/activities
8. Decline in interest in positive leisure pursuits

(DJJ, 2016, p. 86-87)

Appendix B: YASI Scores and Levels

YASI produces numerical scores and levels as described in Table 38 and Table 39.

Table 38. YASI Overall Scores and Levels

Overall	Pre-Screen	Full Assessment
Risk	X	
Static Risk		X
Dynamic Risk		X
Protective		X
Static Protective		X
Dynamic Protective		X

Table 39. YASI Domain Scores and Levels from the Full Assessment

Domain	Static Risk	Dynamic Risk	Static Protective	Dynamic Protective
Legal History	X			
Family	X	X		X
School	X	X		X
Community/Peers	X	X		X
Alcohol/Drugs	X	X		
Mental Health*				
Aggression	X	X		X
Attitudes	X	X	X	X
Skills		X		X
Employment/Free Time	X	X	X	X

Note: The *Mental Health* domain results in a flag to indicate a need for further assessment and/or intervention without indicating an increased risk.

Appendix C. Summary Report to DJJ: Validation Study of YASI

Purpose: The Virginia Department of Juvenile Justice (DJJ) utilizes the Youth Assessment and Screening Instrument (YASI); however, risk assessment instruments do not always generalize across populations. This study focused on the accuracy in predicting recidivism, equity across racial groups, and usage of YASI as a case planning tool in the state of Virginia.

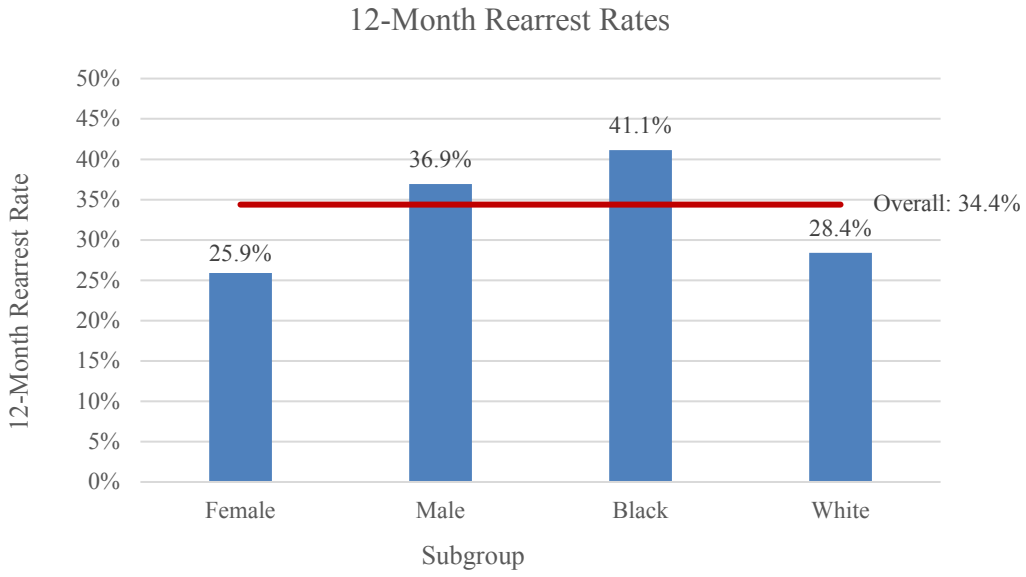
Summary of Findings: Of 11,888 youth on probation or parole, 34.4% were rearrested within 12 months. A higher percentage of males (36.9%) were rearrested than females (25.9%), and a higher percentage of Black youth (41.1%) were rearrested than White youth (34.4%). (See Attachment A.) Females (44.7%) were more likely to be low risk than males (23.2%), and Black youth (23.6%) were more likely to be high risk than White youth (18.9%). (See Attachment B.)

