IMPROVING JUVENILE RISK ASSESSMENT MEASUREMENT MODELS:
A PSYCHOMETRIC COMPARISON OF SCORING METHODS

By

Mary Katherine Kitzmiller

A DISSERTATION

Submitted to
Michigan State University
in partial fulfillment of the requirements
for the degree of

Psychology – Doctor of Philosophy

2022

ABSTRACT

IMPROVING JUVENILE RISK ASSESSMENT MEASUREMENT MODELS:
A PSYCHOMETRIC COMPARISON OF SCORING METHODS

By

Mary Katherine Kitzmiller

Juvenile risk assessments are standardized rating instruments that measure criminogenic risk in court-involved youth. Juvenile court practitioners use scores from risk assessments to inform judicial decisions throughout case processing. It is critically important that risk scores accurately reflect court-involved youths' latent level of criminogenic risk; both artificially high and low scores incur significant detriments to youths, courts, and communities. In light of the consequences of risk misevaluation, there is urgent need to develop and evaluate alternate juvenile risk assessment measurement models

The current study aspired to improve measurement of criminogenic risk through the development of a Novel Scoring Algorithm which innovated upon current juvenile risk assessment scoring twofold: (1) it adjusted the weights of assessment items and domain sub-scores to reflect their correlation with latent constructs of criminogenic risk; and (2) it integrated the mitigating impact of prosocial protective factors into cumulative risk scores. Drawing upon a sample of 559 youth who entered the supervision of a county-level juvenile circuit court for the first time, the Novel Scoring Algorithm outperformed the current method of scoring (i.e., summing all unweighted risk factors) in both absolute and relative model fit. However, the Novel Scoring Algorithm yielded no incremental improvement in diagnostic accuracy, affirming the Scoring-as-Usual method as an acceptable procedure for assessing likelihood of recidivism in court-involved youth. Implications for effectively and equitably managing risk are discussed.

ACKNOWLEDGMENTS

It takes a village to write a dissertation, and I want to recognize my village for their contributions to this work. First, thank you to my advisor and co-Chair, Dr. Caitlin Cavanagh. I have grown so much as a researcher and a writer because of your thoughtful and compassionate mentorship. You are my favorite editor, co-author, and giver-of-advice, and I know that our work together will not end here. I also want to thank my co-Chair, Dr. Rebecca Campbell, for your persistent advocacy on my behalf over the last five years. To my committee members, Drs. Julie Krupa and Cris Sullivan, thank you for believing in this project from start to finish and for your thoughtful feedback along the way. It has been a joy to learn from all of you.

Special thanks to all of the past and current student members of the Juvenile Risk Assessment Team, who have worked tirelessly over the last 18 years to produce the high-quality data that I had the privilege of using in this dissertation. Thank you to our juvenile court collaborators, especially Scott, for many thought-provoking conversations on risk assessment over the years. Your relentless commitment to upholding best practice is laudable and humbling.

To my Eco-Community family, thank you for supporting and challenging me to get this right. Jen, Isi, and Funmi, your friendship has gotten me through the most difficult parts of this process. The time spent with you all – the Saturday morning long runs, the Monday night reality TV watch parties, and the midday trips to Sparty's – are the parts of this journey that I cherish the most. Thank you.

To my parents, thank you for nourishing my love of learning and encouraging me to forge my own path. Dad, your genuine interest and investment in my research means the world to me. Mom, I am forever changed by your love, support, and encouragement. I wish you were here for this moment.

To my partner, Jacob. Time and time again, you have grounded me with patience, perspective, and love. I am so incredibly grateful that you were willing to embark on this Michigan odyssey together. I know we will look back on this season of life with fondness.

Onward.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

INTRODUCTION

In 2019, an estimated 690,000 minors were arrested in the United States for the first time (Office of Juvenile Justice & Delinquency Prevention [OJJDP], 2019). Adolescents are uniquely primed to engage in law breaking by virtue of their psychosocial characteristics (e.g., impulsivity, susceptibility to peer pressure) (Steinberg et al., 2015). However, systems of oppression have reinforced disparate outcomes at every stage in the juvenile case processing, including arrest, conviction, and detention (Birckhead, 2012; Piquero, 2008). A substantial body of literature has attributed these outcomes to "differential selection": the juvenile justice system upholds systems of oppression by imposing more punitive forms of court supervision on marginalized youths (Piquero, 2008). One of the most documented mechanisms for differential selection lies in discretion-based methods of risk evaluation, wherein case processing decisions reflect youths' perceived threat of future harm and receptibility to available services (Mulvey & Iselin, 2008)

In order to promote fair and equitable justice administration for court-involved youths, 46 states have instituted statutes that support or require juvenile risk assessment utilization (Juvenile Justice Geography, Policy, Practice & Statistics [JJGPS], 2020). Juvenile risk assessments are standardized instruments that estimate the likelihood of recidivism based upon empirically validated criteria. Assessment items reflect criminogenically-linked characteristics of the individual youths (e.g., personality, attitudes) and their proximal social environment (e.g., school, family, community) (Andrews & Bonta, 2010). Subsequently generated risk scores, which correspond to the unweighted sum of all risk factors identified, can inform several important judiciary decisions, including type and duration of court supervision (Vincent et al., 2012).

Juvenile risk assessment utilization is considered favorable over discretion-based methods of risk evaluation for a number of reasons: (1) risk assessments are more accurate in predicting general delinquent recidivism (Bonta & Andrews, 2007; Oleson et al., 2011); (2) they ensure youth are evaluated a consistent set of empirically supported criteria (Peck & Jennings, 2016; St. John et al., 2020); and (3) they are often administered with a separate, but complementary, protective factors assessment, which facilitates wholistic case planning (Vincent et al., 2012). Court jurisdictions that utilize juvenile risk assessments witness lower rates of recidivism and higher rates of treatment compliance, signaling the importance of these tools in facilitating effective case management (Schwalbe, 2007).

Despite these advantages, advocates of justice system reform have raised concern that juvenile risk assessments sustain, rather than prevent, discriminatory judicial decisions (Green, 2020). Risk scores reflect population-level inequities, legitimizing inappropriately punitive, and ultimately harmful, justice system sanctions directed towards marginalized youths (Harcourt, 2010; Miron et al., 2021). Furthermore, juvenile risk assessments were developed and calibrated using predominantly male delinquent samples, which calls into question their appropriateness for measuring criminogenic risk among status offenders and girls (Onifade et al., 2009; Van Voorhis et al., 2010). While it is unlikely that risk assessment will eradicate structural oppression upheld through the juvenile justice system, courts can reduce related harm by ensuring that risk scores accurately and consistently reflect youths' latent trait of criminogenic risk.

Inaccurate, inconsistent risk assessment scores directly inhibit effective service delivery. Processing decisions based on artificially low risk scores enable youth to re-enter their communities with unaddressed criminogenic needs, placing them at higher likelihood of reoffending (McCarter, 2016). On the other hand, processing decisions based on artificially high

scores justifies the prescription of inappropriately restrictive or intensive services. In addition to misallocating court resources, these inappropriate services may harm youth by damaging their self-perception, disrupting their at-home routines, and increasing their association with higher risk peers (Cecile & Born, 2009; Gatti et al., 2009; Leve & Chamberlain, 2005). Current methods of juvenile risk assessment scoring may contribute to systematic misevaluation of court-involved youth. Therefore, there is urgent need to develop more precise, strengths-based, and ecologically informed juvenile risk assessment measurement models.

The overall objective of this dissertation was to develop and evaluate a novel juvenile risk assessment scoring algorithm (hereinafter "Novel Scoring Algorithm") with the intention of improving the current method of measuring criminogenic risk (hereinafter "Scoring-as-Usual Method"). Using second-order Confirmatory Factor Analysis (CFA), analyses drew upon risk assessment records from 559 court-involved youths who had been formally petitioned to juvenile court for the first time. Results have immediate implications towards accurately and equitably measuring criminogenic risk in youth via juvenile risk assessment.

LITERATURE REVIEW

Developmental & Ecological Perspectives on Juvenile Delinquency

Juvenile delinquency is common during adolescence; an estimated 1,909 minors are arrested each day in the United States (OJJDP, 2019). Key social and physiological characteristics prime adolescents to engage in law breaking in ways that distinguish them from adults (Littlefield et al., 2010; Kuhn, 2009, Cauffman & Steinberg, 2000). Adolescents are entrusted with more responsibilities and freedoms than they once were as children; however, they lack psychosocial maturity, rendering them unable to regulate strong emotions, foresee future consequences, and resist peer pressure (Cauffman et al., 2016; O'Brien et al., 2011; Sebastian et al., 2010). As a result, adolescents are drawn to high-risk behaviors, which in many cases includes delinquency. Fortunately, most teens' law breaking is contained within adolescence, even among those who commit serious crimes; as a result, juvenile delinquency has been termed both temporary and situational (Moffitt, 1993).

While all young people experience roughly the same psychosocial changes during adolescence, contact and interactions with the juvenile justice system vary notably by race/ethnicity, socioeconomic status, and gender. Rates of disproportionate minority contact (DMC) have been reported at every stage of justice system involvement, with racial disparities widening as youths advance through the stages of court processing (i.e., arrest, formal processing, conviction, incarceration) (Piquero, 2008; Zane & Pupo, 2021). Similarly, while socioeconomic indicators are not reported nationally, some jurisdictions indicate that as many as 60% of youths under court supervision live below the poverty line (Birckhead, 2012). While girls are underrepresented within the general delinquent population, the juvenile justice system has historically failed to respond to the unique social context which primes them towards

4

delinquency, further inhibiting successful rehabilitation (Hubbard & Matthews, 2008). These disparate experiences exemplify how the juvenile justice system is reflective of deeply entrenched social, cultural, historical inequities; therefore, solely conceptualizing them within the framework of individual development is limiting.

Community psychologists have advocated for ecologically informed models of understanding, addressing, and preventing juvenile delinquency (Fountain & Mahmoudi, 2021; Javdani & Allen, 2016; Roesch, 1988). Ecological inquiry here refers to an umbrella of multidisciplinary theories and concepts which describe human behavior as the product of reciprocal interactions between an individual and their environment (McBride & McCoy, 1981). Person-environment interactions are often identified and studied within different contextual systems, including the immediate social environment (e.g., family, school, peers) as well as broader social systems (e.g., community, society, culture) (Bronfenbrenner, 1977; Kumpfer & Turner, 1990). A well-substantiated body of research has documented how ecological contexts can increase and reduce the likelihood of deviant behavior in adolescence, including substance use (Hawkins et al., 2004), violence (Gorman-Smith et al., 1996; Tarter et al., 2002), and delinquency (Moon et al., 2010; Windle, 2000).

The disparate contact and treatment of youth in juvenile justice system can be better understood from an ecological perspective. Although scholars have used different terms to describe structural forms of oppression (e.g., systemic, institutional), these terms center the idea that white supremacist, capitalistic, and patriarchal values are codified into our society's policies, laws, practices, structures, and institutions (Homan, 2019; Rucker & Richeson, 2021). Structural forms of oppression interact to produce differential access to power and essential resources, resulting in differential access to high-quality education, safe housing, employment

5

opportunities, healthcare, and wealth (Bailey et al., 2017; Bonilla-Silva, 1997; Jones, 2000; Powell, 2007). Within the context of juvenile delinquency, structural oppression can be understood as both an on-ramp to law breaking (i.e., differential involvement in crime) as well as a contributing factor to disparate justice system outcomes (i.e., differential selection in the justice system).

Differential Involvement in Juvenile Delinquency. Structural oppression exercised through federal subsidies, predatory mortgage lending restrictions, and subsidized housing locations has created and maintained racially segregated neighborhoods marked by "disinvestment and concentrated poverty" (Williams & Mohammed, 2013; Powell, 2007). Youths who reside within these neighborhoods often affected by chronic unemployment, inadequate living conditions, and under resourced schools. In the absence of legally viable pathways to achieve upward social and economic mobility, these youths may engage in law breaking as a means of survival and financial security (Nunn, 2001). Furthermore, racial profiling, increased neighborhood surveillance, and other heavy-handed police tactics render racially and socioeconomically marginalized youths acutely vulnerable to contact with law enforcement resulting in arrest (Feinstein, 2015).

Differential Selection in the Juvenile Justice System. Structural oppression is also enacted through the operations of the juvenile court system. The juvenile court was developed under the legal doctrine of *parens patriae* (the State as Parent): unlike the criminal justice system, actions of the juvenile court are intended rehabilitate youth from deviant behavior and facilitate prosocial transitions into adulthood (Bilchik, 1998; Center on Juvenile and Criminal Justice [CJCJ], 2021). In practice, *parens patriae* has allowed juvenile court actors near-unchecked levels of discretion in court processing, viewed originally as a favorable relaxation of

the formal procedures carried out by the criminal justice system (Stohs, 2003). Under discretion-based methods of risk evaluation, case processing decisions reflect youths' perceived threat of future harm and receptibility to available services (Mulvey & Iselin, 2008). However, the subsequent lack of procedural safeguards has often come at a detriment to court-involved youths, as their experience and outcomes can vary widely by virtue of the legal actors they encounter. While the consequences of discretion-based biases in the juvenile justice system are both complex and intersectional, an abundance of research has documented its distinct harms to both youth of color and girls:

Empirical research highlights the ways in which racial discrimination, particularly anti-Black racism, is carried out though discretion-based methods of risk evaluation. Bridges and Steen (1998) found that probation officers describe Black and White youth differently in their unstructured evaluations of criminogenic risk: narratives of Black youth were more likely to reference negative personality traits (e.g., unremorseful), while narratives of White youth were more likely to include descriptions of negative environmental influences (e.g., peer pressure). Furthermore, when decisions are guided by legal actors' discretion, youths of color are more likely to be placed in pretrial detention (Bishop & Frazier, 1995; Johnson & Secret, 1990; Wordes et al., 1994), formally petitioned to juvenile court (Bortner & Wornie, 1985; DeJong & Jackson, 1998), and receive more punitive sanctions (McGarrell, 1993; Thomas & Sieverdes, 1975), after controlling for the severity of their offense.

Additionally, the decisions of legal actors have often upheld traditional patriarchal ideologies at the expense of court-involved girls (Chesney-Lind, 1977). The term *judicial paternalism* describes how the actions of the juvenile court both protect and punish young women for violating gendered behavioral expectations (Chesney-Lind, 1977; Spivak et al.,

7

2014). The direction of judicial paternalism's influence on court outcomes is a subject of debate among scholars: some have found evidence that legal actors apply a chivalrous bias towards girls in the justice system, granting them leniency from the otherwise imposed penalties of delinquency (Blackwell et al., 2008; Daly, 1994). Others have asserted that girls under juvenile court jurisdiction are doubly punished: first for the actual offense, and again for violating patriarchal expectations for appropriate feminine behavior (Crew, 1991; Erez, 1992; Spohn, 1999). Regardless, the influence of judicial paternalism on risk evaluation inhibits fair, equitable, and effective case processing for court-involved girls.

Inappropriate case processing decisions can disrupt normative adolescent development, inhibit prosocial transitions to adulthood, and predispose youths to persistent criminal trajectories (Liberman & Fontaine, 2015). In sum, steering justice administration by discretion of legal actors institutionalizes racism, paternalism, and other forms of oppression, at significant costs to court-involved youths (Liberman & Fontaine, 2015).

Development of Juvenile Risk Assessment

Actuarial juvenile risk assessments were developed to remedy the harm incurred by discretion-based methods of risk evaluation (St. John et al., 2020). Beginning in the 1990s, several studies systematically investigated the characteristics that distinguished adolescents who follow persistent criminal trajectories from their peers (Hoge et al., 1996; Howell & Hawkins, 1998). Drawing upon these findings, experts developed several semi-structured interview paradigms, checklists, and rating tools for the purpose of identifying these characteristics in court-involved youths (Hoge & Andrews, 2010).

Risk-Needs-Responsivity Model. The fundamental theory of change guiding juvenile risk assessments is the Risk-Needs-Responsibility (RNR) Model, a widely adopted approach to

corrections in the criminal justice system that has been generalized to juvenile court settings (Bonta & Andrews, 2007). The RNR steers court responses to delinquency using three principles: (1) the *risk principle*, which states that the level of restriction in court-sanctioned services must match the youth's cumulative level of criminogenic risk; (2) the *needs principle*, which states that the types of services must match the youth's unique profile of criminogenic needs; and (3) the *responsivity principle*, which states that the court must maximize the likelihood of personal growth by adapting services to relevant individual and community characteristics (Bonta & Andrews, 2007).

Juvenile risk assessments play a key role in translating these principles into practice. First, juvenile risk assessments yield cumulative risk scores, which correspond to the number of risk factors identified. In accordance with the risk principle, more intensive services should be reserved for youths with elevated risk scores, while those with lower scores should be eligible for less involved sanctions (e.g., community service) or dismissed from court supervision altogether (Andrews et al., 1990). Next, juvenile risk assessments parse out criminogenic risk in distinct domains, including prior involvement in the justice system, family and peer relations, school conduct, and leisure time management. The needs principle states that courts should provide a menu of services and strategies that address these distinct areas of need; a "one size fits all" approach to intervention will yield limited success (Bonta & Andrews, 2007). Finally, many juvenile risk assessments identify individual- and community-level strengths that may deter youths from future offending. The responsivity principle states that courts can maximize the likelihood of treatment success by incorporating identified strengths into case planning (Hoge et al., 1996).

Innovations in Juvenile Risk Assessment

Overtime, many juvenile risk assessment instruments have evolved to reflect a more ecologically informed and strengths-based correctional framework (Barnes-Lee, 2020; Jacobs et al., 2020). Both innovations represent significant gains in the evaluation and treatment of court-involved youths, over discretion-based methods of risk evaluation.

Ecologically Informed Measurement. Standardized juvenile risk assessments measure relevant criminogenic characteristics at both the individual (e.g., antisocial attitudes, emotional regulation) and microsystemic (e.g., family, school, peer group) level (Bronfenbrenner, 1979). Identifying areas of need within youths' immediate social environment enables court practitioners to involve other relevant actors and settings in rehabilitation efforts (Singh & Azman, 2020). For example, family members, school personnel, and coaches may play key roles in youths' treatment success by monitoring changes in behavior and participating in therapeutic interventions (Singh & Azman, 2020).

Integration of Strengths-Based Assessment. Many juvenile risk assessments are paired with a separate, but complementary, assessment of criminogenic protective factors (Hoge et al., 1996). Identifying protective factors represents a critical step towards integrating strengths-based assessment into judicial decision-making (Nissen, 2006). Strengths-based assessments measure positive qualities, capacities, and resources that could play meaningful roles in youths' desistance from crime (Nissen, 2006). Strengths-based assessments are not designed to replace conventional "deficit-based" risk assessments; rather, they help practitioners identify and treat problem behaviors from a more wholistic and humanistic perspective (Graybeal, 2001; Nissen, 2006). In practice, juvenile case managers may use protective factors to enhance treatment

responsivity by referring youths into programs that showcase their talents, goals, interests, and abilities (Barnes-Lee, 2020; de Vogel et al., 2011; Ward & Brown, 2004).

Advantages of Juvenile Risk Assessment Utilization

Since their inception, juvenile risk assessments have become an increasingly integral component of risk evaluation across the United States: as of 2020, 46 states have implemented statutes which support or require their utilization (JJGPS, 2020). Their popularity can be attributed to several favorable outcomes over discretion-based methods of evaluation:

Accurate Assessment. Estimates from juvenile risk assessments have proven more accurate than discretion-based methods in distinguishing chronic juvenile offenders from their peers (Grove & Meehl, 1996; Harris, 2006; Oleson et al., 2011). Improved accuracy in risk evaluations allows courts to operate in alignment with the RNR's risk principle: time- and cost-intensive resources are allocated only to the youth who exhibit the greatest cumulative risk (Bonta & Andrews, 2007). Accordingly, meta-analytic evidence indicates that overtime, juvenile risk assessment utilization lowers courts' reliance on incarceration without compromising public safety (Viljoen et al., 2019).

Consistent Appraisal. Juvenile risk assessments are composed of clearly operationalized risk factors; when implemented correctly, youths' risk score should be invariant of the legal actor administering the assessment (Peck & Jennings, 2016). This represents significant improvement over discretion-based methods of risk evaluation, wherein case processing decisions can vary based upon the mental heuristics, political agendas, and personal biases of the evaluator (Oleson et al., 2011).

Improved Service Delivery. Juvenile risk assessments facilitate wholistic and humanistic case management by identifying a host of criminogenic risks and protective factors at

multiple ecological levels (Bonta & Andrews, 2007). Case managers can maximize the potential

for treatment success by matching youths to services that address areas of need and leverage

existing talents, community resources, and capacities. As a result, jurisdictions that utilize

juvenile risk assessments witness higher rates of treatment compliance and lower rates of

recidivism (Schwalbe, 2007).

Challenges to Risk Assessment Utilization

      While juvenile risk assessments have improved court outcomes in several regards, some

scholars have raised concern that risk assessment scores serve as proxy indicators for racism,

trauma, poverty, and other consequences of structural oppression (Harcourt, 2010; Miron et al.,

2021). As previously noted, structural oppression predisposes youth to differential delinquent

involvement; consequently, many characteristics of structural oppression are measured, either

directly or indirectly, as predictors of reoffending via juvenile risk assessment (Vincent &

Viljoen, 2020). Accordingly, a substantial body of literature has investigated the ways in which

juvenile risk assessments may sustain, rather than prevent, discriminatory judicial decision-

making, with explicit attention to consequences for youth of color and girls:

      Racism in Risk Assessment. Speaking specifically to the relationship between racism

and juvenile risk assessment, Brown (2007) writes: "Each category [of criminogenic risk] builds

a bias towards youth of color by neglecting how urban geographies affect these standardized

measures". For example, juvenile risk assessments flag defiance towards legal authority and pro-

criminal attitudes as criminogenic risk factors (Hoge, 2020). However, these attitudes and beliefs

may be logical responses for youth of color, especially Black youth, whose communities have

been generationally harmed by mass incarceration, racial profiling, and police sanctioned murder

(Glover, 2008; Outland, 2021; Tucker, 2014). By virtue of their positionality, youth of color may

be at systematically classified as high risk, regardless of their actual likelihood of reoffending. In this regard, juvenile risk assessments may do little more than discretion-based method of risk evaluation to prevent racially disparate sentencing decisions.

Gender Bias in Risk Assessment. Most widely utilized juvenile risk assessment instruments were developed and calibrated using predominantly male samples, which may come at a significant detriment to court-involved girls (Belisle & Salisbury, 2021). Feminist criminologists have identified distinct gendered factors and life experiences which prime girls towards and away from delinquency. Specifically, court-involved girls are more frequently exposed to trauma and victimization, which can be tied to their offense directly (e.g., running away from an abusive home) or indirectly (e.g., substance use as a coping mechanism for post-traumatic stress disorder) (Kerig & Becker, 2012). These experiences also hold relevance when measuring criminogenic risk, as related features of trauma may be flagged as familial dysfunction (e.g., poor relations with mother/father) or emotion dysregulation (i.e., short attention span) (Van Voohis et al., 2010). Importantly, girls' delinquency is categorically less chronic, violent, and severe when compared to boys', highlighting how juvenile risk assessment scores may conflate actual likelihood of reoffending with non-criminogenic trauma (Holtfreter & Morash, 2003). By failing to account for gendered socialization processes and life experiences, juvenile risk assessments may justify overly restrictive sentencing decisions directed towards girls.

Juvenile Risk Assessment & Systematic Misevaluation

These critiques bring to light how certain groups of youth are at heightened risk for systematic misevaluation via juvenile risk assessment. Directly inhibiting the RNR's risk principle, systematic misevaluation occurs when youths' cumulative risk scores are misaligned

with their actual likelihood of reoffending. While no risk assessment will predict recidivism with complete accuracy, misevaluation that is *systematic* indicates that certain criminogenic characteristics are consistently and methodologically mismeasured. Two forms of misevaluation, as well as their consequences, are discussed below (Butcher et al., 2014):

False Negatives. When juvenile risk assessments underestimate criminogenic risk (i.e., artificially low scores), youths' actual likelihood of reoffending exceeds the level estimated by their risk score. In these circumstances, courts may automatically divert youths from formal processing, rendering them ineligible for rehabilitative treatment and wraparound services. While it may not be the ideal venue for service delivery, contact with the juvenile justice system can catalyze youths' first opportunity to receive mental healthcare, addiction recovery treatment, and intensive school support (Pullmann et al., 2006). Accordingly, systematic misevaluation of these "false negatives" allows youths to re-enter their communities with unaddressed criminogenic needs (McCarter, 2016); as a result, these youths are at elevated likelihood for reoffending.

False Positives. When juvenile risk assessments overestimate criminogenic risk (i.e., artificially high scores), youths' actual likelihood of reoffending falls short of the level estimated by their risk score. In these circumstances, courts prescribe inappropriately punitive, cost- and time-intensive sanctions to youths who do not need them. Rather, these sanctions may unintentionally raise the criminogenic risk level of these "false positives". Prescription of highly restrictive sanctions (e.g., juvenile detention) promotes the development of deviant self-perception and disrupts participation in normative at-home routines (Cecile & Born, 2009; Gatti et al., 2009). Furthermore, association with higher risk peers through court-sanctioned programming may increase their propensity towards delinquent behavior (Leve & Chamberlain, 2005). Ergo, systematic misevaluation of these "false positives" indicates that the juvenile risk

assessment is erroneously misallocating court resources, at a potential detriment to youths' well-being.

THE CURRENT STUDY

Given these severe consequences, it is critical that juvenile risk assessment scores closely align with youths' latent level of criminogenic risk. Juvenile risk assessment scores correspond to the unweighted sum of all risk factors identified, a process hereinafter referred to as the Scoring-as-Usual method. Despite its near universal application, the Scoring-as-Usual method may generate imprecise risk estimates, enabling systematic misevaluation (Butcher, 2014; Vincent et al., 2012). The current study sought to improve measurement of criminogenic risk through the development of a Novel Scoring Algorithm, tailored to local patterns in data from a county-level analytic sample.

The Novel Scoring Algorithm was developed using initial scores from court-involved youth who received the Youth Level of Service/Case Management Inventory (YLS/CMI), the most widely utilized proprietary juvenile risk assessment instrument (Andrews & Bonta, 2010; Schwalbe, 2007). The YLS/CMI is a 41-item assessment that measures criminogenic risk in eight domains: Prior Dispositions/Offenses, Education, Leisure & Recreation, Attitudes & Orientation, Personality & Behavior, Peer Relations, Substance Abuse, and Family & Parenting (Andrews & Bonta, 2010). Innovations to YLS/CMI scoring are described below:

Innovation I: Adjust Item and Domain Weights

Early developers of juvenile risk assessments distinguished between "static risk factors", which are less amenable to change and more significantly tied to reoffending (e.g., prior offense history), and "dynamic risk factors" which are more easily remedied through effective court-sanctioned intervention (e.g., substance use) (OJJDP, 2015). Furthermore, assessment items on the YLS/CMI identify the same behavior at different frequencies or intensities (e.g., occasional versus chronic substance usage). Theoretically, dynamic risk factors at low frequencies should

16

have lower correlation with criminogenic risk, relative to static risk factors at high frequencies (Hoge & Andrews, 1996). However, the Scoring-as-Usual method constrains all risk factors to contribute equally to estimates of criminogenic risk. The Novel Scoring Algorithm tailored item weights to their correlation with latent domains of criminogenic risk, based upon the patterns in juvenile risk assessment scores from the county-level analytic sample.

Weighting was doubly necessary in the present context, as risk domain subscales on the YLS/CMI are unequally sized. For example, the Leisure & Recreation subscale encompasses three risk factors while the Personality & Behavior subscale encompasses seven. By failing to weight domain sub-scores, cumulative risk estimates are inherently biased towards domains with more risk factors (McNeish & Wolf, 2020). The Novel Scoring Algorithm weighted domain sub-scores to ensure that their relative contribution to cumulative risk estimates was empirically grounded, rather than reflective of arbitrary scale composition.

Innovation II: Integrate Risk and Protective Factors

Contemporary juvenile risk assessments are advantageous over other forms of risk evaluation because they include a separate, complementary assessment of criminogenic protective factors. While the protective factors identified can be used to facilitate wholistic case planning, they are omitted from estimates of cumulative criminogenic risk in Scoring-as-Usual procedures (Barnes-Lee, 2020). This creates significant challenges for interpretation and eliminates the potential for protective factors to enhance risk evaluation: the extent to which the presence of protective factors mitigates the influence of risk factors is unclear. To remedy this challenge, the Novel Scoring Algorithm integrated risk and protective factors into a single estimate of criminogenic risk.

PLAN OF WORK

In effort to improve juvenile risk assessment measurement models, the overall objective of this dissertation was to develop and evaluate a Novel Scoring Algorithm which innovates upon the conventions of Scoring-as-Usual method. In light of these innovations, it was hypothesized that risk estimates generated by Novel Scoring Algorithm would significantly improve the measurement of criminogenic risk (i.e., model fit) and prediction of recidivism (i.e., diagnostic accuracy) of the juvenile risk assessment instrument, over and above the Scoring-as-Usual method. To maximize responsivity to local criminogenic patterns, the Novel Scoring Algorithm was informed by official risk assessment and recidivism records collected from a county-level juvenile circuit court. Data sources and analytic methods are discussed below.

Methods

Sample. The study drew upon archival risk assessment and recidivism records from a pooled sample of 559 court-involved youth between the ages of 10 and 18 (mean [M] = 14.6 years, standard deviation [SD] = 1.4 years). Thirty-nine youths in the sample exceeded 16 years of age at the time of initial risk assessment which, at the time of data collection, was the maximum age of juvenile court jurisdiction. These 39 youth were initially petitioned to criminal court, and later waived to juvenile court via prosecutorial discretion. They were ultimately retained in the current analytic sample so that the present results encompass all youth who were evaluated via the YLS/CMI during the window of data collection.

Youth entered the supervision of a juvenile circuit court in a single mid-sized Midwestern County via truancy (34.2%) or delinquency (65.8%) division. The truancy division is a specialized branch of the court which offers targeted services designed to remove barriers to school attendance and promote educational success. Youth with chronically unexcused absences

were referred to truancy court by school truancy officers. Alternatively, youth processed in the delinquency division came under court contact through conventional means (i.e., via arrest or police referral), and were matched to individualized treatment plans designed to reduce likelihood of future delinquent involvement. Despite these distinctions, youth in the truancy and delinquency division participated in the same juvenile risk assessment process. Youth processed through both divisions were retained in analyses to emulate the generalist application and interpretation of juvenile risk assessment scores. However, to account for distinction in the selection and treatment of youth across divisions, post-hoc analyses examined differences in the performance of the risk assessment instrument between truant and delinquent subsamples.

The analytic sample represents all youth who were formally petitioned to juvenile court and adjudicated as delinquent or truant for the first time between September 2015 and December 2018. Sample demographic and charge information, collected via self-report at the time of initial risk assessment, is reported in Table 1. Ten youths in the analytic sample (1.8%) were missing indicators of race/ethnicity. Indicators of race/ethnicity in were found to be missing at random, as these cases did not otherwise differ from sample in cumulative number of risk factors, protective factors, or recidivism rates. Accordingly, they were retained in all aggregated analyses, as well in gender and court division comparisons. However, they were omitted from analyses comparing risk assessment performance across racial/ethnic cohorts.

Table 1.
Demographics and charge types of study sample.

Characteristic	N (%)
Gender	
Girls	194 (34.7%)
Boys	365 (65.3%)
Race/Ethnicity	
Caucasian/White	149 (26.7%)
Hispanic/Latinx	57 (10.2%)
African American/Black	231 (41.3%)

Table 1 (cont'd).

Characteristic	
Race/Ethnicity	
Multi-Racial	107 (19.1%)
Other	5 (<1.0%)
Charge Type	
Status	204 (36.5%)
Property	151 (27.0%)
Person	117 (20.9%)
Sex	29 (5.2%)
Public Ordinance	24 (4.3%)
Weapon	14 (2.5%)
Drug	7 (1.3%)
Other	9 (1.6%)

Measures. The core constructs of the proposed study were represented by the risk

assessment and recidivism measures utilized by the county-level juvenile circuit court at the time

of data collection.

The Youth Level of Services/Case Management Inventory (YLS/CMI) is an adapted

youth version of the Level of Service Inventory – Revised (LSI-R), an instrument designed to

evaluate criminogenic risk in court-involved adults (Andrews & Bonta, 2010). The psychometric

properties of the YLS/CMI have been rigorously investigated (see Andrews et al., 1986; Shields

& Simourd, 1991; Simourd et al., 1991, 1994); across these studies, 41 items have been retained,

having consistently demonstrated significant correlation with juvenile reoffending. Using factor

analysis, these items have been grouped into eight domains of criminogenic risk: Prior

Dispositions/Offenses, Education, Leisure & Recreation, Attitudes & Orientation, Personality &

Behavior, Peer Relations, Substance Abuse, and Family & Parenting. In line with criminological

theory, these domains represent both static and dynamic characteristics of the youth and their

proximal social environment (OJJDP, 2015).

Each of the eight domains on the YLS/CMI includes between 3 and 7 risk factors, which are scored dichotomously using a set of concretized, pre-determined criteria. (see Appendices A and B for a list of YLS/CMI items and intradomain bivariate correlations between items). In the current research setting, youth are classified at one of three risk levels based upon the cumulative number of unweighted risk factors identified across domains: low risk (8 or fewer risk factors), moderate risk (between 9-22 risk factors), or high risk (23 or more risk factors). This risk level classification informs several judicial decisions, including eligibility for diversion, duration of court supervision, and level of restrictiveness in sanctions. Descriptive information regarding sample risk scores is presented in Table 2.

The Protective Factors for Reducing Juvenile Reoffending (PFRJR) is a novel strengths-based assessment. It was developed in response to a growing desire to understand how prosocial characteristics may reduce odds of reoffending and increase responsivity to treatment in court-involved youth (Barnes-Lee & Campbell, 2020). The 22-item scale maps on to seven of the eight domains of the YLS/CMI (Prior Dispositions/Offenses, Education, Leisure & Recreation, Attitudes & Orientation, Personality & Behavior, Peer Relations, Substance Abuse, and Family & Parenting), and includes an additional domain identifying community-level strengths (see Appendices A and B for a list of PFRJR items and intradomain bivariate correlations between items) (Barnes-Lee, 2020). The factor structure of the PFRJR was confirmed via cross-validation and found to have strong internal consistency (Barnes-Lee, 2020). Like the YLS/CMI, the PFRJR is administered as part of the initial risk assessment and scored as a summative checklist of dichotomous protective factor items. Importantly, scores from the PFRJR are not integrated into estimates of cumulative criminogenic risk using the Scoring-as-Usual method, and do not influence the youth's risk level classification. Rather, PFRJR scores are used at case managers'

21

discretion to enhance treatment responsivity. Descriptive information regarding sample protective factor scores is presented in Table 2.

The Novel Scoring Algorithm and Scoring-as-Usual method were evaluated, in part, based on their diagnostic accuracy in correctly classifying youth as recidivant or desistant. Recidivism here refers to any additional adult or juvenile petitions (including felonies, misdemeanors, or criminal violations of probation) received within the 24 months immediately following the date of the youth's initial risk assessment. Given that the sample encompasses youth whose initial risk assessment took place between September 2015 and December 2018, the two-year window for measuring recidivism concluded in December 2020. Table 2 summarizes sample recidivism rates, both overall and by risk level classification.

Table 2.
Sample YLS/CMI scores, PFRJR scores, and recidivism rates.

			Risk Classification N (%)		
	Mean (*SD*)	Range	Low Risk	Moderate Risk	High Risk
YLS/CMI Score	16.1 (*6.8*)	1-42	80 (14.3%)	368 (66.1%)	108 (19.4%)
PFRJR Score	7.6 (*5.2*)	0-22	–	–	–
	Overall		Low Risk	Moderate Risk	High Risk
Recidivism Rate	44.4%		30.0%	44.8%	54.6%

Data Collection. Official risk assessment and recidivism records were obtained through a collaborative research partnership involving Michigan State University and the county-level juvenile circuit court.

The initial YLS/CMI and PFRJR were administered together via structured interview between a highly trained case manager and a recently adjudicated youth. Case managers scored each item dichotomously based upon youths' self-report, using a set of predetermined criteria. Criteria operationalize each risk and protective factor into concrete, easily identifiable terms. For example, to fulfill the criteria for chronic drug use, youth must disclose illegal drug use at least

twice per week or have a drug-related problem in one or more major life areas (e.g., drug-related arrest, school/employment citations, withdrawal symptoms). Case managers calculated cumulative risk scores using the Scoring-as-Usual method. Each risk assessment was reviewed by a trained research assistant for quality and completion, and entered into a secure database housed on the court's computers. All identifying information was removed from the risk assessment records prior to analysis.