- YASI in Virginia performed as expected in comparison to existing research on youth risk assessment instruments. (See Attachment C for risk scores and levels by rearrest rates.)
- The predictive validity of the overall and dynamic risk scores and levels was statistically equivalent for males and females. The *Community/Peers* and *Family* domains had stronger predictive validity for males than females.
- The predictive validity was statistically equivalent for White and Black youth for overall risk levels and dynamic risk scores and levels; however, the predictive validity for the overall risk score was statistically significantly higher for White youth than Black youth. The *Alcohol/Drugs* domain had stronger predictive validity for White youth than Black youth.
- Each domain had a positive correlation between risk and priority area assignment, though some were stronger correlations than others. The strongest correlations were for the *Alcohol/Drugs* domain, followed by *Aggression*. The *Employment/Free Time* domain had the lowest prevalence of moderate to high risk, lowest occurrence of assigned priority area, and the lowest correlation between domain level and priority area assignment.

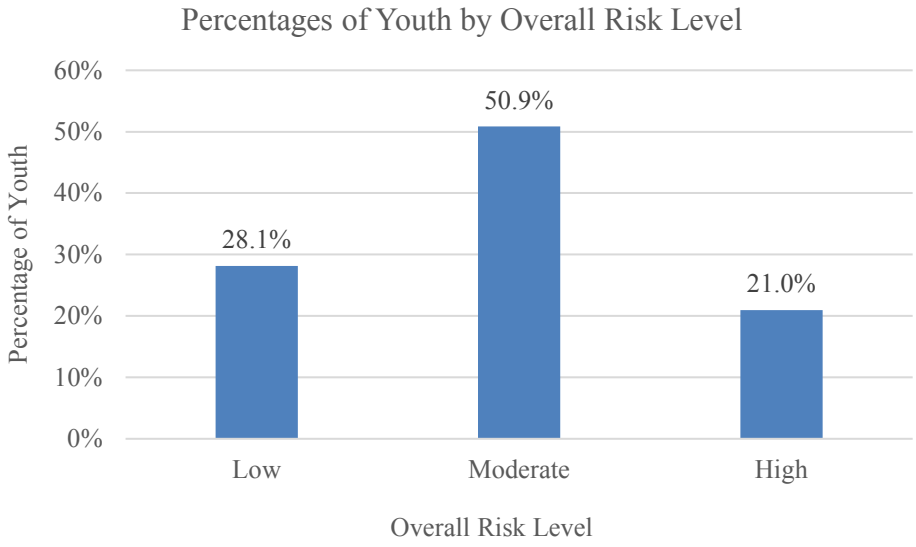
Research Recommendations:
- Instrumental validity: Do individual items load onto their designated domains? Are the domains separate and distinct concepts?
- Inter-rater reliability: Have training efforts maintained consistency in assessments?
- Interactions between domains: Do patterns exist in how domains present?
- Different outcomes: What is the predictive ability for different reoffense severities?
- Additional groups: Are there other group or geographical differences?
- Weighting and scoring: Can modifications improve the predictive ability?
- Service matching: Is the tool used by practitioners to match services to the risk profile?
- Recidivism reduction: Does the use of risk assessment tools decrease reoffending?
- Program evaluations: Do services improve dynamic risk?