Recidivism records include any additional juvenile or adult petitions accrued within two years of the initial risk assessment date. Adult petitions were obtained and synthesized with juvenile petitions through an integrated data management system involving the criminal and juvenile branches of the county circuit court. This information is obtained, merged with risk assessment records, and evaluated by the Michigan State University research team on an annual basis.

Analytic Plan. This dissertation was executed in three sequential analytic phases: development, evaluation, and cohort comparison. All latent variable modeling was conducted in MPlus Version 8.5 (Muthén & Muthén, 2020). Diagnostic testing was conducted in SPSS Version 27, and post-hoc comparisons of diagnostic accuracy were performed in R (v.4.1.2; R Core Team 2022).

The development of the Novel Scoring Algorithm was achieved through second-order Confirmatory Factor Analysis (CFA). CFA is a statistical technique used to examine how latent factors influence responses on measured variables (Kline, 2016; Takane & Deleeuw, 1987). Second-order CFA is utilized when a general construct (i.e., criminogenic risk) accounts for the variation between the latent factors (i.e., domains on the YLS/CMI and PFRJR) (Gould, 2015). CFA is best suited to addressing the specific aim, as opposed to other methods of factor analysis

(e.g., Exploratory Factor Analysis, Principal Components Analysis), because the factor structure was specified a priori by the nine domains on the YLS/CMI and the PFRJR. CFA is the only method of factor analysis which analyzes multi-dimensional constructs based on established theory (Kline, 2016).

The execution of the second order CFA model included two levels. At the first level, the 64 collective item indicators on the YLS/CMI and PFRJR were loaded onto nine first-order factors, reflective of the nine domains of the YLS/CMI and PFRJR (i.e., Prior Dispositions/Offenses, Education, Leisure & Recreation, Attitudes & Orientation, Personality & Behavior, Peer Relations, Substance Abuse, Family & Parenting, and Community) (see Figure 1). Factor loadings among the first-order factors (represented by single-arrowed lines between the nine latent domains and observed item indicators) indicated the magnitude and direction of the association between each item and its corresponding domain. Factor loadings from the first-order factors were used as assessment item weights in the Novel Scoring Algorithm.

At the second level, the nine first-order factors were loaded onto one second-order factor, reflecting cumulative criminogenic risk. Factor loadings from the second-order factor (represented by the single-arrowed lines between criminogenic risk and the nine latent domains) indicated the magnitude and direction of the association between each domain and cumulative criminogenic risk. Factor loadings from the second-order factor weighted domain scores in the Novel Scoring Algorithm. Weighting scores at both the item- and domain-level was necessary, given the variation in number of items per domain. Failing to adjust the weight of the domain scores would bias cumulative estimates toward the domains with more assessment items.

Figure 1.

Factor model describing the relationship between assessment items and domains on the YLS/CMI and PFRJR.
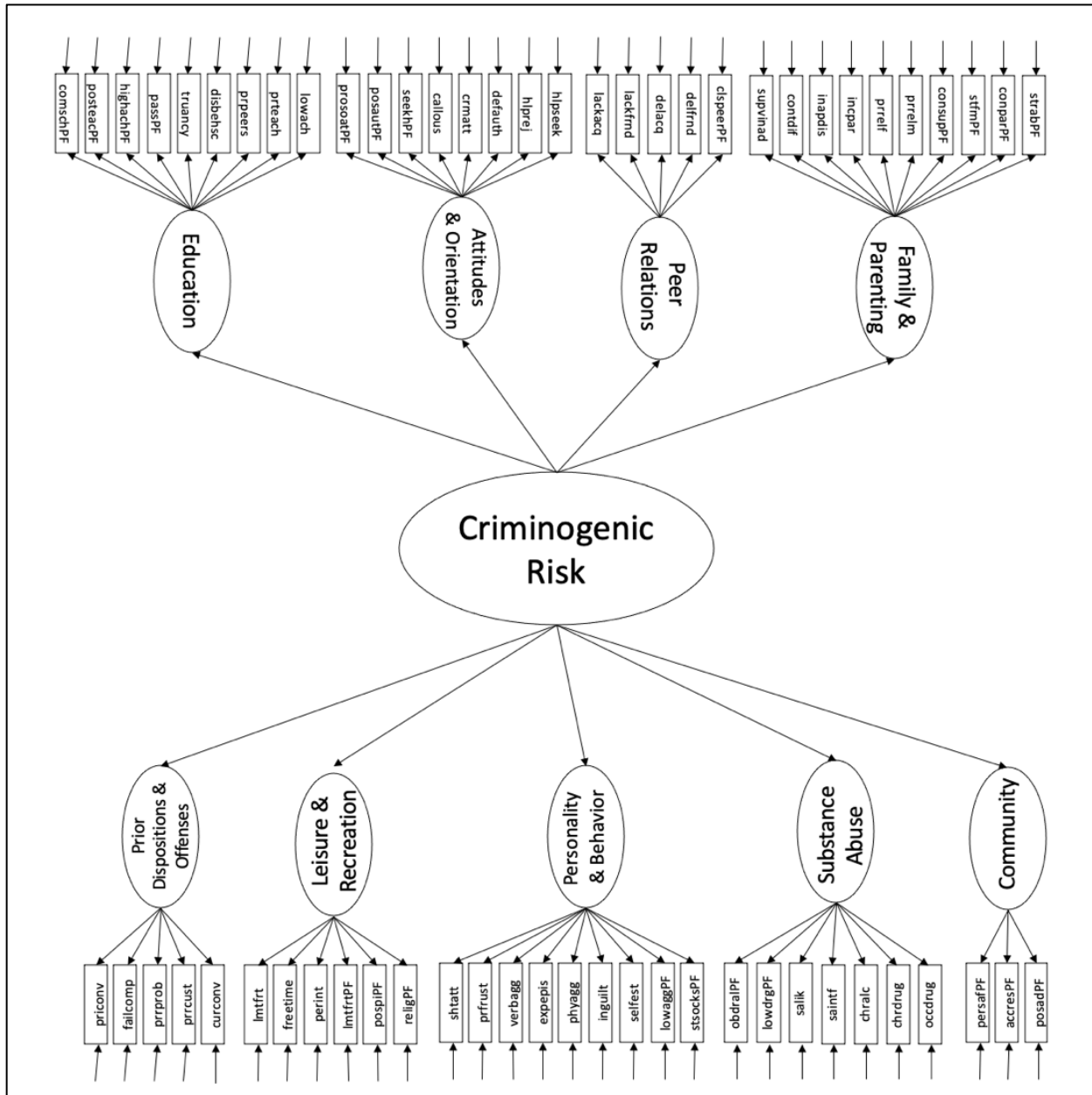
Note. Items ending in "PF" denote protective factors.

The Scoring-as-Usual method involves summing all unweighted risk factors identified. While sum scoring and factor analysis are typically vetted as competing methods, McNeish and

Wolf (2020) argue that both are forms of latent variable modeling. Accordingly, the Scoring-as-Usual model was estimated as a constrained form of the Novel Scoring Algorithm, wherein all first- and second-order factor loadings of the YLS/CMI were set to one. The PFRJR was omitted from this model because protective factors are not included in composite estimates of risk using the Scoring-as-Usual method.

The Novel Scoring Algorithm and the Scoring-as-Usual method were evaluated on several criterion: (1) *absolute fit*, indicating how well the models explained variation in the observed data (Kline, 2016); (2) *relative* or *incremental fit*, indicating how well the models improved fit of the data relative to a null model (Kline, 2016); and (3) *diagnostic accuracy,* measuring the performance of the models in correctly classifying youth as recidivant or desistant (Rice & Harris, 2005). Given the known penalties of basing evaluation on one indicator alone, multiple fit indices were calculated and interpreted. Table 3. describes the evaluation metrics and their criteria for acceptable fit.

Table 3.
Latent variable model evaluation metrics and criteria for acceptable fit.

Index	Type of Index	Description	Criteria for Acceptable Fit
Chi-Square Fit Index	Absolute fit	Indicates discrepancy in the covariance and mean matrices between the specified model and the observed data. Non-significant test statistics indicate low discrepancy, signaling good fit (McNeish & Wolf, 2020)	$p > 0.05$
Standardized Root Mean Square Residual (SRMR)	Absolute fit	Badness-of-fit index; indicates the overall discrepancy between the observed and predicted variable correlations (Kline, 2016)	SRMR < 0.10
Root Mean Square Error of Approximation (RMSEA)	Absolute fit	Badness-of-fit index; indicates the discrepancy between the specified and observed covariance matrices, adjusting for model complexity (Cangur & Ercan, 2015)	RMSEA < 0.08

Table 3 (cont'd).

Index	Type of Index	Description	Criteria for Acceptable Fit
Comparative Fit Index/Tucker Lewis Index (CFI/TLI)	Relative fit	Goodness-of-fit index; indicates whether the specified model improves the fit of the data by 90%, relative to the null model (Kline, 2016)	CFI $\geq$ 0.90
Area Under the Receiver Operating Characteristic (AUROC)	Clinical performance	Performance metric for model discrimination; indicates the specified model's diagnostic accuracy in predicting outcomes over chance (Rice & Harris, 2005)	AUC = 0.55, .64, & 0.71 for small, moderate, and large effect size

Composite scores yielded by juvenile risk assessments may have differential implications for recidivism by virtue of youths' demographic (e.g., race/ethnicity, gender) and charge-related (e.g., truancy, delinquency) characteristics. Differential diagnostic performance directly inhibits fair and equitable court decision-making (Anderson et al., 2016; Campbell et al., 2018; Onifade et al., 2009). To detect variation in diagnostic accuracy, a series of cohort comparisons were conducted for youth across gender, race/ethnicity, and division of the court (i.e., truancy, delinquency). Within cohort comparisons were conducted through a series of pairwise test of independent-group area differences, performed on both the Novel Scoring Algorithm and the Scoring-as-Usual method. Between model comparisons were conducted through a series of DeLong tests (DeLong et al., 1988).

These metrics of absolute fit, relative fit, and diagnostic accuracy were used to identify and select the model that most accurately measures criminogenic risk in court-involved youth. For the purpose of this inquiry, in order for the Novel Scoring Algorithm to be retained as a procedure for estimating likelihood of recidivism, it must improve metrics in all three domains (i.e., absolute fit, relative fit, diagnostic accuracy) over the Scoring-as-Usual method.

RESULTS

Phase Ia: Development of Novel Scoring Algorithm

Congruent with the multidimensional structure of the YLS/CMI and PFRJR, the Novel

Scoring Algorithm loaded 63 discreet binary assessment items (i.e., the risk and protective factor

items) onto nine latent first-order factors (i.e., the risk and protective factor domains). The nine

latent first-order factors were subsequently loaded onto a single second order factor, reflective of

composite criminogenic risk. The first factor loading for each first- and second-order latent

variable was constrained to one, representing the unit loading identification (ULI) constraint

(Kline, 2016). ULI constraints scale the latent factors to the YLS/CMI's and PFRJR's units of

measurement. In a single sample analysis, the indicator selected as the ULI is arbitrary and holds

no bearing on the fit of the model (Kline, 2016). All latent variable means, intercepts, and error

variances were freely estimated, meaning that they reflected the corresponding parameters

observed in the data.

To account for redundancy between discreet variables, modification indices

recommended covariances between the following pairs of items: disruptive classroom behavior

with problems with teachers ($r(557) = 0.57$, $p < 0.05$), passing (protective factor) with low

achievement ($r(557) = -0.66$, $p < 0.05$), involvement in organized activities (protective factor)

with lack of organized activities ($r(557) = -0.84$, $p < 0.05$), lack of positive friends with lack of

positive peer acquaintances ($r(557) = 0.57$, $p < 0.05$), some delinquent friends with some

delinquent peer acquaintances ($r(557) = 0.64$, $p < 0.05$), consistent supervision (protective factor)

with inadequate supervision ($r(557) = 0.58$, $p < 0.05$), and actively seeking help with not seeking

help ($r(557) = 0.50$, $p < 0.05$). It is ill-advised to add paths to the model based upon modification

indices without first consulting theory. Nevertheless, these covariances are logically justified, as

the item pairs represent the same characteristic at different intensities (e.g., delinquent peers versus delinquent friends) or the same characteristic at opposite ends of the spectrum (e.g., actively seeking help versus not seeking help). The theoretical relationship between these items is further confirmed by the strong positive and inverse correlations observed between each item pair (see Appendix B). A summary of the modified model, including all first- and second-order factor loadings, is presented in Appendix C.

Using the Novel Scoring Algorithm, possible composite estimates of criminogenic risk ranged from -39.67 (i.e., having all protective factors and no risk factors) to 53.86 (i.e., having all risk factors and no protective factors). However, observed estimates within the sample ranged from -36.14 to 38.14 ($M = 8.26$; $SD = 16.64$ points). Table 4 describes the sample scores rendered by the Novel Scoring Algorithm, in aggregate and across gender, racial/ethnic, and court division cohorts.

Phase Ib: Development of Scoring-as-Usual Method

The Scoring-as-Usual model loaded the 41 items of the YLS/CMI onto eight latent first-order factors. The ninth first-order factor, as modeled in the Novel Scoring Algorithm, represents the Community domain on the PFRJR, which is composed solely of protective factors. The eight first-order factors were in turn loaded onto one second-order factor, reflective of composite criminogenic risk. Based upon modification indices, the Scoring-as-Usual method additionally specified covariances between the three following pairs of assessment items: disruptive classroom behavior with problems with teachers, some delinquent friends with some delinquent acquaintances, and lack of positive friends with lack of positive peer acquaintances. The four additional covariances specified in the Novel Scoring Algorithm were not relevant to the Scoring-as-Usual model, as they involved protective factors.

To emulate the process of summing the unweighted total of all YLS/CMI assessment items, all unweighted factor loadings of the Scoring-as-Usual model were fixed to one. A summary of the model, including all first- and second-order factor loadings, is presented in Appendix D. Using the Scoring-as-Usual method, possible composite estimates of criminogenic risk ranged from 0 (i.e., having no risk factors) to 41 (i.e., having all risk factors). However, observed estimates within the sample ranged from 1 to 35, with a mean score of 16.18 and a standard deviation of 6.76 points. Table 4 describes the sample scores rendered by the Scoring-as-Usual method, in aggregate and across gender, racial/ethnic, and court division cohorts.

Table 4.
Composite risk estimates using the Novel Scoring Algorithm and Scoring-as-Usual method.

	Novel Scoring Algorithm		
	Mean Risk Score	Standard Deviation	Range
All Youth	8.26	16.64	-36.16 – 38.14
Gender			
Girls	8.32	16.16	-29.61 – 38.14
Boys	8.22	16.91	-36.14 – 37.02
Race/Ethnicity			
African American/Black	10.61	15.96	-34.16 – 38.14
Caucasian/White	3.80	17.50	-36.14 – 35.16
Multi-Racial	10.23	16.21	-28.99 – 36.86
Hispanic/Latinx	8.77	14.39	-31.99 – 27.95
Other	-11.23	21.40	-34.08 – 23.97
Division			
Delinquency	8.81	17.33	-36.14 – 38.14
Truancy	7.19	15.1	-28.99 – 34.41
	Scoring-as-Usual Method		
	Mean Risk Score	Standard Deviation	Range
All Youth	16.18	6.76	1 – 35
Gender			
Girls	15.68	6.63	2 – 31
Boys	16.45	6.83	1 – 35
Race/Ethnicity			
African American/Black	17.60	6.70	2 – 35
Caucasian/White	14.12	6.01	1 – 29
Multi-Racial	16.62	6.63	1 – 29
Hispanic/Latinx	15.63	6.85	2 – 25
Other	10.60	6.11	2 – 19

Table 4 (cont'd).

	Scoring-as-Usual Method		
	Mean Risk Score	Standard Deviation	Range
Division			
Delinquency	17.03	6.94	1 – 35
Truancy	14.56	6.10	1 – 30

Comparing Risk Scores Across Models

A Pearson correlation was conducted to estimate the linear relationship between the composite risk scores generated by the two measurement models. The correlation was strong, positive, and statistically significant ($r(557)=0.91$, $p <0.01$), indicating that only 17.19% ($\sigma = 1 - r^2$) of the variance in scores differed between the Novel Scoring Algorithm and the Scoring-as-Usual method. Figure 2 below visualizes the strong, positive linear relationship between scores generated by the two models.

Taken together, these results indicate that youths' relative position on the spectrum of criminogenic risk remained largely unchanged, regardless of the scoring method employed. The high degree of shared variance between the two models highlights the strong, inverse relationship between risk and protective factors. In other words, the youth whose risk scores were lowered by protective factors using the Novel Scoring Algorithm were already low risk using the Scoring-as-Usual method. Concurrently, patterns in unit weighting reflect the observed relationship between each assessment item and its intradomain constituents. As a result, the youth whose risk scores were raised by heavily weighted items using the Novel Scoring Algorithm were already high risk using the Scoring-as-Usual method.

Figure 2.