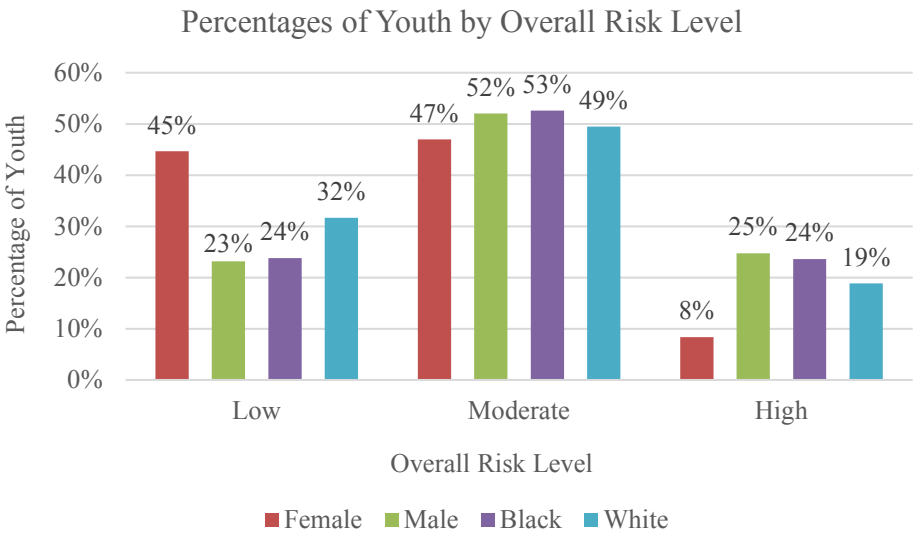
Ongoing Policy Recommendations:
- Determine purpose and function of the assessment within the system
- Conduct staff and stakeholder training
- Test the tool's performance
- Calibrate and modify the instrument
- Repeat