Relationship between composite risk scores estimated by the Novel Scoring Algorithm and Scoring-as-Usual method.
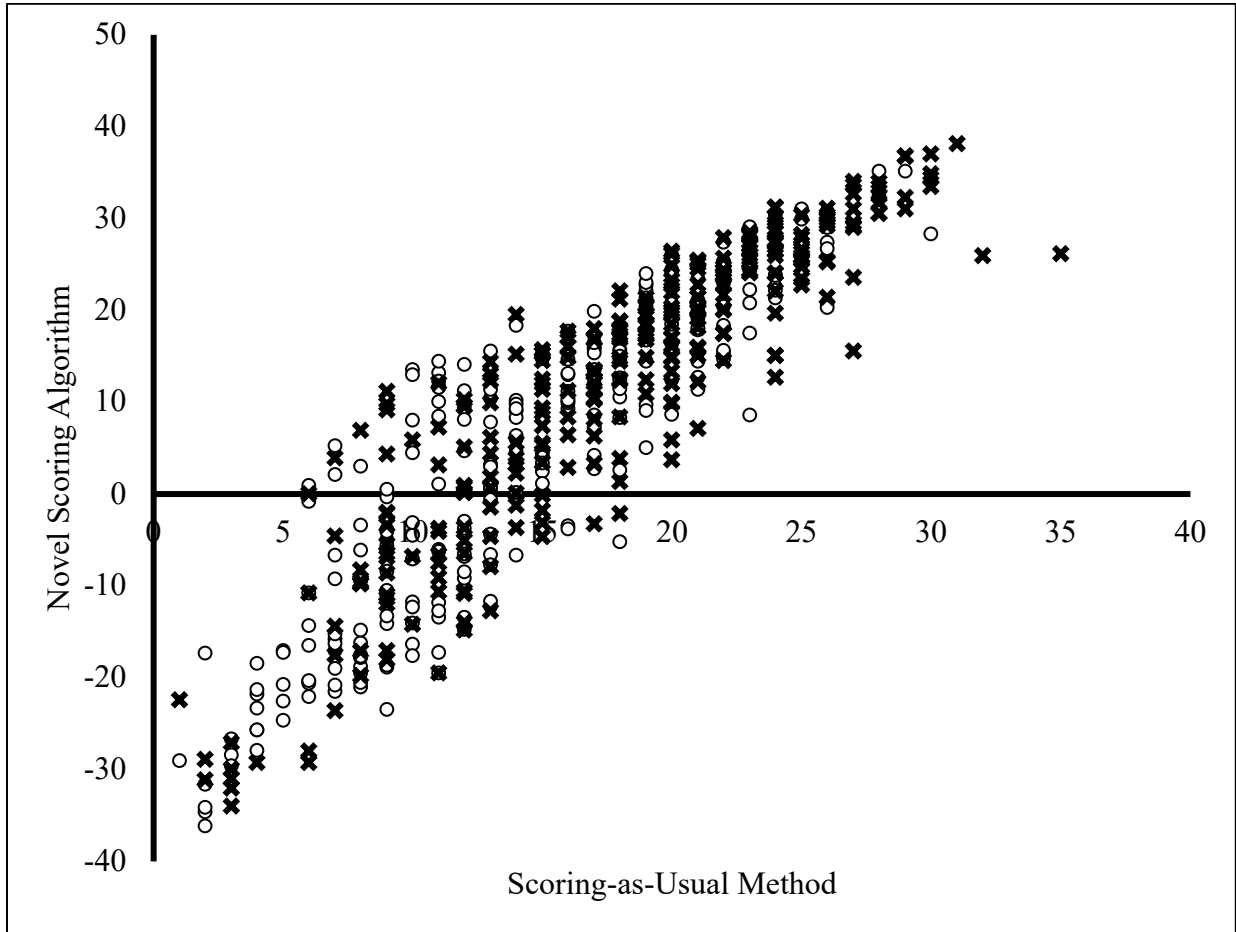
Note. Open circles denote youth who desisted and Xs denote youth who recidivated.

Phase II: Evaluation of Scoring Methods

   Absolute Fit. Absolute fit indices are used to evaluate how well a latent variable model explains variation in the observed data (Kline, 2016). The Novel Scoring Algorithm and the Scoring-as-Usual method were compared on absolute fit using the chi-square fit index, the Standardized Root Mean Square Residual (SRMR), and the Root Mean Square Error of Approximation (RMSEA) (see Table 3 for description of fit indices and thresholds for acceptable fit).

Likely penalized by the large sample size, both the Novel Scoring Algorithm and the Scoring-as-Usual method yielded significant chi-square values (Novel Scoring Algorithm: $\chi^2(1,873)=3,072$, $p < 0.05$; Scoring-as-Usual: $\chi^2(808)=1,932.63$, $p < 0.05$) (Bentler & Bonett, 1980). Significant chi-square values indicate discrepancies between the covariance matrices in the specified models and the observed data (McNeish & Wolf, 2020). Additionally, both models exceeded the threshold for acceptable SRMR (Novel Scoring Algorithm: SRMR = 0.11; Scoring-as-Usual: SRMR = 0.17), indicating significant discrepancies between the observed and predicted variable correlations (Kline, 2016).

However, both models yielded acceptable RMSEA values (Novel Scoring Algorithm: RMSEA = 0.03; Scoring-as-Usual: RMSEA = 0.05), indicating that, after adjusting for model complexity, the specified and observed covariance matrices in both models were comparable. Even though both models yielded acceptable RMSEA values, the constraints imposed in the Scoring-as-Usual method significantly worsened the absolute fit of the data (ΔRMSEA > 0.015) (Chen, 2007; Cheung & Rensvold, 2002). Taken together, the analysis of absolute fit indicated that the Novel Scoring Algorithm better explained variation in the observed data over and above the Scoring-as-Usual method.

Relative Fit. Relative or incremental fit indices are used to evaluate how well a latent variable model improves fit of the data over a null model (Kline, 2016). The Novel Scoring Algorithm and the Scoring-as-Usual method were compared on absolute fit using the Comparative Fit Index/Tucker Lewis Index (CFI/TLI) (see Table 3 for description of fit indices and thresholds for acceptable fit). Only the Novel Scoring Algorithm yielded acceptable relative fit (CFI/TLI = 0.94), indicating that the estimated model improved overall fit by 94% over a null model. Conversely, the Scoring-as-Usual method failed to achieve acceptable fit (CFI/TLI =

0.83). Using thresholds recommended by Chen (2007) and Cheung & Rensvold (2002), the

constraints imposed in the Scoring-as-Usual method significantly worsened the relative fit of the

data (ΔCFI/TLI > 0.01). Taken together, the analysis of relative fit indicated that the Novel

Scoring Algorithm yielded greater improvement over a null model when compared to the

Scoring-as-Usual method. Both absolute and relative fit indices for the measurement models are

summarized in Table 5.

Table 5.
Relative and absolute fit indices for the Novel Scoring Algorithm and Scoring-as-Usual
method.

| | # param | χ^2 | | | SRMR | RMSEA | CFI/TLI |
		Est.	df	p			
Novel Scoring Algorithm	142	3,072.35	1,874	.00	.11	.03	.94
Scoring-as-Usual	53	1,932.63	808	.00	.17	.05	.83

Diagnostic Accuracy. The final step in the evaluation process compared the diagnostic

accuracy of the measurement models in correctly classifying youth as either recidivant (i.e.,

received one or more petitions in the two years post-initial risk assessment) or desistant (i.e., did

not receive any petitions in the two years post-initial risk assessment) (Rice & Harris, 2005).

Overall diagnostic accuracy was assessed using Area Under the Curve (AUC) values derived

from a univariate logistic regression predicting recidivism from composite estimates of

criminogenic risk. The direction and magnitude of risk misclassification was further probed

through an analysis of model sensitivity (i.e., true positive rate) and 1 – specificity (i.e., false

positive rate). Results for the Novel Scoring Algorithm and Scoring-as-Usual method are

presented below.

Drawing upon the study sample of 559 youth, composite criminogenic risk scores derived

from the Novel Scoring Algorithm method significantly predicted recidivism in the first two

years following the youths' initial contact with the court. The estimated odds ratio indicated that

a one-unit increase in composite risk score increased the likelihood of recidivism by 3.00%

(*Exp[B]* = 1.03, *p* < 0.01, *95% CI* [1.02 1.04]) (see Table 6. for model summary).

Next, a confusion matrix was generated to assess the accuracy of the Novel Scoring Algorithm as a classification method. Overall, the model correctly classified 59.53% of court-involved youths as either recidivant or desistant. Model sensitivity was 0.59, indicating that 59.43% of youth who reoffended were correctly classified as recidivant (i.e., "true positives"). Additionally, 1 – specificity was 0.54, indicating that 53.58% of youth who did not reoffend were incorrectly classified as recidivant (i.e., "false positives"). The Area Under the Curve (AUC) index, which summarizes the overall diagnostic accuracy of the Novel Scoring Algorithm, was 0.64, which equates to a moderate effect size in violence risk assessment literature (Rice & Harris, 2005).

 Drawing upon the same sample of 559 youth, estimates of criminogenic risk derived from the Scoring-as-Usual method significantly predicted recidivism in the first two years following the youths' initial contact with the court. The estimated odds ratio indicated that each additional unweighted risk factor increased the likelihood of recidivism by 8.00% (*Exp[B]* = 1.08, *p* < 0.01, *95% CI* [1.05 1.11]) (see Table 6. for model summary).

Next, a confusion matrix was generated to assess the accuracy of the Scoring-as-Usual procedure as a classification method. Overall, the model correctly classified 61.33% of court-involved youths as either recidivant (i.e., received at least one petition in the two years following initial court contact) or desistant (i.e., did not receive any additional petitions in the two years following initial court contact). Model sensitivity was 0.62, indicating that 61.66% of youth who did reoffend were correctly classified as recidivant (i.e., "true positives"). However, 1 – specificity was 0.47, indicating that 47.17% of youth who did not reoffend were incorrectly

35

classified as recidivant (i.e., "false positives"). The Area Under the Curve (AUC) index for the

Scoring-as-Usual method was 0.65, which again falls within disciplinary standards of a moderate

effect size (Rice & Harris, 2005).

Table 6.
Diagnostic accuracy of the Novel Scoring Algorithm and Scoring-as-Usual method.

Novel Scoring Algorithm					
				95% CI for Odds Ratio	
	B (*SE*)	Wald	Odds Ratio	Upper	Lower
Intercept	-0.15** (*0.09*)	2.40	0.86		
Criminogenic Risk Score	0.03** (*0.01*)	30.26	1.03	1.02	1.04
	Predicted Desistant			Predicted Recidivant	
Observed Desistant	123			142	
Observed Recidivant	83			208	
Scoring-as-Usual					
				95% CI for Odds Ratio	
	B (*SE*)	Wald	Odds Ratio	Upper	Lower
Intercept	-1.18** (*0.23*)	25.32			
Criminogenic Risk Score	0.08** (*0.01*)	34.07	1.08	1.05	1.11
	Predicted Desistant			Predicted Recidivant	
Observed Desistant	140			125	
Observed Recidivant	90			201	

**$p < 0.01$

The AUC estimates yielded by the Novel Scoring Algorithm (AUC = 0.64) and Scoring-

as-Usual method (AUC = 0.67) were compared against each other using a DeLong test (DeLong,

1988). In diagnostic testing, DeLong tests are used to evaluate multiple AUC estimates derived

from different predictors (i.e., composite risk estimates from the two measurement models) on

the same set of data (DeLong, 1988). Results from the DeLong test indicated no statistically

significant differences in the AUC estimates produced between the Novel Scoring Algorithm and

the Scoring-as-Usual method (z-score = 0.32, p = 0.75, 95% CI [-0.02 0.02]).

Summary of Evaluation. In sum, the Novel Scoring Algorithm better models the observed variation in measured juvenile risk assessment data, as evidenced by appreciably better indices of relative and absolute fit. These findings suggest that risk and protective factors covary with one another to different degrees, and failing to account for this in risk measurement creates significant psychometric imprecision. Concurrently, the Novel Scoring Algorithm held no relative advantage over the Scoring-as-Usual method in accurately distinguishing youth who recidivated from those who desisted. Due to these comparable rates of diagnostic accuracy, results of the evaluation ultimately affirm Scoring-as-Usual as an acceptable method of estimating likelihood of recidivism.

Phase III: Cohort Comparisons

Most juvenile risk assessments, including the YLS/CMI, are designed to be administered and interpreted using the same procedure, regardless of youths' demographics (i.e., gender, race/ethnicity) or charge-related (i.e., delinquency, truancy) characteristics. Accordingly, it is critical to ensure that the psychometric properties and diagnostic accuracy of these generalist risk assessment instruments are consistent for all youth. Due to small cell sizes across gender, racial/ethnic, and court division cohorts, comparing the relative and absolute fit of the YLS/CMI between these groups via multigroup CFA lies beyond the scope of this dissertation. However, a series of subgroup analyses were conducted to detect variation in overall diagnostic accuracy between and within the Novel Scoring Algorithm and the Scoring-as-Usual method (see Table 7). Assessing diagnostic accuracy holds the most immediate relevance for equitable justice administration, as results denote patterns in under- and overestimation of risk.

Table 7.
Diagnostic accuracy across sample cohorts.

	Novel Scoring Algorithm			
	Two-Year Recidivism Rate	Sensitivity (True Positive Rate)	1 - Specificity (False Positive Rate)	AUC
All Youth	52.33%	0.59	0.54	0.64
Gender				
Girls	43.46%	0.35	0.22	0.60
Boys	56.99%	0.81	0.60	0.66
Race/Ethnicity				
African American/Black	63.16%	0.88	0.71	0.65
Caucasian/White	40.27%	0.08	0.04	0.59
Multi-Racial	51.40%	0.67	0.46	0.68
Hispanic/Latinx	50.87%	0.66	0.71	0.57
Other	0.00%	--	--	--
Division				
Delinquency	60.33%	0.87	0.71	0.63
Truancy	36.70%	0.32	0.11	0.67
	Scoring-as-Usual Method			
	Two-Year Recidivism Rate	Sensitivity (True Positive Rate)	1 – Specificity (False Positive Rate)	AUC
All Youth	52.33%	0.62	0.47	0.65
Gender				
Girls	43.46%	0.35	0.22	0.59
Boys	56.99%	0.75	0.51	0.67
Race/Ethnicity				
African American/Black	63.16%	0.90	0.71	0.65
Caucasian/White	40.27%	0.03	0.01	0.56
Multi-Racial	51.40%	0.38	0.70	0.69
Hispanic/Latinx	50.87%	0.66	0.40	0.62
Other	0.00%	--	--	--
Division				
Delinquency	60.33%	0.87	0.77	0.61
Truancy	36.70%	0.32	0.13	0.67

Gender Variation in Diagnostic Accuracy. When analyzed independently, both models produced statistically comparable rates of diagnostic accuracy between boys and girls (see Table 8). Additionally, when comparing models against each other, the Novel Scoring Algorithm and the Scoring-as-Usual method predicted recidivism to equivalent degrees of diagnostic accuracy for both girls and boys (see Table 9). Taken together, these findings reflect the results of the

overall sample: the Novel Scoring Algorithm held no relative advantage over the Scoring-as-Usual method in predicting recidivism for youth across gender.

While AUC is considered best overall indicator of diagnostic accuracy, taking stock of model sensitivity (i.e., true positive rate) and 1 – specificity (i.e., false positive rate) offers greater insight on the direction and magnitude of risk misclassification (Mossman, 1994; Swets et al., 2000; Rice & Harris, 2005). Model sensitivity among boys was elevated relative to the full sample (Novel Scoring Algorithm: Sensitivity = 0.81; Scoring-as-Usual: Sensitivity = 0.75), indicating that most recidivist boys are correctly identified using both scoring methods. Additionally, 1 – specificity among boys was slightly elevated relative to the full sample (Novel Scoring Algorithm: 1 – Specificity = 0.60; Scoring-as-Usual: 1 – Specificity = 0.51), indicating that over half of desistant boys are incorrectly identified as recidivist using both scoring methods (see Table 7). In other words, composite juvenile risk assessment scores slightly overpredicted recidivism in court-involved boys, regardless of the scoring method employed.

Conversely, model sensitivity among girls fell short of the level estimated for the full sample (Novel Scoring Algorithm & Scoring-as-Usual: Sensitivity = 0.35), signaling that most girls who recidivate are not correctly identified using either scoring method. Concurrently, 1 – specificity among girls also fell short of the level estimated for the full sample (Novel Scoring Algorithm & Scoring-as-Usual: Sensitivity = 0.22), indicating that fewer than 25% of desistant girls are incorrectly classified as recidivist using both scoring methods. Taken together, composite juvenile risk assessment scores underpredicted recidivism in court-involved girls, regardless of the scoring method employed.

Racial/Ethnic Variation in Diagnostic Accuracy. When analyzed independently, both models produced statistically comparable rates of diagnostic accuracy for youth across

racial/ethnic groups (i.e., African American/Black, Caucasian/White, Multi-Racial, Hispanic/Latinx, and Other) (see Table 8). Additionally, when comparing models against each other, the Novel Scoring Algorithm and the Scoring-as-Usual method predicted recidivism to equivalent degrees of diagnostic accuracy for all racial/ethnic groups (see Table 9). In congruence with previous findings, these results suggest that the Novel Scoring Algorithm neither improves nor worsens the overall diagnostic accuracy of the Scoring-as-Usual method for youth across race/ethnicity.

For African American/Black youth, both models correctly identified most recidivists (Novel Scoring Algorithm: Sensitivity = 0.88; Scoring-as-Usual: Sensitivity = 0.90); however, nearly three quarters of desistant youth were incorrectly classified as recidivist (Novel Scoring Algorithm & Scoring-as-Usual: 1 - Specificity = 0.71). A similar, but less extreme, pattern was observed among Multi-Racial youth (Novel Scoring Algorithm: Sensitivity = 0.67, 1 - Specificity=0.46; Scoring-as-Usual: Sensitivity = 0.38, 1 – Specificity = 0.70) and Hispanic/Latinx youth (Novel Scoring Algorithm: Sensitivity = 0.66, 1 - Specificity=0.71; Scoring-as-Usual: Sensitivity = 0.66, 1 – Specificity = 0.40). Taken together, composite risk estimates overpredicted recidivism among youth of color, with the most significant degree of misevaluation observed in African American/Black youth.

Conversely, model sensitivity among Caucasian/White youth was the lowest of all racial/ethnic groups (Novel Scoring Algorithm: Sensitivity = 0.08; Scoring-as-Usual: Sensitivity = 0.03), signaling that the overwhelming majority of Caucasian/White youth who recidivate are not correctly identified using either scoring method. Additionally, 1 – specificity among Caucasian/White youth was also exceedingly low (Novel Scoring Algorithm: 1 – Specificity: 0.04; Scoring-as-Usual: 1 - Specificity = 0.01), indicating that fewer than 5% of desistant

40

Caucasian/White youth are incorrectly classified as recidivist. Accordingly, composite risk estimates underpredicted recidivism among White youth, regardless of the scoring method employed.

Court Division Variation in Diagnostic Accuracy. When analyzed independently, both models produced statistically comparable rates of diagnostic accuracy for youth across court divisions (i.e., delinquency, truancy) (see Table 8). Additionally, when comparing models against each other, the Novel Scoring Algorithm and the Scoring-as-Usual method predicted recidivism to equivalent degrees of diagnostic accuracy for both truant and delinquent youth (see Table 9). Once more, these results indicate that Novel Scoring Algorithm and the Scoring-as-Usual method are equally accurate predictors of recidivism across court division.