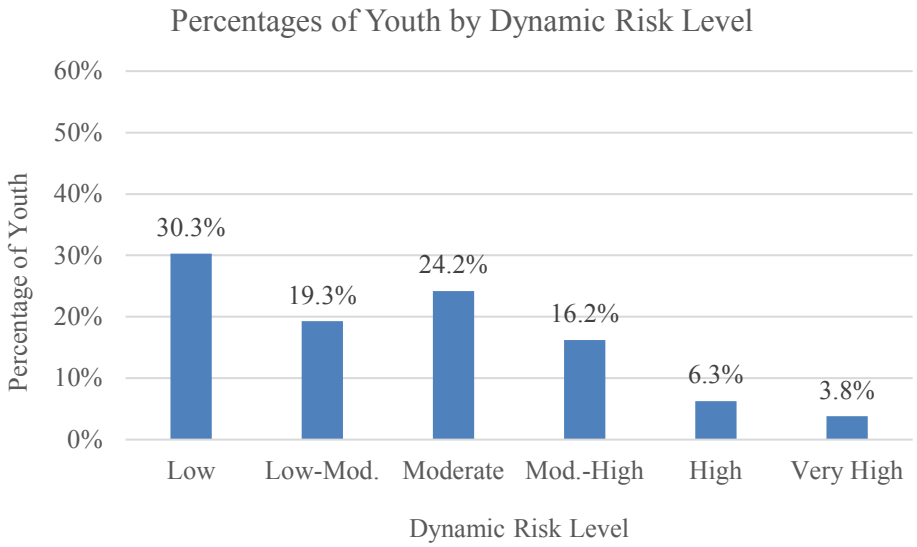
Attachment A: Recidivism

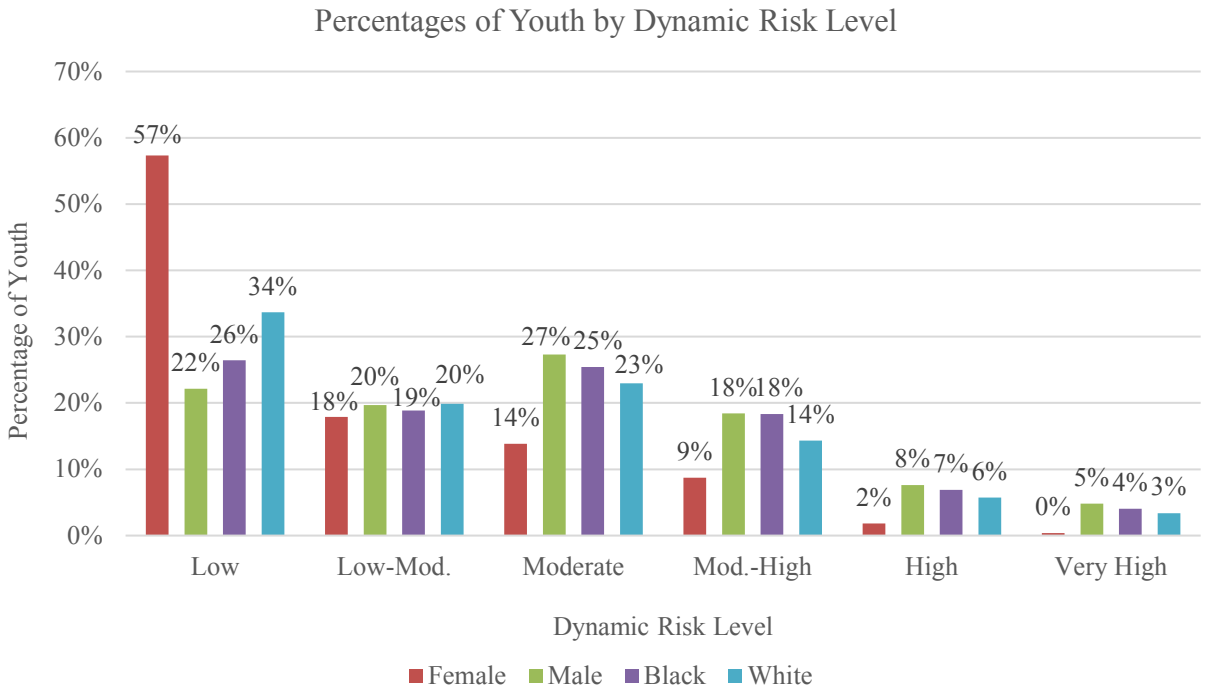


Attachment B: Risk Assessment Distributions






Percentages of Youth by Dynamic Risk Level
60%
50%
40%
30%
20%
10%
0%
Percentage of Youth
30.3%
19.3%
24.2%
16.2%
6.3%
3.8%
Low
Low-Mod.
Moderate
Mod.-High
High
Very High
Dynamic Risk Level

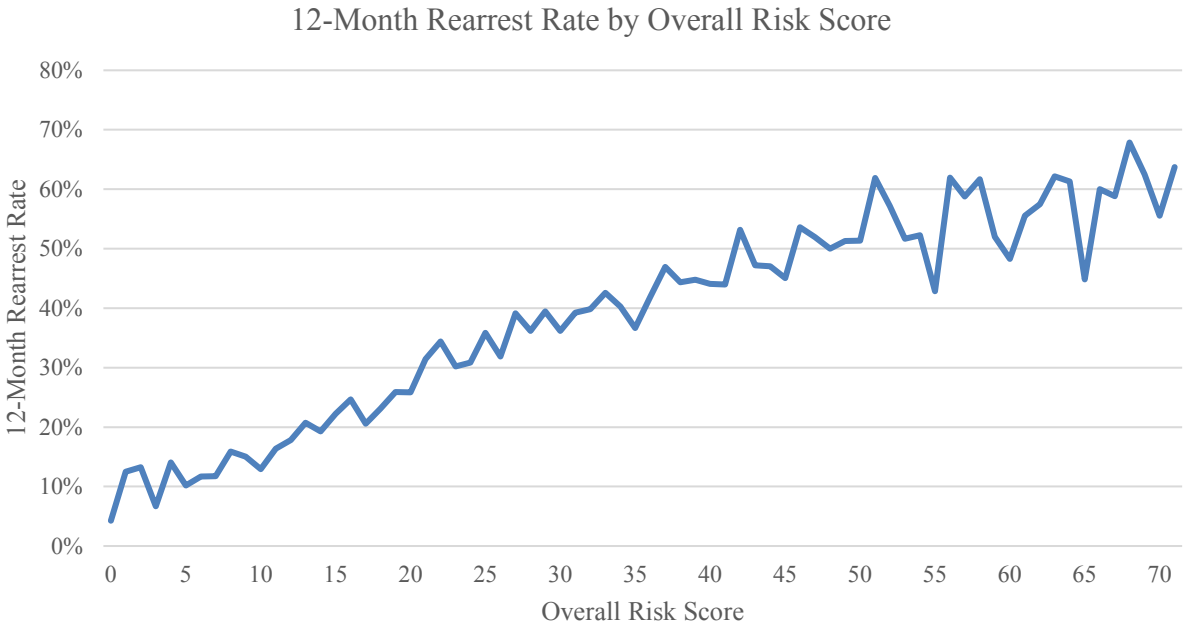

Percentages of Youth by Dynamic Risk Level
70%
60%
50%
40%
30%
20%
10%
0%
Percentage of Youth
57%
22%
26%
34%
18%
20%
19%
20%
14%
27%
25%
23%
9%
18%
18%
14%
2%
8%
7%
6%
0%
5%
4%
3%
Low
Low-Mod.
Moderate
Mod.-High
High
Very High
Dynamic Risk Level
Female
Male
Black
White

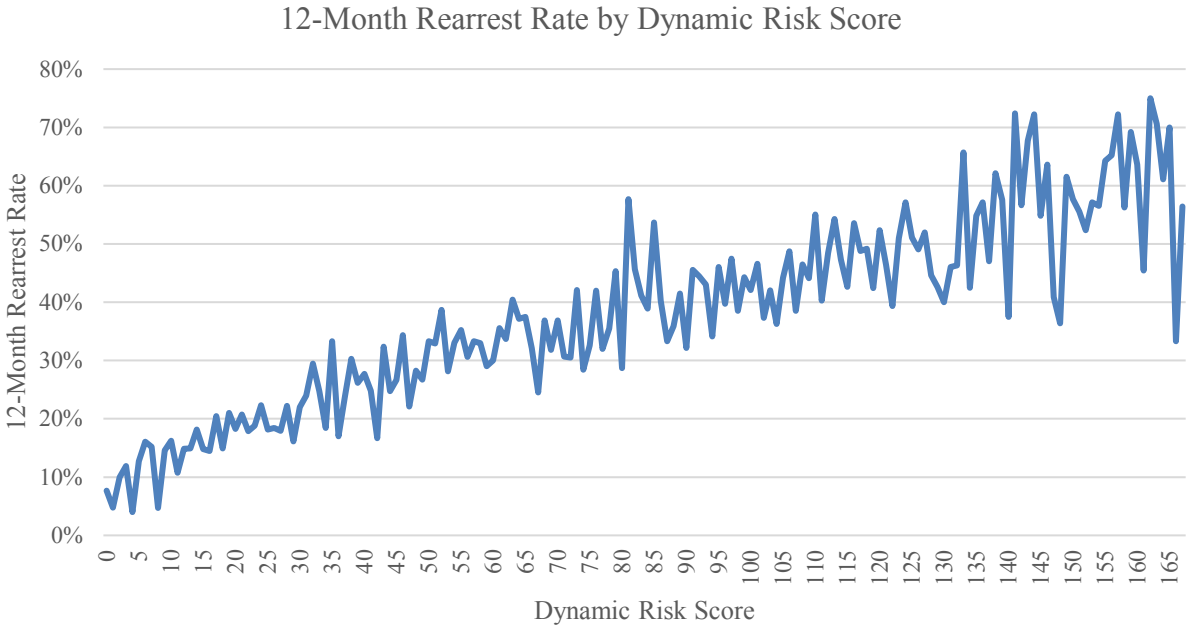
Summary of Statistically Significant Assessment Differences by Subgroup