Model sensitivity among delinquent youth was elevated relative to the full sample (Novel Scoring Algorithm: Sensitivity & Scoring-as-Usual Method = 0.87), indicating that 87% recidivist delinquent youth are correctly identified using both scoring methods. Additionally, 1 – specificity among delinquent youth was elevated relative to the full sample (Novel Scoring Algorithm: 1 – Specificity = 0.71; Scoring-as-Usual: 1 – Specificity = 0.77), indicating that over two thirds of desistant delinquent youth were incorrectly classified using both scoring methods. In sum, these results indicate that composite risk estimates overpredicted recidivism among delinquent youth, regardless of the scoring method employed.

Conversely, model sensitivity among truant youth fell short of the level estimated for the full sample (Novel Scoring Algorithm & Scoring-as-Usual: Sensitivity = 0.32), signaling that only 32% of truant youth who recidivate are correctly identified using both scoring methods. However, 1 – specificity among truant youth also fell short of the level estimated for the full sample (Novel Scoring Algorithm: 1 – Specificity = 0.13; Scoring-as-Usual: 1- Specificity =

0.11), indicating that no more than 13% of desistant truant youth are incorrectly classified as recidivist. Taken together, these results indicate that composite risk estimates underpredicted recidivism among truant youth, regardless of the scoring method employed.

Table 8.
Comparing within-model diagnostic accuracy across cohorts.

Novel Scoring Algorithm					
				95% CI for Z-Score	
	AUC Difference	Z-Score	p	Lower	Upper
Gender					
Boys & Girls	0.06	1.24	0.22	-0.16	0.04
Race/Ethnicity					
African American/Black & Caucasian/White Youth	0.06	0.98	0.33	-0.18	0.06
African American/Black & Multi-Racial Youth	0.03	0.67	0.51	-0.17	0.08
African American/Black & Hispanic/Latinx Youth	0.08	0.94	0.35	-0.25	0.08
Caucasian/White & Multi-Racial Youth	0.11	1.22	0.22	-0.23	0.52
Caucasian/White & Hispanic/Latinx Youth	0.02	0.24	0.81	-0.16	0.20
Multi-Racial & Hispanic/Latinx Youth	0.11	1.16	0.25	-0.08	0.29
Division					
Delinquency & Truancy	-0.04	-.78	0.44	-0.14	0.06
Scoring-as-Usual Method					
				95% CI for Z-Score	
	AUC Difference	Z-Score	p	Lower	Upper
Gender					
Boys & Girls	0.08	1.54	0.12	-0.18	0.02
Race/Ethnicity					
African American/Black & Caucasian/White Youth	0.09	1.41	0.16	-0.21	0.03
African American/Black & Multi-Racial Youth	0.03	0.42	0.68	-0.15	0.10

Table 8 (cont'd).

	Scoring-as-Usual Method				
				95% CI for Z-Score	
	AUC Difference	Z-Score	p	Lower	Upper
African American/Black & Hispanic/Latinx Youth	0.03	0.37	0.71	-0.20	0.13
Caucasian/White & Multi-Racial Youth	0.13	1.83	0.07	-0.27	0.01
Caucasian/White & Hispanic/Latinx Youth	0.06	0.63	0.53	-0.23	0.12
Multi-Racial & Hispanic/Latinx Youth	0.07	0.81	0.07	-0.11	0.25
Division					
Delinquency & Truancy	0.06	2.24	0.21	-0.16	0.04

Table 9.

DeLong tests comparing between-model performance across sample cohorts.

	AUC Difference	Z-Score	p	95% CI for Z-Score	
				Lower	Upper
All Youth	0.01	-0.32	0.75	-0.02	0.02
Gender					
Boys	0.01	0.32	0.75	-0.02	0.02
Girls	0.01	0.57	0.57	-0.03	0.05
Race/Ethnicity					
African American/Black	<0.01	0.02	0.98	-0.03	0.03
Caucasian/White	0.03	0.21	0.21	-0.01	0.07
Hispanic/Latinx	0.05	1.31	0.19	-0.13	0.03
Multi-Racial	0.01	0.69	0.49	-0.06	0.03
Division					
Delinquency	0.01	1.57	0.12	<-0.01	0.04
Truancy	<0.01	0.28	0.78	-0.04	0.0.3

DISCUSSION

Juvenile risk assessment has become an increasingly integral component of evaluating and treating court-involved youths (JJGPS, 2020). The purpose of this study was to develop and evaluate a Novel Scoring Algorithm for estimating composite criminogenic risk, based upon patterns of risk and protective factors in a county-level sample. Composite criminogenic risk estimates generated from the Novel Scoring Algorithm were highly correlated with those generated from Scoring-as-Usual (r(557)=0.91, p <0.01), indicating substantial shared variance between the two scoring methods. Put simply, the Novel Scoring Algorithm generally replicated, rather than altered, youths' Scoring-as-Usual risk scores in relation to their peers.

Indices of absolute and relative model fit favored the Novel Scoring Algorithm (χ^2(1,874) = 3,072.35, p < 0.01, SRMR = 0.11, RMSEA = 0.03; CFI/TLI = 0.94), highlighting significant psychometric imprecision incurred by the Scoring-as-Usual method (χ^2(808) = 1,932.63, p < 0.01, SRMR = 0.17, RMSEA = 0.05; CFI/TLI = 0.83). However, differences in AUC estimates rendered by the two models were not statistically significant, indicating that the Novel Scoring Algorithm (AUC=0.64) holds no relative advantage over the Scoring-as-Usual method (AUC=0.65) in classifying youth as recidivant or desistant. AUC values derived from both the Novel Scoring Algorithm and the Scoring-as-Usual method were remarkably similar to average meta-analytic estimates for third generation juvenile risk assessment instruments (AUC=0.65, k=21, N=4,965) (Schwalbe, 2007). While both scoring methods predicted juvenile recidivism with expected levels of diagnostic accuracy, results ultimately do not support the full hypothesized advantages of the Novel Scoring Algorithm over the Scoring-as-Usual method.

Taken together, results ultimately affirm the Scoring-as-Usual method as an acceptable method of estimating likelihood of recidivism in court-involved youths. Nevertheless, the

magnitude and form of risk misclassification observed across gender, court division, and racial/ethnic cohorts highlight the penalties of juvenile risk assessment utilization on fair and equitable decision-making. Drawing from the factor structure of the Novel Scoring Algorithm, preliminary recommendations for risk management and measurement are discussed.

Patterns in Diagnostic Accuracy

In studies of prediction, AUC corresponds conceptually to the probability that a random score drawn from one sample (e.g., youth who recidivate) exceeds another score drawn from a separate sample (e.g., youth who desist) (Mossman, 1994; Swets et al., 2000; Rice & Harris, 2005). Following conversion procedures from Cohen's (1988) thresholds, AUC estimates of 0.55, 0.64, and 0.71 respectively correspond to small, moderate, and large effect sizes in violence risk assessment literature (Rice & Harris, 2005). Ergo, the AUC estimates yielded by the Novel Scoring Algorithm (AUC=0.64) and the Scoring-as-Usual method (AUC=0.65) correspond to a moderate effect size (Rice & Harris, 2005). As previously noted, these estimates additionally fall in line with meta-analytic estimates of both juvenile risk assessment performance (AUC=0.64) (Schwalbe, 2007), and adult risk assessment performance (AUC=0.67) (Gendreau et al., 1996). Taken together, both scoring methods predicted recidivism to an expected degree of overall diagnostic accuracy.

For the aggregated sample, model sensitivity (i.e., true positive rate) was 0.59 for the Novel Scoring Algorithm and 0.62 for the Scoring-as-Usual method, indicating that 59% and 62% of youth who recidivated were correctly identified as recidivist by the respective models. Concurrently, 1 – specificity (i.e., false positive rate) was 0.54 for the Novel Scoring Algorithm and 0.47 for the Scoring-as-Usual method, indicating that 54% and 47% of desistant youth were incorrectly identified as recidivist by the respective models. While the differences in overall

diagnostic accuracy between models were not statistically significant (see Table 9), these comparisons indicate that the Scoring-as-Usual yielded slightly more true positives and fewer false positives when compared to the Novel Scoring Algorithm. These results further affirm the Scoring-as-Usual method as the preferred method of estimating likelihood of recidivism in court-involved youths.

Cohort Comparisons. While results from the at-large sample affirm the Scoring-as-Usual method as a valid method of predicting recidivism, it is additionally important to investigate how certain subgroups (i.e., gender, racial/ethnic, court division cohorts) fare. Juvenile risk assessments were developed, in part, to alleviate discriminatory, paternalistic, and otherwise harmful biases incurred through discretionary court decision-making (Peck & Jennings, 2016). However, these standardized risk assessment instruments may conflate likelihood of recidivism with related characteristics of structural oppression (e.g., racism, poverty, trauma), justifying inappropriate treatment outcomes for marginalized youth (Green, 2007; Harcourt, 2010; Holtfreter & Morash, 2003).

Cohort comparisons of overall diagnostic accuracy (i.e., AUC) revealed no statistically significant gender, racial/ethnic, or court division differences between measurement models (see Tables 9 and 10). However, upon assessing forms of misclassification, several patterns emerged: composite risk scores overpredicted recidivism among boys, youth of color, and youth processed in the delinquency division of the court, regardless of the scoring method employed. Concurrently, composite risk scores underpredicted recidivism among girls, Caucasian/White youth, and youth processed in the truancy division. Both forms of risk misevaluation directly inhibit effective service delivery and diminish the likelihood of successful rehabilitation.

Previous research on gender in juvenile risk assessment contextualizes the observed differences between boys and girls. Many third-generation juvenile risk assessment instruments, including the original version of the YLS/CMI, were developed and validated in the 1990s, when girls represented approximately 1 in 5 juvenile court cases (Office of Juvenile Justice & Delinquency Prevention [OJJDP], 2019). While girls still account for far fewer arrests, in the time since then, they have become the fastest growing cohort in the juvenile justice system (OJJDP, 2019; Schwartz & Steffensmeier, 2012). Accordingly, some facets of measured risk on the YLS/CMI may be less sensitive to how girls present criminogenic risk. Research suggests that girls are often socialized into delinquency through distinct pathways (e.g., via intimate partner relationships), which are drawn out of focus or omitted entirely from generalist juvenile risk assessment instruments (Eklund et al., 2010; Kerig, 2014). Courts may benefit from utilizing gender-responsive evaluation approaches to estimate criminogenic risk more accurately in girls (Van Voohris et al., 2010).

Some feminist scholars have cautioned against general application of juvenile risk assessment, as features of non-criminogenic trauma may be incorrectly flagged as risk (Holtfreter & Morash, 2003). Girls may be acutely vulnerable to risk overprediction, given the elevated prevalence of previous trauma and victimization experiences (Hennessey et al., 2004). The present results were not compatible with this prior literature: girls' risk scores underpredicted their actual likelihood of recidivism, such that only 35% of those who recidivated were correctly identified. The low percentage of true positives among girls may instead reflect the court's failure to adequately address girls' criminogenic needs. Girls enter juvenile court supervision with qualitatively different risk profiles, as identified via initial juvenile risk assessment, with greater needs centered in familial and behavioral domains (Kitzmiller et al.,

2022). Effective court-sanctioned intervention for girls should generally be both: (1) minimally restrictive, given that most girls enter court supervision with low to moderate cumulative levels of risk; and (2) complementary to these distinct differences in types of needs. It is possible that the court is failing to provide effective intervention in one or both of these regards, thus increasing girls' actual criminogenic risk level over the course of court supervision (De La Rue & Ortega, 2019).

Finally, results indicate that juvenile risk assessment scores slightly overpredicted recidivism in boys, relative to the aggregated sample. As previously noted, juvenile risk assessments may be more attuned to typical features of criminogenic risk in boys. For example, several risk factors indirectly or directly identify externalizing behaviors (e.g., disruptive classroom behavior, explosive episodes, physical aggression), which are more commonly observed coping mechanisms in adolescent boys (Hoffmann & Su, 1998; Maschi et al., 2008). These externalizing behaviors, left unchecked, closely resemble delinquency (Maschi et al., 2008); however, they also represent relatively normative characteristics of psychosocial immaturity, which tend to digress naturally as youth enter late adolescence and early adulthood (Liu, 2004). While both boys and girls experience these normative psychosocial changes, it is perhaps less likely that girls would be flagged for externalizing characteristics of criminogenic risk at the onset of court supervision. In congruence with the current study's findings, measuring externalizing behaviors via juvenile risk assessment may provide well-reasoned impetus for referral to adjacent wraparound services; however, it may additionally contribute to overestimation of risk.

Results observed across court divisions closely resembles those across gender: namely, juvenile risk assessment scores overpredicted recidivism among delinquent youth, and

underpredicted recidivism for truant youth, regardless of the scoring method employed. It is important to note that gender and court division are intertwined: in the current sample, girls represent 50.79% of youth in the truancy division compared to 35.79% of youth in the delinquency division ($\chi^2(1)=33.11$, p<0.01). Previous research substantiates this pattern: while boys and girls commit status offenses (e.g., truancy) at comparable frequencies, girls are more likely than boys to fall under juvenile court jurisdiction for a status offense (Chesney-Lind & Sheldon, 2004; Onifade et al., 2010). Accordingly, it is possible that the courts' failure to adequately mitigate criminogenic risk in girls is reflected again in the rates of diagnostic accuracy for the truancy division.

Concurrently, while the YLS/CMI has made significant inroads in predicting repeat delinquent offending, less is known regarding its appropriateness in truancy specialty courts (Onifade et al., 2010). Truancy has increasingly been addressed through the juvenile court system, rather than the education system, in effort to stymie future criminogenic development more effectively (Baker et al., 2001; Onifade et al., 2010). In the current sample, youth processed via truancy division recidivated at the lowest rates (36.70%) relative to all other cohorts. Even so, their initials risk scores indicate that this recidivism rate is higher-than-expected. Thus, addressing truancy in a juvenile justice context may be ineffective and iatrogenic. This speculation is in line with other research highlighting the harmful effects of overprocessing low risk youth (Cecile & Born, 2009; Gatti et al., 2009). While the current study is not designed to examine the effects of truancy court specifically, future research should investigate whether truancy courts likewise represent a form of ineffective, and ultimately harmful, overprocessing.

Variation by race/ethnicity. Results observed across race/ethnicity underscore one of the central most criticisms of risk assessment: namely, that risk assessments forecast future justice

system contact, which is deeply informed by racism and other overlapping systems of oppression (Green, 2020; Hannah-Moffat et al., 2009; Maurutto & Hannah-Moffat, 2007). As a result, risk scores overpredicted recidivism among youth of color, justifying the continued over prescription of restrictive court sanctions to this cohort. While evidence of artificially high risk scores was observed in all youth of color, African American/Black youth appear to be acutely vulnerable to overprediction of recidivism: out of the 86 African American/Black youth who desisted, 61 were incorrectly identified as recidivist using both scoring methods. These findings are supported by previous research which note the unique effects of anti-Black racism on standardized juvenile risk assessment instruments (Miller et al., 2021).

Because many facets of measured risk appear to be linked to racial marginalization, juvenile risk assessment scores demonstrated exceedingly poor accuracy in correctly identifying recidivist Caucasian/White youth. Of the 60 Caucasian/White youth who recidivated, five were correctly identified using the Novel Scoring Algorithm and two were correctly identified using Scoring-as-Usual. The high prevalence of "false negatives" within this cohort may represent a missed opportunity to provide much needed intervention and wraparound services, as Caucasian/White youth with artificially low risk scores may re-enter their communities with unaddressed criminogenic needs.

Moving Towards Equitable Decision-Making

The widespread utilization of juvenile risk assessments reflects a growing effort towards implementing evidence-based evaluation and treatment standards in juvenile court settings (National Research Council, 2013; Singh et al., 2014; Vincent et al., 2012). However, the patterns of misclassification yielded from the Novel Scoring Algorithm and the Scoring-as-Usual method suggest that risk assessments provide may evidence-based justification for deeply

50

entrenched oppressive ideologies upheld through the justice system (Butcher & Kretschmar, 2020). One of the goals of this dissertation is to support equitable decision-making through improved juvenile risk assessment measurement. Ergo, eliminating risk assessment from juvenile justice administration altogether is likely not the appropriate solution; after all, risk assessments provide court practitioners with valuable and consistent information regarding youths' criminogenic risks and needs (Oleson et al., 2011; Peck & Jennings, 2016).

Some scholars posit that juvenile risk assessments can minimize contribution to systems-level inequity through the process of community norming. Community norming is the process by which an off-the-shelf risk assessment instrument is modified from its original form to improve performance, based upon local patterns of risk and recidivism (Lovins et al., 2018). While most criminal and juvenile justice agencies use off-the-shelf instruments without local norming or validation, some research suggests that juvenile risk assessments' performance is highly variable across jurisdictions (Wright et al., 1984). In response, experts posit that community norming via data mining and machine learning techniques will become hallmark characteristics as risk assessments enter their fifth generation (Duwe, 2014; Wormith, 2017).

While the process of community norming is not currently standardized, frequent steps include: (1) collecting responses from a large pool of potential items drawn existing off-the-shelf measures (Barnoski & Drake, 2007); (2) selecting items with strong predictive association with the outcome of interest via stepwise logistic regression (Austin & Tu, 2004; Hamilton et al., 2015); (3) weighting items appropriately based on predictive association (Hamilton et al., 2015); (4) conducting thorough review by subject matter experts; and (5) ensuring robustness to change overtime via cross-validation (Silver et al., 2000). While the process of community norming lies well beyond the scope of the current results, the factor model yielded from the Novel Scoring

Algorithm provides a promising launching point to refine risk measurement and management in court-involved youths.

Recommendations for Effective Risk Management

Prior to discussing the implications for effective risk management, it is firstly important to discuss the conceptual implications of the Novel Scoring Algorithm. Through CFA, the Novel Scoring Algorithm estimated latent criminogenic risk from the shared covariance among measured risk and protective factor items and domains. The Novel Scoring Algorithm yielded favorable absolute and relative model fit over the Scoring-as-Usual method, signaling that its parameters appropriately represent the observed covariance between item indicators. Accordingly, first-order factor loadings reflect shared covariance between a given assessment item and the other items within its domain. Assessment items with large first-order factor loadings have a stronger and more predictable "pull" on their constituents while those with small factor loadings have little bearing on other related facets of risk (Comrey & Lee, 1992). Taken together, the observed factor loadings generated by the Novel Scoring Algorithm can help court practitioners expedite risk reduction by centering areas with large factor loadings in treatment, while bringing items with low factor loadings out of focus.