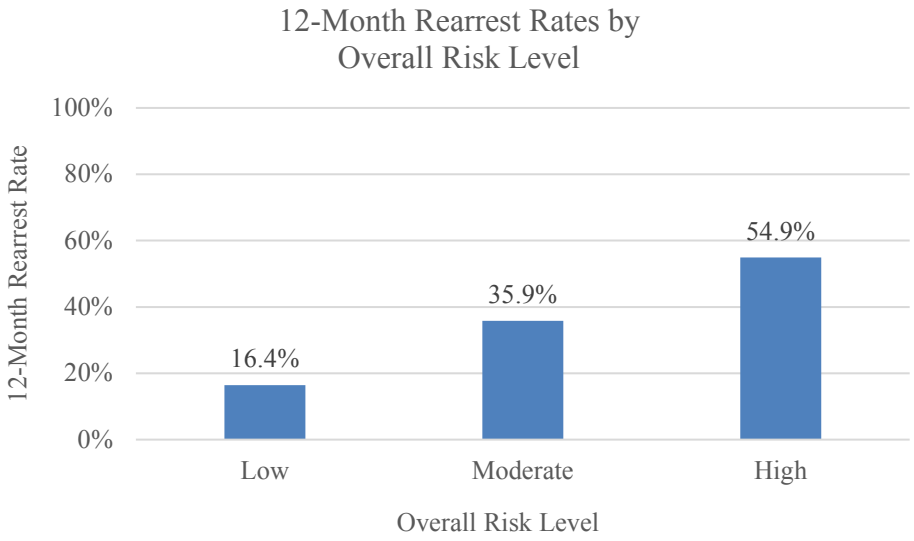
	Sex		Race	
	Score	Level	Score	Level
Overall	--	Male	Black	Black
Dynamic	--	Male	Black	Black
Domains				
Legal History	Male	Male	Black	Black
Family	Female	Male	White	--
School	Female	Male	Black	Black
Community/Peers	Male	Male	Black	Black
Alcohol/Drugs	Male	Male	White	White
Aggression	Female	Male	Black	Black
Attitudes	--	Male	Black	Black
Skills	--	Male	Black	Black
Employment/Free Time	Female	--	Black	Black

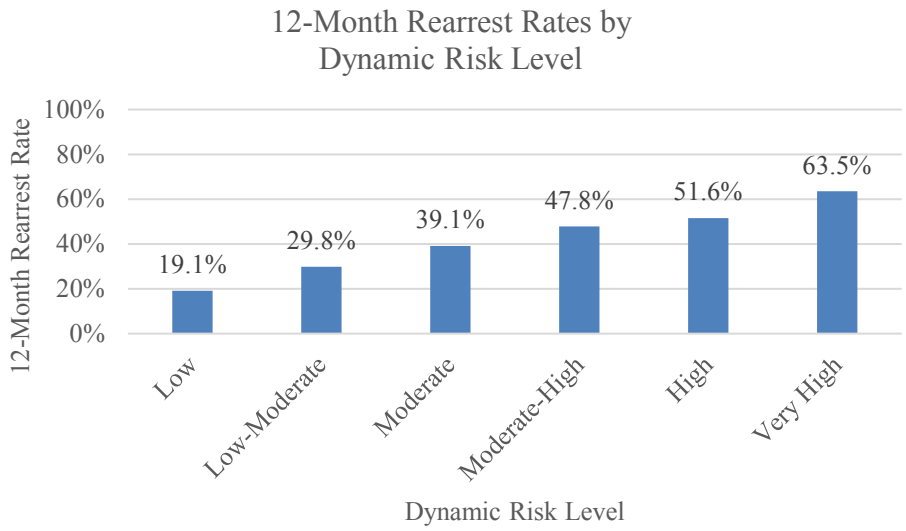
Note: The subgroup with the higher value is listed.