Importance of Protective Factors. In six of the seven domains that include both risk and protective factors, the assessment items with the largest factor loadings were protective factors (Education: positive relationships with teachers (λ = -0.86); Peer Relations: close bonds with positive peers (λ = -0.86); Family & Parenting: strong family management (λ = -0.93); Attitudes & Orientation: prosocial attitudes (λ = -0.88); Personality & Behavior: low aggression (λ = -0.97); Substance Abuse: low availability to drugs (λ = -0.69)). Only within the Leisure &

Recreation domain was the assessment item with the largest factor loading a risk factor (i.e., could make better use of time ($\lambda = 0.87$)).

The presence of a protective factor denotes two similar, but distinct, pieces of information about a court-involved youth. First, they indicate that the youth does not have an unaddressed need in an area which may contribute to repeat offending. This inverse association with risk factors was clearly apparent in the present study, as evidenced by the strong, negative correlations observed between risk and protective factor items in Appendix B. Perhaps more importantly, protective factors indicate that the youth has an existing strength in an area that may play a role in their desistance from delinquency (Fergus & Zimmerman, 2005). The presence of a strength, coupled with the absence of a deficit, likely explains why protective factors were often the items with the largest "pull" on others within a risk domain. These results suggest that rehabilitative efforts are best devoted to cultivating new and existing strengths, rather than mitigating youths' deficits.

The importance of protective factors is substantiated by the growing development and implementation of strengths-based, restorative approaches to curriculum design and programming for court-involved youths. Protective factors identified via juvenile risk assessment provide a menu of youths' goals, capabilities, and assets which, in turn, can be incorporated into individually tailored treatment plans (Rennie & Dolan, 2010). Related studies have shown favorable effects of strengths-based programming on youths' self-efficacy and relationships with program staff (Akiva et al., 2017). While their effects on recidivism have yet to be systematically investigated, the current results provide preliminary support that strengths-based programming may additionally yield promising returns on criminogenic risk score reduction.

Identifying Extraneous Assessment Items. Concurrently, assessment items with low factor loadings should be brought out of focus from rehabilitative treatment, as they have weak association with other related facets of criminogenic risk. In concert with other community norming processes (e.g., incremental changes in predictive validity, cross-validation), items with low factor loadings may additionally be considered for removal from composite estimates of criminogenic risk. The results of the current study do not serve as conclusive evidence for assessment item removal; rather, this discussion contextualizes the following risk factors in the extant literature and weigh their implications for equitable decision-making. Given that juvenile risk assessments overpredicted recidivism among cohorts of youth that are already overrepresented in the juvenile justice system (e.g., boys, delinquent youth, and youth of color), identifying risk items which artificially raise composite scores is an issue of immediate importance.

Using Comrey and Lee's (1992) criteria, standardized factor loadings that fall below 0.38 are considered poor indicators of the specified latent construct. Results from the second-order CFA revealed that the following eight items fell below this threshold: three or more current convictions ($\lambda=0.30$), chronic alcohol use ($\lambda=0.27$), substance use linked to offense(s) ($\lambda=0.34$), poor relations with father ($\lambda=0.33$), poor relations with mother ($\lambda=0.34$), not seeking help ($\lambda=0.30$), inadequate guilt feelings ($\lambda=0.35$), and inflated self-esteem ($\lambda=0.09$).

Within the Prior/Current Offenses domain, three or more current convictions was endorsed in 9 (1.6%) cases; it is therefore not likely a frequent contributor to composite risk estimates. Nevertheless, many scholars contend that quantifying risk based using prior and current justice system involvement biases risk assessment tools against people of color (Harcourt, 2010). Regarding the item at hand, the number of convictions on a youths' current

docket reflects both their participation in delinquency and the decisions of justice officials (e.g., police decision to arrest, prosecutor decision to approve petitions) (Skeem & Lowenkamp, 2016). The weak factor loadings attributed to this item suggests that youth who had three or more current convictions did not necessarily have previous justice system contact. Therefore, by considering omission of three or more current convictions, the court could reduce the impact of differential selection on youths' risk scores without losing other related information on their previous justice system involvement.

Within the Substance Abuse domain, assessment items chronic alcohol use (N=16; 2.9%) and substance use linked to offense(s) (N=86; 15.4%) yielded weak factor loadings. Experimentation with alcohol is widely considered to be a common feature of adolescent risk taking, with little to no serious or long-term consequences (Bonomo et al., 2001). However, youth who consume alcohol with high frequency are more likely to report adverse outcomes concerning the justice system (e.g., trouble with police) and beyond (e.g., trouble at school or work, trips to the emergency room, trouble at home) (Colder et al., 2002). Concurrently, youth who consume alcohol chronically in early adolescence are more likely to perpetrate or be victimized by violence in adulthood (Popovici et al., 2012). The weak factor loading attributed to chronic alcohol use indicates that this characteristic does not predictably covary with other facets of measured risk pertaining to substance abuse; however, in accordance with the extant literature, chronic alcohol use in adolescents may signal elevated risk of future justice system contact, and may be an important indicator for referral to alcohol dependence treatment (Popovici et al., 2012).

The risk factor substance use linked to offense(s) diverges from the other items within the Substance Abuse domain, as it pertains to a characteristic of the offense, rather than the youths'

self-reported behavior. Quantifying risk based upon characteristics of the offense introduces opportunity for penalty based on differential selection; the endorsement criteria hinges upon a decision from justice officials to arrest and petition the youth for a substance use-related charge. Prior research indicates that self-reported rates of substance use are consistent across racial/ethnic cohorts (Rosenberg, 2018). Despite this, youth of color, particularly African American/Black youth, are disproportionately arrested, adjudicated, and incarcerated for substance use related charges (Rosenberg, 2018; Rovner, 2016). In the present sample, less than one quarter (22.43%) of youth who used substances occasionally or chronically met the criteria for substance use linked to offense(s), indicating that this assessment item has little bearing on youths' habitual substance usage, and thus provides little information on their need for substance use related treatment. Taken together, by considering omission of substance use linked to offense(s), the court could further reduce the impact of differential selection without losing other relevant information on youths' substance use tendencies.

Within the Family & Parenting domain, assessment items poor relations with mother (N=152; 27.2%) and poor relations with father (N=346; 61.9%) yielded weak factor loadings. A substantial body of literature holds that dysfunctional family environments can cause, sustain, or worsen adolescent delinquent involvement (Simons et al., 2005; Stern & Smith, 1995); in turn, mobilizing the family as a therapeutic influence is among the most common goals of juvenile court intervention (Buel, 2002; Diamond et al., 2011; Woolfenden et al., 2001; Woolfenden et al., 2002). However, estimating family risk through family configuration (e.g., relationships with biological parents) reflects the antiquated notion that so-called "broken homes" can be identified based upon kinship form (Parsons, 1943; Wells & Rankin, 1991). This assumption has been widely critiqued by race and gender scholars, who argue that, among other deficiencies, it fails to

account for support from extended family and community networks, a common feature in African American/Black communities(Collins, 1990; Love & Morris, 2019; Stack, 1974). Indeed, the results of the current study indicate that the relationship between the child and their biological parents holds little bearing on other measured components of family risk (e.g., inadequate supervision, inappropriate discipline, inconsistent parenting). Accordingly, the court may consider removing poor relations with mother and poor relations with father as indicators of family risk.

The three remaining assessment items include not seeking help (N=299; 53.5%), inadequate guilt feelings (N=91; 16.3%), and inflated self-esteem (N=32; 5.7%). These characteristics all pertain to attitudes, personality, and behavioral tendencies, which may be more difficult for court practitioners to accurately assess in on-the-job risk evaluations. While little is known on participants' experience of risk assessment specifically, participating in justice system procedures can be stressful and traumatizing for both youth (Branson et al., 2017; Ko et al., 2008; Pilnik & Kendall, 2012) and adults (Covington, 2022; Maschi et al., 2011). The resulting fear and confusion may further obscure youths' true personality traits and behavioral tendencies. It is also worth noting that probation officers have interpreted youths' behavior differently based upon race: report narratives of African American/Black youth were more likely to include descriptions of negative personality traits, while narratives of Caucasian/White youth were more likely to include descriptions of negative environmental influences (Bridges & Steen, 1998). Importantly, these three items are not exhaustive of all potentially difficult-to-assess personality and behavioral characteristics. However, their lack of predictable covariance with other facets of related risk flag them as potential areas to consider for omission.

Summary

It is critical that juvenile court processing decisions are appropriately tailored to youths' latent cumulative level of criminogenic risk to reduce recidivism. Aspiring to improve courts' measurement of risk, the present study compared two juvenile risk assessment measurement models: one derived from the unweighted sum score of all endorsed risk factors (i.e., Scoring-as-Usual method), and one weighted to correspond to a freely estimated factor model (i.e., Novel Scoring Algorithm). While the Novel Scoring Algorithm improved the overall fit of the data, composite risk estimates predicted recidivism with equivalent degrees of diagnostic accuracy to the Scoring-as-Usual method. Accordingly, these results endorse Scoring-as-Usual as an acceptable method of predicting recidivism.

Both measurement models yielded rates of diagnostic accuracy which fell in line with meta-analytic estimates for third generation juvenile risk assessment tools (Schwalbe, 2007). However, the form and magnitude of risk misclassification varied widely in accordance with demographic and charge-related characteristics: juvenile risk assessment scores overpredicted recidivism among boys, youth of color, and youth processed via delinquency division. Concurrently, juvenile risk assessment scores underpredicted recidivism among girls, Caucasian/White youth, and youth processed via truancy division.

While the current results cannot parse apart the mechanisms responsible for the divergent patterns in risk misclassification, it is possible that juvenile risk assessments may not be responsive to the characteristics which prime girls and status offenders to reoffending. Furthermore, ineffective court intervention may yield iatrogenic effects among these cohorts, rendering them more likely to recidivate upon court supervision exit than they were at entry. The overprediction of recidivism among boys, youth of color, and youth processed via delinquency

suggests that certain characteristics of measured risk may correspond to non-criminogenic features of adolescent developments and consequences of structural racism. These findings highlight an urgent need to critically examine juvenile risk assessment items and eliminate those with marginal implications for risk and recidivism.

While the process of community norming lies well beyond the scope of the current study, parameter estimates yielded by the Novel Scoring Algorithm serve as an optimal launching point for this work. Specifically, estimates indicate that leveraging new and existing protective factors in juvenile programming and case management may be the most efficient means of expedient risk reduction. Additionally, the following eight risk factors yielded marginal covariance with other related indicators of risk: three or more current convictions (λ=0.30), chronic alcohol use (λ=0.27), substance use linked to offense(s) (λ=0.34), poor relations with father (λ=0.33), poor relations with mother (λ=0.34), not seeking help (λ=0.30), inadequate guilt feelings (λ=0.35), and inflated self-esteem (λ=0.09). In concert with other community norming procedures (e.g., cross-validation, stepwise logistic regression), these items may be considered for removal to reduce artificially high composite risk scores.

Strengths & Limitations

The results of the current study are bolstered by several strengths. First, the measurement of criminogenic risk and recidivism is highly ecologically valid, as the data collected represents official juvenile risk assessment and recidivism records retained by court practitioners in the field. Relatedly, the assessment instrument utilized (i.e., the YLS/CMI) is among the most the widely adopted actuarial juvenile risk assessment tools in juvenile court settings. Taken together, the findings have immediate implications towards understanding and refining local measurement

of criminogenic risk among court-involved youth. Additionally, given the popularity of the YLS/CMI, findings create opportunity for cross-validation in different settings.

Concurrently, the current study employs a novel methodological approach to measuring criminogenic risk via juvenile risk assessment. Sum scoring is among the most common method of estimating a variable of interest that is not directly measurable (e.g., criminogenic risk) (Bauer & Curran, 2015). However, sum scoring may be insufficient depending on the context and the stakes involved (McNeish & Wolf, 2020). The current study is the first of its kind to weigh the tradeoffs in psychometric precision and predictive validity incurred by sum scoring in juvenile risk assessment. Results affirm that sum scoring yields no detriment in estimating youths' likelihood of recidivism when compared to a freely estimated factor model. Accordingly, results serve as a necessary robustness check on a near-universally utilized method of estimating composite criminogenic risk.

Despite these strengths, findings from the current study are tempered by several methodological and theoretical shortcomings. The data collected represents patterns in risk assessment scores and recidivism from a single juvenile circuit court jurisdiction. Utilizing a single county sample optimizes the study's responsivity to the local ecology of delinquency, and therefore maximizes the relevance of implications on court practices. However, patterns in risk assessment scores and recidivism vary widely by geography (Feld, 1991); thus, the parameter estimates yielded by the Novel Scoring Algorithm and their implications for recidivism are not generalizable beyond the single county sample. Future research drawing from additional court jurisdictions is warranted to rigorously evaluate the factor structure and diagnostic performance of the YLS/CMI.

Results are further tempered by premising prediction of recidivism solely on youths' initial juvenile risk assessment score. Functionally, the initial risk assessment score is analogous to a court's first impression of a newly adjudicated youth, and therefore has the greatest influence over processing and treatment decisions. However, many facets of criminogenic risk in adolescents are subject to change over time. For instance, scores among those initially classified as high risk may decline over the period of court supervision, either naturally or in response to effective intervention. Likewise, scores among youths initially classified as low risk may increase over the period of court supervision, sometimes in response to iatrogenic court responses. In any case, these initial risk score may not align with the youths' true likelihood of recidivism at the end of their period of court supervision. In the current study, misclassification caused by fluxuations in criminogenic risk over time was indistinguishable from measurement error in the risk assessment tool. Future research may disentangle the confounding effect of risk score fluxuations by instead premising prediction of recidivism on youths' final risk assessment score.

Finally, the comparison of diagnostic performance across sample cohorts was limited in its lack of intersectional scope. Youth hold multiple identities, each of which may have compounding or contradicting implications for risk misevaluation. For example, results indicate that African American/Black girls in the delinquency division are simultaneously vulnerable to artificially low and high composite risk scores. It is likely, therefore, that discussions of race/ethnicity, gender, and court division are overly simplistic, and obscure heterogeneous within-cohort patterns. Future research should consider replicating analyses with larger samples, allowing intersectional cohort comparisons to be drawn.

Conclusions

Over the last few decades, courts have increasingly relied upon juvenile risk assessments to inform case processing and treatment decisions (JJGPS, 2020). Results of the current study both affirm and challenge their continued use. First, findings suggest that the Scoring-as-Usual method, the near-universally implemented procedure for calculating composite risk, predicts recidivism with acceptable levels of diagnostic accuracy, based upon disciplinary standards (Rice & Harris, 2005). Concurrently, findings highlight distinctly different patterns in risk misevaluation based upon youths' demographic and charge characteristics, suggesting that risk scores provide evidence-based justification for deeply entrenched oppressive ideologies upheld through the justice system. Importantly, this central criticism of risk assessment reflects system-level inequities and will likely persist without systems-level change. Nonetheless, the present results provide preliminary evidence that courts may be able to reduce immediate harms by leveraging protective factors in case management, while drawing other extraneous facets of risk out of focus.

APPENDICES

Appendix A: Frequency of YLS and PFRJR Item Endorsement

Table 10.
Frequency of YLS and PFRJR item endorsement.

Assessment Item	n	Frequency of Endorsement	
		Endorsed (%)	Not Endorsed (%)
Prior/Current Offenses			
Three or more prior convictions	559	7 (1.3%)	552 (98.7%)
Two or more prior failures to comply	559	21 (3.8%)	538 (96.2%)
Prior probation	559	51 (9.1%)	508 (90.9%)
Prior custody	559	37 (6.6%)	552 (93.4%)
Three or more current convictions	559	9 (1.6%)	550 (98.4%)
Education			
Low achievement	559	453 (81.0%)	106 (19.0%)
Problems with teachers	559	222 (39.7%)	337 (60.3%)
Problems with peers	559	253 (45.3%)	306 (54.7%)
Disruptive classroom behavior	559	264 (47.2%)	295 (52.8%)
Disruptive behavior on school property	559	338 (60.5%)	221 (39.5%)
Truancy	559	424 (75.8%)	135 (24.2%)
Passing*	559	83 (14.8%)	476 (85.2%)
High achievement*	559	28 (5.0%)	531 (95.0%)
Positive relationships with teachers*	559	214 (38.3%)	345 (61.7%)
Commitment to school/education*	559	229 (41.0%)	330 (59.0%)
Leisure & Recreation			
Lack of organized activities	559	387 (69.2%)	172 (30.8%)
Could make better use of time	559	454 (81.2%)	105 (18.8%)
No personal interests	559	51 (9.1%)	508 (90.9%)
Involvement in organized activities*	559	162 (29.0%)	397 (71.0%)
Positive personal interests*	559	337 (60.3%)	222 (39.7%)
Religiosity*	559	156 (27.9%)	403 (72.1%)
Peer Relations			
Lack of positive peer acquaintances	559	209 (37.4%)	350 (62.6%)
Lack of positive friends	559	242 (43.3%)	317 (56.7%)
Some delinquent peer acquaintances	559	425 (76.0%)	134 (24.0%)
Some delinquent friends	559	344 (61.5%)	215 (38.5%)
Close bonds with positive peers*	559	168 (30.1%)	391 (69.9%)
Substance Abuse			
Occasional drug use	559	365 (65.3%)	194 (34.7%)
Chronic drug use	559	187 (33.5%)	372 (66.5%)
Chronic alcohol use	559	16 (2.9%)	543 (97.1%)
Substance abuse interferes with life	559	152 (27.2%)	407 (72.8%)
Substance use linked to offense(s)	559	86 (15.4%)	473 (84.6%)
Low availability to drugs*	559	130 (23.3%)	429 (76.7%)
Actively abstaining from drugs/alcohol*	559	214 (38.3%)	345 (61.7%)
Family & Parenting			
Inadequate supervision	559	227 (40.6%)	332 (59.4%)
Difficulty in controlling behavior	559	348 (62.3%)	211 (37.7%)
Inappropriate discipline	559	256 (45.8%)	303 (54.2%)
Inconsistent parenting	559	235 (42.0%)	324 (58.0%)
Poor relations with father	559	346 (61.9%)	213 (38.1%)
Poor relations with mother	559	152 (27.2%)	407 (72.8%)

Table 10 (cont'd).