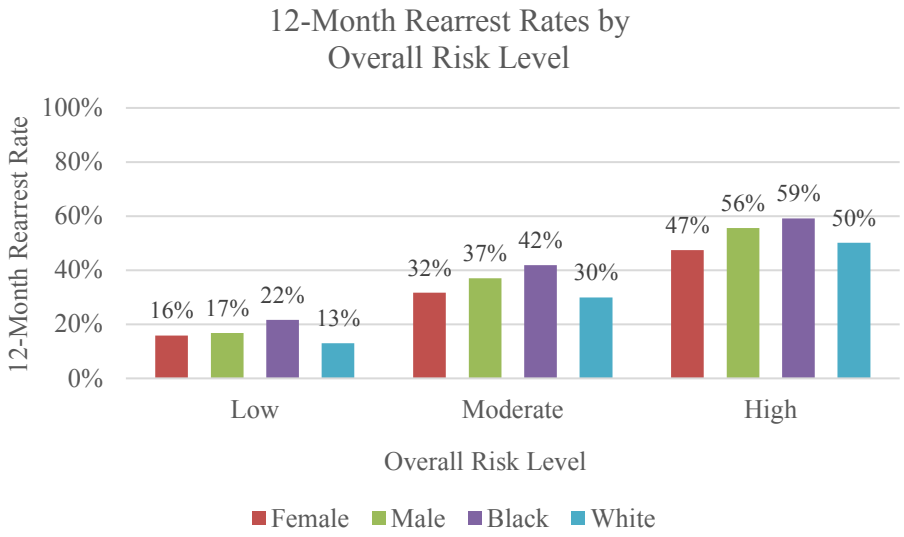
Attachment C: Risk Assessment and Recidivism