Consistent supervision*	559	203 (36.3%)	356 (63.7%)
Strong family management*	559	151 (27.0%)	408 (73.0%)
Consistent parenting*	559	111 (27.3%)	294 (72.4%)
Strong adult bonds*	559	313 (56.0%)	246 (44.0%)
Attitudes & Orientation			
Not seeking help	559	299 (53.5%)	260 (46.5%)
Actively rejecting help	559	74 (13.2%)	485 (86.8%)
Defies authority	559	67 (12.0%)	492 (88.0%)
Antisocial/pro-criminal attitudes	559	178 (31.8%)	381 (68.2%)
Callous, little concern for others	559	76 (13.6%)	483 (86.4%)
Actively seeking help*	559	79 (19.5%)	279 (81.3%)
Positive response to authority*	559	185 (33.1%)	374 (66.9%)
Prosocial attitudes*	559	133 (23.8%)	426 (76.2%)
Personality & Behavior			
Short attention span	559	341 (61.0%)	218 (39.0%)
Poor frustration tolerance	559	416 (74.4%)	143 (25.6%)
Verbally aggressive/intimidating	559	364 (65.1%)	195 (34.9%)
Explosive episodes	559	263 (47.0%)	296 (53.0%)
Physically aggressive	559	264 (47.2%)	295 (52.8%)
Inadequate guilt feelings	559	91 (16.3%)	468 (83.7%)
Inflated self-esteem	559	32 (5.7%)	527 (94.3%)
Low aggression*	559	135 (24.2%)	424 (75.8%)
Strong social skills*	559	112 (21.8%)	437 (78.2%)
Community			
Perceived safety*	559	397 (71.0%)	162 (29.0%)
Access to resources*	559	343 (61.4%)	216 (38.6%)
Positive adults*	559	275 (49.2%)	284 (50.8%)

*Denotes that item is a protective factor.

Appendix B: Correlations Between YLS and PFRJR Assessment Items

Table 11.
Correlations between YLS and PFRJR assessment items.

Prior/Current Offenses

	1	2	3	4	5
1. Three or more prior convictions	--				
2. Two or more failures to comply	.32*	--			
3. Prior probation	.30*	.36*	--		
4. Prior custody	.23*	.36*	.57*	--	
5. Three or more current convictions	.11*	.05	.01	.02	--

Education

	1	2	3	4	5	6	7	8	9	10
1. Low achievement	--									
2. Problems with teachers	.16*	--								
3. Problems with peers	.12*	.33*	--							
4. Disruptive classroom behavior	.10*	.57*	.36*	--						
5. Disruptive behavior on school property	.10*	.27*	.35*	.37*	--					
6. Truancy	.29*	.06	.02	.03	.01	--				
7. Passing	-.66*	-.13*	-.13*	-.17*	-.14*	-.28*	--			
8. High achievement	-.39*	-.10*	-.08	-.12*	-.10*	-.20*	.48*	--		
9. Positive relationships with teachers	-.27*	-.37*	-.17*	-.30*	-.16*	-.23*	.31*	.22*	--	
10. Commitment to school/education	-.24*	-.19*	-.07	-.14*	-.11*	-.33*	.34*	.24*	.45*	--

Leisure & Recreation

	1	2	3	4	5	6
1. Lack of organized activities	--					
2. Could make better use of time	.41*	--				
3. No personal interests	.10*	.09*	--			
4. Involvement in organized activities	-.84*	-.45*	-.12*	--		
5. Positive personal interests	-.22*	-.22*	-.30*	.28*	--	
6. Religiosity	-.24*	-.15*	-.09*	.26*	.18*	--

Peer Relations

	1	2	3	4	5
1. Lack of positive peer acquaintances	--				
2. Lack of positive friends	.59*	--			
3. Some delinquent peer acquaintances	.17*	.20*	--		
4. Some delinquent friends	.22*	.27*	.64*	--	
5. Close bonds with positive peers	-.35*	-.48*	-.27*	-.28*	--

Substance Abuse

	1	2	3	4	5	6	7
1. Occasional drug use	--						
2. Chronic drug use	.48*	--					
3. Chronic alcohol use	.08	.15*	--				
4. Substance abuse interferes with life	.39*	.50*	.14*	--			
5. Substance use linked to offense(s)	.25*	.32*	.17*	.27*	--		

Table 11 (cont'd).

	Substance Abuse						
	1	2	3	4	5	6	7
6. Low availability to drugs	-.60*	-.38*	-.09*	-.29*	-.22*	--	
7. Actively abstaining from drugs/alcohol	-.52*	-.45*	-.11*	-.37*	-.22*	.49*	--

	Family & Parenting									
	1	2	3	4	5	6	7	8	9	10
1. Inadequate supervision	--									
2. Difficulty in controlling behavior	.28*	--								
3. Inappropriate discipline	.21*	.38*	--							
4. Inconsistent parenting	.24*	.27*	.48*	--						
5. Poor relations with father	.09*	.14*	.15*	.08	--					
6. Poor relations with mother	.05	.24*	.17*	.12	.02	--				
7. Consistent supervision	-.58*	-.41*	-.18*	-.19*	-.12*	-.18*	--			
8. Strong family management	-.16*	-.54*	-.34*	-.31*	-.23*	-.17*	.40*	--		
9. Consistent parenting	-.17*	-.44*	-.37*	-.42*	-.21*	-.11*	.39*	.63*	--	
10. Strong adult bonds	-.03	-.32*	-.10*	-.10*	-.07	-.18*	.25*	.42*	.30*	--

	Attitudes & Orientation							
	1	2	3	4	5	6	7	8
1. Not seeking help	--							
2. Actively rejecting help	.13*	--						
3. Defies authority	.04	.13*	--					
4. Antisocial/pro-criminal attitudes	.04	.18*	.23*	--				
5. Callous, little concern for others	.07	.11*	.16*	.24*	--			
6. Actively seeking help	-.50*	-.16*	-.12*	-.09*	-.09*	--		
7. Positive response to authority	-.18*	-.17*	-.26*	-.20*	-.10*	.22*	--	
8. Prosocial attitudes	-.15*	-.16*	-.15*	-.34*	-.10*	.24*	.49*	--

	Personality & Behavior								
	1	2	3	4	5	6	7	8	9
1. Short attention span	--								
2. Poor frustration tolerance	.25*	--							
3. Verbally aggressive/intimidating	.23*	.47*	--						
4. Explosive episodes	.18*	.41*	.43*	--					
5. Physically aggressive	.13*	.32*	.37*	.33*	--				
6. Inadequate guilt feelings	.08*	.08	.08	.02	.17*	--			
7. Inflated self esteem	-.01	.04	.10*	.14*	<.01	.04	--		
8. Low aggression	-.24*	-.55*	-.58*	-.49*	-.46*	-.07	-.03	--	
9. Strong social skills	-.24*	-.37*	-.27*	-.25*	-.20*	-.08	<.01	.42*	--

Table 11 (cont'd).

	Community		
	1	2	3
1. Perceived safety	--		
2. Access to resources	.47*	--	
3. Positive adults	.41*	.41*	--

*$p < 0.05$

Appendix C: Summary of the Novel Scoring Algorithm

Table 12.
Summary of the Novel Scoring Algorithm.

First-Order Factor Loadings			
	Unstd. Est. (*S.E.*)	Std. Est. (*S.E.*)	p
Prior/Current Offenses BY			
Three or more prior convictions	1.00 (*.00*)	.98 (*.13*)	.00
Two or more prior failures to comply	.82 (*.14*)	.80 (*.08*)	.00
Prior probation	.87 (*.14*)	.85 (*.07*)	.00
Prior custody	.97 (*.16*)	.95 (*.07*)	.00
Three or more current convictions	.31 (*.18*)	.30 (*.18*)	.09
Education BY			
Low achievement	1.00 (*.00*)	.57 (*.06*)	.00
Problems with teachers	1.03 (*.13*)	.58 (*.05*)	.00
Problems with peers	.87 (*.12*)	.49 (*.05*)	.00
Disruptive classroom behavior	.97 (*.13*)	.55 (*.05*)	.00
Disruptive behavior on school property	1.00 (*.14*)	.56 (*.05*)	.00
Truancy	.89 (*.12*)	.51 (*.06*)	.00
Passing*	-1.31 (*.11*)	-.74 (*.05*)	.00
High achievement*	-1.32 (*.16*)	-.75 (*.06*)	.00
Positive relationships with teachers*	-1.53 (*.16*)	-.86 (*.03*)	.00
Commitment to school/education*	-1.39 (*.15*)	-.78 (*.04*)	.00
Leisure & Recreation BY			
Lack of organized activities	1.00 (*.00*)	.69 (*.05*)	.00
Could make better use of time	1.26 (*.11*)	.87 (*.05*)	.00
No personal interests	.58 (*.11*)	.40 (*.07*)	.00
Involvement in organized activities*	-1.07 (*.05*)	-.74 (*.04*)	.00
Positive personal interests*	-1.08 (*.11*)	-.74 (*.05*)	.00
Religiosity*	-.61 (*.11*)	-.42 (*.07*)	.00
Peer Relations BY			
Lack of positive peer acquaintances	1.00 (*.00*)	.65 (*.04*)	.00
Lack of positive friends	1.03 (*.07*)	.67 (*.04*)	.00
Some delinquent peer acquaintances	1.05 (*.10*)	.68 (*.05*)	.00
Some delinquent friends	1.05 (*.09*)	.68 (*.04*)	.00
Close bonds with positive peers*	-1.33 (*.09*)	-.86 (*.03*)	.00
Substance Abuse BY			
Occasional drug use	1.00 (*.00*)	.59 (*.03*)	.00
Chronic drug use	1.07 (*.06*)	.63 (*.03*)	.00
Chronic alcohol use	.45 (*.12*)	.27 (*.10*)	.00
Substance abuse interferes with life	.89 (*.06*)	.53 (*.04*)	.00
Substance use linked to offense(s)	.57 (*.08*)	.34 (*.06*)	.00
Low availability to drugs*	-1.17 (*.06*)	-.69 (*.03*)	.00
Actively abstaining from drugs/alcohol*	-1.02 (*.06*)	-.60 (*.03*)	.00
Family & Parenting BY			
Inadequate supervision	1.00 (*.00*)	.49 (*.05*)	.00
Difficulty in controlling behavior	1.75 (*.19*)	.86 (*.03*)	.00
Inappropriate discipline	1.22 (*.15*)	.60 (*.04*)	.00
Inconsistent parenting	1.20 (*.15*)	.59 (*.04*)	.00

Table 12 (cont'd).

First-Order Factor Loadings			
	Unstd. Est. (*S.E.*)	Std. Est. (*S.E.*)	p
Family & Parenting BY			
Poor relations with father	.67 (*.13*)	.33 (*.06*)	.00
Poor relations with mother	.70 (*.14*)	.34 (*.06*)	.00
Consistent supervision*	-1.47 (*.12*)	-.72 (*.04*)	.00
Strong family management*	-1.88 (*.20*)	-.93 (*.02*)	.00
Consistent parenting*	-1.82 (*.20*)	-.90 (*.03*)	.00
Strong adult bonds*	-1.23 (*.16*)	-.61 (*.04*)	.00
Attitudes & Orientation BY			
Not seeking help	1.00 (*.00*)	.30 (*.06*)	.00
Actively rejecting help	1.55 (*.32*)	.47 (*.06*)	.00
Defies authority	2.13 (*.44*)	.64 (*.05*)	.00
Antisocial/pro-criminal attitudes	2.02 (*.41*)	.61 (*.04*)	.00
Callous, little concern for others	1.56 (*.36*)	.47 (*.07*)	.00
Actively seeking help*	-1.44 (*.25*)	-.44 (*.06*)	.00
Positive response to authority*	-2.78 (*.53*)	-.84 (*.03*)	.00
Prosocial attitudes*	-2.92 (*.57*)	-.88 (*.03*)	.00
Personality & Behavior BY			
Short attention span	1.00 (*.00*)	.41 (*.06*)	.00
Poor frustration tolerance	1.91 (*.29*)	.79 (*.04*)	.00
Verbally aggressive/intimidating	1.99 (*.30*)	.82 (*.03*)	.00
Explosive episodes	1.69 (*.27*)	.70 (*.04*)	.00
Physically aggressive	1.61 (*.26*)	.67 (*.04*)	.00
Inadequate guilt feelings	.86 (*.23*)	.35 (*.08*)	.00
Inflated self-esteem	.23 (*.26*)	.09 (*.11*)	.40
Low aggression*	-2.35 (*.35*)	-.97 (*.03*)	.00
Strong social skills*	-2.19 (*.33*)	-.90 (*.05*)	.00
Community BY			
Perceived safety*	1.00 (*.00*)	.68 (*.05*)	.00
Access to resources*	1.13 (*.13*)	.77 (*.05*)	.00
Positive adults*	1.40 (*.15*)	.96 (*.06*)	.00
Second-Order Factor Loadings			
	Unstd. Est.	Std. Est.	p
Criminogenic Risk BY			
Prior/Current Offenses	1.00 (*.00*)	.35 (*.05*)	.00
Education	1.43 (*.32*)	.86 (*.02*)	.00
Leisure & Recreation	1.53 (*.32*)	.75 (*.04*)	.00
Peer Relations	1.81 (*.35*)	.95 (*.03*)	.00
Substance Abuse	1.80 (*.36*)	.72 (*.03*)	.00
Family & Parenting	1.25 (*.27*)	.86 (*.02*)	.00
Attitudes & Orientation	.81 (*.22*)	.91 (*.03*)	.00
Personality & Behavior	.84 (*.21*)	.69 (*.03*)	.00
Community	-1.01 (*.23*)	-.50 (*.05*)	.00

*Denotes that item is a protective factor.

Appendix D: Summary of the Scoring-as-Usual Method

Table 13.

Summary of the Scoring-as-Usual method.

First-Order Factor Loadings				
	Std. Est	Unstd. Est.	S.E.	p
Prior/Current Offenses BY				
Three or more prior convictions	.98	1.00	.00	999.00
Two or more prior failures to comply	.80	1.00	.00	999.00
Prior probation	.85	1.00	.00	999.00
Prior custody	.95	1.00	.00	999.00
Three or more current convictions	.30	1.00	.00	999.00
Education BY				
Low achievement	.57	1.00	.00	999.00
Problems with teachers	.58	1.00	.00	999.00
Problems with peers	.49	1.00	.00	999.00
Disruptive classroom behavior	.55	1.00	.00	999.00
Disruptive behavior on school property	.56	1.00	.00	999.00
Truancy	.51	1.00	.00	999.00
Leisure & Recreation BY				
Lack of organized activities	.69	1.00	.00	999.00
Could make better use of time	.87	1.00	.00	999.00
No personal interests	.40	1.00	.00	999.00
Peer Relations BY				
Lack of positive peer acquaintances	.65	1.00	.00	999.00
Lack of positive friends	.67	1.00	.00	999.00
Some delinquent peer acquaintances	.68	1.00	.00	999.00
Some delinquent friends	.68	1.00	.00	999.00
Substance Abuse BY				
Occasional drug use	.85	1.00	.00	999.00
Chronic drug use	.91	1.00	.00	999.00
Chronic alcohol use	.39	1.00	.00	999.00
Substance abuse interferes with life	.76	1.00	.00	999.00
Substance use linked to offense(s)	.49	1.00	.00	999.00
Family & Parenting BY				
Inadequate supervision	.49	1.00	.00	999.00
Difficulty in controlling behavior	.86	1.00	.00	999.00
Inappropriate discipline	.60	1.00	.00	999.00
Inconsistent parenting	.59	1.00	.00	999.00
Poor relations with father	.33	1.00	.00	999.00
Poor relations with mother	.34	1.00	.00	999.00
Attitudes & Orientation BY				
Not seeking help	.30	1.00	.00	999.00
Actively rejecting help	.47	1.00	.00	999.00
Defies authority	.64	1.00	.00	999.00
Antisocial/pro-criminal attitudes	.61	1.00	.00	999.00
Callous, little concern for others	.47	1.00	.00	999.00
Personality & Behavior BY				
Short attention span	.41	1.00	.00	999.00
Poor frustration tolerance	.79	1.00	.00	999.00

Table 13 (cont'd).

First-Order Factor Loadings				
	Std. Est	Unstd. Est.	S.E.	p
Verbally aggressive/intimidating	.82	1.00	.00	999.00
Explosive episodes	.70	1.00	.00	999.00
Physically aggressive	.67	1.00	.00	999.00
Inadequate guilt feelings	.35	1.00	.00	999.00
Inflated self-esteem	.09	1.00	.00	999.00
Second-Order Factor Loadings				
	Std. Est.	Unstd. Est.	S.E.	p
Criminogenic Risk BY				
Prior/Current Offenses	.35	1.00	.00	999.00
Education	.86	1.00	.00	999.00
Leisure & Recreation	.75	1.00	.00	999.00
Peer Relations	.95	1.00	.00	999.00
Substance Abuse	.72	1.00	.00	999.00
Family & Parenting	.86	1.00	.00	999.00
Attitudes & Orientation	.91	1.00	.00	999.00
Personality & Behavior	.69	1.00	.00	999.00

*Denotes that item is a protective factor.

REFERENCES

REFERENCES

Akiva, T., Li, J., Martin, K. M., Horner, C. G., & McNamara, A. R. (2017). Simple interactions: Piloting a strengths-based and interaction-based professional development intervention for out-of-school time programs. *Child & Youth Care Forum, 46*(3), 285-305.

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

Andrews, D. A., Kiessling, J. J., Mickus, S., & Robinson, D. (1986). The construct validity of interview-based risk assessment in corrections. *Canadian Journal of Behavioral Science*, *18*(4), 460.

Austin, P. C., & Tu, J. V. (2004). Bootstrap methods for developing predictive models. *The American Statistician, 58*(2), 131-137.

Bailey, Z. D., Krieger, N., Agénor, M., Graves, J., Linos, N., & Bassett, M. T. (2017). Structural racism and health inequities in the USA: Evidence and interventions. *The Lancet, 389*(10077), 1453–1463. https://doi.org/10.1016/S0140-6736(17)30569-X

Baker, M. L., Sigmon, J. N., & Nugent, M. E. (2001). Truancy Reduction: Keeping Students in School. Juvenile Justice Bulletin.