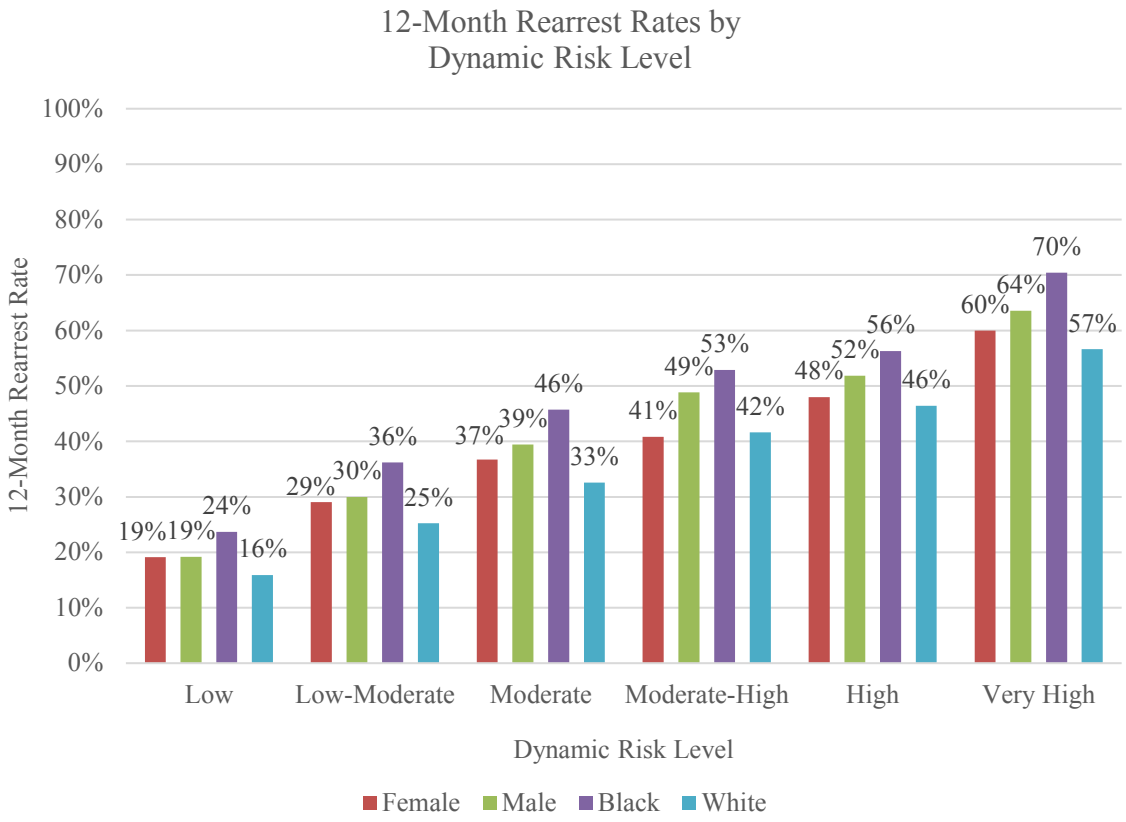





12-Month Rearrest Rates by
Overall Risk Level
12-Month Rearrest Rate
100%
80%
60%
40%
20%
0%
16.4%
35.9%
54.9%
Low
Moderate
High
Overall Risk Level


12-Month Rearrest Rates by
Dynamic Risk Level
12-Month Rearrest Rate
100%
80%
60%
40%
20%
0%
19.1%
29.8%
39.1%
47.8%
51.6%
63.5%
Low
Low-Moderate
Moderate
Moderate-High
High
Very High
Dynamic Risk Level


12-Month Rearrest Rates by
Overall Risk Level
12-Month Rearrest Rate
100%
80%
60%
40%
20%
0%
16%
17%
22%
13%
32%
37%
42%
30%
47%
56%
59%
50%
Low
Moderate
High
Overall Risk Level
Female
Male
Black
White


12-Month Rearrest Rates by
Dynamic Risk Level
12-Month Rearrest Rate
100%
90%
80%
70%
60%
50%
40%
30%
20%
10%
0%
19%
19%
24%
16%
29%
30%
36%
25%
37%
39%
46%
33%
41%
49%
53%
42%
48%
52%
56%
46%
60%
64%
70%
57%
Low
Low-Moderate
Moderate
Moderate-High
High
Very High
Dynamic Risk Level
Female
Male
Black
White

YASI Predictive Validity Results of Current Study Compared to Previous Studies

	Orbis, 2007	Jones, 2011	Jones et al., 2016	Baird et al., 2013	Current - Scores	Current - Levels
Pre-Screen	0.65	0.64	0.79	0.68	0.68	0.66
Female	0.61	0.60	0.68	0.67	0.67	0.63
Male	0.68	0.64	0.82	0.71	0.68	0.65
Black				0.66	0.65	0.64
White				0.68	0.69	0.67
Dynamic Risk	0.62	0.63			0.65	0.66
Female	0.59	0.62			0.64	0.62
Male	0.64	0.63			0.66	0.65
Domain Dynamic Risk	0.55-0.63		0.54-0.73		0.56-0.64	0.56-0.64
Female		0.50-0.60			0.55-0.63	0.55-0.61
Male		0.50-0.62			0.57-0.65	0.57-0.64

Note: Values represent the Area Under the Receiver Operating Characteristic Curve (AUC), a measure of predictive validity. Studies varied in population, follow-up timeframes, and recidivism measures and may have reported additional AUCs not displayed in the summary table above. Orbis (2007) and Jones (2011) both studied New York populations, resulting in similar AUCs. In the Orbis (2007) study, "negative outcome" was defined as a new referral/arrest, violation of probation, or adjudication/conviction; AUCs for the pre-screen after item weight and cut-off point revisions are displayed.

Summary of Statistically Significant AUC Differences by Subgroup

	Sex		Race	
	Score	Level	Score	Level
Overall	--	--	White	--
Dynamic	--	--	--	--
<u>Domains</u>				
Legal History	--	--	--	--
Family	--	Male	--	--
School	--	--	--	--
Community/Peers	Male	Male	--	--
Alcohol/Drugs	--	--	White	White
Aggression	--	--	--	--
Attitudes	--	--	--	--
Skills	--	--	--	--
Employment/Free Time	--	--	--	--

Note: The subgroup with the higher AUC value is listed.