Barnes-Lee, A. R. (2020). Development of protective factors for reducing juvenile reoffending: a strengths-based approach to risk assessment. *Criminal Justice and Behavior*, *47*(11), 1371-1389.

Barnes-Lee, A. R., & Campbell, C. A. (2020). Protective factors for reducing juvenile reoffending: an examination of incremental and differential predictive validity. *Criminal Justice and Behavior*, *47*(11), 1390-1408.

Barnoski, R., & Drake, E. (2007). Washington's Offender Accountability Act: Department of Correction's static risk instrument. Washington State Institute for Public Policy.

Bauer, D., & Curran, P. (2015). The discrepancy between measurement and modeling in longitudinal data analysis. *Advances in multilevel modeling for educational research: Addressing practical issues found in real-world applications*, 3-38.

Beaulac, J., Bouchard, D., & Kristjansson, E. (2009). Physical activity for adolescents living in a disadvantaged neighbourhood: Views of parents and adolescents on needs, barriers, facilitators, and programming. *Leisure/Loisir*, *33*(2), 537-561.

Belisle, L. A., & Salisbury, E. J. (2021). Starting with girls and their resilience in mind: Reconsidering risk/needs assessments for system-involved girls. *Criminal Justice and Behavior*, *48*(5), 596-616.

Birckhead, T. R. (2012). Delinquent by reason of poverty. *Wash. UJL & Pol'y*, *38*, 53.

Bishop, D. M., & Frazier, C. E. (1995). Race effects in juvenile justice decision-making: Findings of a statewide analysis. *Journal of Criminal Law & Criminology, 86,* 392.

Blackwell, B. S., Holleran, D., & Finn, M. A. (2008). The impact of the Pennsylvania sentencing guidelines on sex differences in sentencing. *Journal of Contemporary Criminal Justice*, *24*(4), 399-418.

Bonomo, Y., Coffey, C., Wolfe, R., Lynskey, M., Bowes, G., & Patton, G. (2001). Adverse outcomes of alcohol use in adolescents. *Addiction*, *96*(10), 1485-1496.

Bonta, J., & Andrews, D. A. (2007). Risk-need-responsivity model for offender assessment and rehabilitation. *Rehabilitation*, *6*(1), 1-22.

Bonilla-Silva, E. (1997). Rethinking racism: Toward a structural interpretation. *American Sociological Review, 62*(3), 465–480. https://doi.org/10.2307/2657316

Bortner, M. A., & Wornie, L. R. (1985). The preeminence of process: An example of refocused justice research. *Social Science Quarterly*, *66*(2), 413.

Branson, C. E., Baetz, C. L., Horwitz, S. M., & Hoagwood, K. E. (2017). Trauma-informed juvenile justice systems: A systematic review of definitions and core components. *Psychological Trauma: Theory, Research, Practice, and Policy, 9*(6), 635.

Bridges, G. S., & Steen, S. (1998). Racial disparities in official assessments of juvenile offenders: Attributional stereotypes as mediating mechanisms. *American Sociological Review*, *63*(4), 554–570. https://doi.org/10.2307/2657267

Bronfenbrenner, U. (1979). *The ecology of human development*. Harvard university press.

Butcher, F., Kretschmar, J. M., Lin, Y., Flannery, D. J., & Singer, M. I. (2014). Analysis of the validity scales in the trauma symptom checklist for children. *Research on Social Work Practice*, *24*(6), 695-704.

Cangur, S., & Ercan, I. (2015). Comparison of model fit indices used in structural equation modeling under multivariate normality. *Journal of Modern Applied Statistical Methods*, *14*(1), 14.

Cauffman, E., Cavanagh, C., Donley, S., & Thomas, A. G. (2016). A developmental perspective on adolescent risk-taking and criminal behavior. *The Handbook of Criminological Theory*, 100-120.

Cauffman, E., & Steinberg, L. (2000). (Im)maturity of judgment in adolescence: Why adolescents may be less culpable than adults. *Behavioral Sciences & the Law, 18*, 741-760.

Cécile, M., & Born, M. (2009). Intervention in juvenile delinquency: Danger of iatrogenic effects?. *Children and Youth Services Review*, *31*(12), 1217-1221.

Chesney-Lind, M. (1977). Judicial paternalism and the female status offender: Training women to know their place. *Crime & Delinquency*, *23*(2), 121-130.

Chesney-Lind, M., & Sheldon, R. G. (2004). Young women, delinquency and juvenile justice.

Colder, C. R., Campbell, R. T., Ruel, E., Richardson, J. L., & Flay, B. R. (2002). A finite mixture model of growth trajectories of adolescent alcohol use: predictors and consequences. *Journal of Consulting and Clinical Psychology*, *70*(4), 976.

Collins, P. H. (1990). *Black feminist thought: Knowledge, consciousness, and the politics of empowerment.* New York: Routledge.

Comrey, L. A., & Lee, H. B. (1992). *A first course in factor analysis (2nd ed.).* Hillside, NJ: Lawrence Erlbaum Associates.

Covington, S. (2022). Creating a trauma-informed justice system for women. *Wiley handbook on what works with female offenders*.

Crew, B. K. (1991). Sex differences in criminal sentencing: Chivalry or patriarchy?.

Daly, K. (1994). *Gender, crime, and punishment*. Yale University Press.

Development Services Group, Inc. (2015). Risk and needs assessment for youths. Washington, D. C.: Office of Juvenile Justice and Delinquency Prevention. Available at https://www.ojjdp.gov/mpg/litreviews/RiskandNeeds.pdf

DeJong, C., & Jackson, K. C. (1998). Putting race into context: Race, juvenile justice processing, and urbanization. *Justice Quarterly*, *15*(3), 487-504.

De La Rue, L., & Ortega, L. (2019). Intersectional trauma-responsive care: A framework for humanizing care for justice involved girls and women of color. *Journal of Aggression, Maltreatment & Trauma*, *28*(4), 502-517.

de Vogel, V., de Vries Robbé, M., de Ruiter, C., & Bouman, Y. H. (2011). Assessing protective factors in forensic psychiatric practice: Introducing the SAPROF. International Journal of Forensic Mental Health, 10(3), 171–177.

Diamond, B., Morris, R. G., & Caudill, J. W. (2011). Sustaining families, dissuading crime: The effectiveness of a family preservation program with male delinquents. *Journal of Criminal Justice, 39*(4), 338-343.

Draelos, R. (2019, February 23). Measuring performance: AUC (AUROC). *Glass Box*. Available at https://glassboxmedicine.com/2019/02/23/measuring-performance-auc-auroc/

Duwe, G. (2014). The development, validity, and reliability of the Minnesota screening tool assessing recidivism risk (MnSTARR). *Criminal Justice Policy Review*, *25*(5), 579-613.

Eklund, J. M., Kerr, M., & Stattin, H. (2010). Romantic relationships and delinquent behaviour in adolescence: The moderating role of delinquency propensity. *Journal of Adolescence*, *33*(3), 377-386.

Erez, E. (1992). Dangerous men, evil women: Gender and parole decision-making. *Justice Quarterly*, *9*(1), 105-126.

Feinstein, R. (2015). A qualitative analysis of police interactions and disproportionate minority contact. *Journal of Ethnicity in Criminal Justice*, *13*(2), 159-178.

Feld, B. C. (1991). Justice by geography: Urban, suburban, and rural variations in juvenile justice administration. *Journal of Criminal Law & Criminology*, *82*, 156.

Fergus, S., & Zimmerman, M. A. (2005). Adolescent resilience: A framework for understanding healthy development in the face of risk. *Annual Review of Public Health*, *26*, 399-419.

Fountain, E. N., & Mahmoudi, D. (2021). Mapping juvenile justice: Identifying existing structural barriers to accessing probation services. *American Journal of Community Psychology*, *67*(1-2), 116-129.

Gatti, U., Tremblay, R. E., & Vitaro, F. (2009). Iatrogenic effect of juvenile justice. *Journal of Child Psychology and Psychiatry*, *50*(8), 991-998.

Gendreau, P., Little, T., & Goggin, C. (1996). A meta-analysis of the predictors of adult offender recidivism: What works!. *Criminology*, *34*(4), 575-608.

Glover, K. S. (2008). Citizenship, hyper-surveillance, and double-consciousness: Racial profiling as panoptic governance. In *Surveillance and governance: Crime control and beyond*. Emerald Group Publishing Limited.

Gorman-Smith, D., Tolan, P. H., Zelli, A., & Huesmann, L. R. (1996). The relation of family functioning to violence among inner-city minority youths. *Journal of Family Psychology*, *10*(2), 115.

Graybeal, C. (2001). Strengths-based social work assessment: Transforming the dominant paradigm. *Families in society*, *82*(3), 233-242.

Green, B. (2020, January). The false promise of risk assessments: epistemic reform and the limits of fairness. In *Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency* (pp. 594-606).

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, M. (2015). Risk-needs assessment: Constitutional and ethical challenges. *American Criminal Law Review*, *52*, 231.

Harcourt, B. E. (2010, September). Risk as a proxy for race. University of Chicago Law & Economics Olin Working Paper 535; University of Chicago Public Law Working Paper 323. Retrieved from http://ssrn.com/ abstract1677654

Harris, P. (2006). What community supervision officers need to know about actuarial risk assessment and clinical judgment. *Federal Probation, 70*(2), 8-14.

Hawkins, J. D., Van Horn, M. L., & Arthur, M. W. (2004). Community variation in risk and protective factors and substance use outcomes. *Prevention Science*, *5*(4), 213-220.

Hennessey, M., Ford, J. D., Mahoney, K., Ko, S. J., & Siegfried, C. B. (2004). Trauma among girls in the juvenile justice system. *Los Angeles, CA: National Child Traumatic Stress Network.*

Hoffmann, J. P., & Su, S. S. (1998). Stressful life events and adolescent substance use and depression: Conditional and gender differentiated effects. *Substance Use & Misuse*, *33*(11), 2219-2262.

Hoge, R. D. (2020). The Youth level of service/Case management inventory. In *Handbook of violence risk assessment* (pp. 191-205). Routledge.

Hoge, R. D., Andrews, D. A., & Leschied, A. W. (1996). An investigation of risk and protective factors in a sample of youthful offenders. *Journal of Child Psychology and psychiatry*, *37*(4), 419-424.

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

Holtfreter, K., & Morash, M. (2003). The needs of women offenders. *Women & Criminal Justice*, *14*(2-3), 137-160.

Homan, P. (2019). Structural sexism and health in the United States: A new perspective on health inequality and the gender system. *American Sociological Review*, *84*(3), 486-516.

Howell, J. C., & Hawkins, J. D. (1998). Prevention of youth violence. *Crime and justice*, *24*, 263-315.

Hubbard, D. J., & Matthews, B. (2008). Reconciling the differences between the "gender-responsive" and the "what works" literatures to improve services for girls. *Crime & Delinquency*, *54*(2), 225-258.

Jacobs, L. A., Ashcraft, L. E., Sewall, C. J., Folb, B. L., & Mair, C. (2020). Ecologies of juvenile reoffending: a systematic review of risk factors. *Journal of criminal justice*, *66*, 101638.

Javdani, S., & Allen, N. E. (2016). An ecological model for intervention for juvenile justice-involved girls: Development and preliminary prospective evaluation. *Feminist Criminology*, *11*(2), 135-162.

Jones, C. P. (2000). Levels of racism: A theoretic framework and a gardener's tale. *American Journal of Public Health, 90*(8), 1212–1215.

Juvenile Justice Geography, Policy, Practice & Statistics. (2020). Juvenile justice services: Risk assessment. *National Center for Juvenile Justice.* Available: http://www.jjgps.org/juvenile-justice-services#risk-assessment

Kerig, P. K. & Becker, S. P. (2012). Trauma and girls' delinquency. *Delinquent Girls*, 119-143.

Kerig, P. K. (2014). Introduction: for better or worse: intimate relationships as sources of risk or resilience for girls' delinquency. *Journal of Research on Adolescence*, *24*(1), 1-11.

Kitzmiller, M. K., Hoskins, K., & Cavanagh, C. (2022). Examining Sex-Based Measurement Invariance in the Youth Level of Service/Case Management Inventory. *Crime & Delinquency*, 00111287211073677.

Ko, S. J., Ford, J. D., Kassam-Adams, N., Berkowitz, S. J., Wilson, C., Wong, M., ... & Layne, C. M. (2008). Creating trauma-informed systems: Child welfare, education, first responders, health care, juvenile justice. *Professional psychology: Research and practice*, *39*(4), 396.

Kline, R. B. (2016). *Principles and Practices of Structural Equation Modeling* (4th ed.). The Guilford Press: New York, NY.

Kuhn, D. (2009). Adolescent thinking. In R.M. Lerner & L. Steinberg (Eds.), Handbook of Adolescent Psychology. Hoboken, NJ: Wiley & Sons.

Kumpfer, K. L., & Turner, C. W. (1990). The social ecology model of adolescent substance abuse: Implications for prevention. *International journal of the addictions*, *25*(sup4), 435-463.

Leve, L. D., & Chamberlain, P. (2005). Association with delinquent peers: Intervention effects for youth in the juvenile justice system. *Journal of abnormal child psychology*, *33*(3), 339-34.

Liberman, A., & Fontaine, J. (2015). *Reducing harms to boys and young men of color from criminal justice system involvement*. Washington, DC: Urban institute.

Littlefield, A.K., Sher, K.J., & Steinley, D. (2010). Developmental trajectories of impulsivity and their association with alcohol use and related outcomes during emerging and young adulthood. Alcoholism: Clinical and Experimental Research, 34(4), 1409–1416. doi: 10.1111/j.1530"0277.2010.01224.x.

Liu, J. (2004). Childhood externalizing behavior: Theory and implications. *Journal of Child and Adolescent Psychiatric Nursing*, *17*(3), 93-103.

Long, J., & Sullivan, C. J. (2017). Learning more from evaluation of justice interventions: Further consideration of theoretical mechanisms in juvenile drug courts. *Crime & Delinquency*, *63*(9), 1091-1115.

Love, T. P., & Morris, E. W. (2019). Opportunities diverted: intake diversion and institutionalized racial disadvantage in the juvenile justice system. *Race and Social Problems*, *11*(1), 33-44.

Lovins, B. K., Latessa, E. J., May, T., & Lux, J. (2018). Validating the Ohio risk assessment system community supervision tool with a diverse sample from Texas. *Corrections*, *3*(3), 186-202.

Mandrekar, J. N. (2010). Receiver operating characteristic curve in diagnostic test assessment. *Journal of Thoracic Oncology*, *5*(9), 1315-1316.

Maschi, T., Morgen, K., Bradley, C., & Hatcher, S. S. (2008). Exploring gender differences on internalizing and externalizing behavior among maltreated youth: Implications for social work action. *Child and Adolescent Social Work Journal*, *25*(6), 531-547.

Maschi, T., Gibson, S., Zgoba, K. M., & Morgen, K. (2011). Trauma and life event stressors among young and older adult prisoners. *Journal of Correctional Health Care*, 17(2), 160-172.

McBride, D. C., & McCoy, C. B. (1981). Crime and drug-using behavior: An areal analysis. *Criminology*, *19*(2), 281-302.

McCarter, S. A. (2016). Holistic representation: A randomized pilot study of wraparound services for first-time juvenile offenders to improve functioning, decrease motions for review, and lower recidivism. *Family Court Review, 54*(2), 250-260.

McGarrell, E. F. (1993). Trends in racial disproportionality in juvenile court processing: 1985-1989. *Crime & Delinquency*, *39*(1), 29-48.

McNeish, D., & Wolf, M. G. (2020). Thinking twice about sum scores. *Behavior Research Methods, 52*, 2287-2305.

McNeish, D., & Wolf, M. G. (2020). Dynamic fit index cutoffs for Confirmatory Factor Analysis models.

Miller, W. T., Campbell, C. A., Papp, J., & Ruhland, E. (2021). The contribution of static and dynamic factors to recidivism prediction for Black and White youth offenders. *International journal of offender therapy and comparative criminology*, 0306624X211022673.

Miron, M., Tolan, S., Gómez, E., & Castillo, C. (2021). Evaluating causes of algorithmic bias in juvenile criminal recidivism. *Artificial Intelligence and Law*, *29*(2), 111-147.

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

Moon, S. S., Patton, J., & Rao, U. (2010). An ecological approach to understanding youth violence: The mediating role of substance use. *Journal of human behavior in the social environment*, *20*(7), 839-856.

Mossman, D. (1994). Assessing predictions of violence: Being accurate about accuracy. *Journal of Consulting and Clinical Psychology, 62*(4), 783.

Muthén, L. K., & Muthén, B. O. (1998-2020). Mplus User's Guide. Sixth Edition. Los Angeles, CA: Muthén & Muthén.

National Research Council (NAS). (2013). *Reforming juvenile justice: A developmental approach*. Washington, D. C.: The National Academics Press.

Nissen, L. (2006). Bringing strength-based philosophy to life in juvenile justice. *Reclaiming Children and Youth*, *15*(1), 40.

Nunn, K. B. (2001). The child as other: Race and differential treatment in the juvenile justice system. *DePaul L. Rev.*, *51*, 679.

O'Brien, L., Albert, D., Chein, J., & Steinberg, L. (2011). Adolescents prefer more immediate rewards when in the presence of their peers. Journal of Research on Adolescence, 21(4), 747–753. doi: 10.1111/j.1532"7795.2011.00738.

Office of Juvenile Justice & Delinquency Prevention. (2019). Estimated number of arrests by offense and age group. *OJJDP Statistical Briefing Book*. Available: https://www.ojjdp.gov/ojstatbb/crime/ucr.asp?table_in=1

Oleson, J. C., VanBenschoten, S. W., Robinson, C. R., & Lowenkamp, C. T. (2011). Training to see risk: Measuring the accuracy of clinical and actuarial risk assessments among federal probation officers. *Fed. Probation*, *75*, 52.

Onifade, E., Smith Nyandoro, A., Davidson, W. S., & Campbell, C. (2010). Truancy and patterns of criminogenic risk in a young offender population. Youth violence and juvenile justice, 8(1), 3-18.

Outland, R. (2021). Why Black and Brown Youth Fear and Distrust Police: An Exploration of Youth Killed by Police in the US (2016/2017), Implications for Counselors and Service Providers. *Open Journal of Social Sciences*, *9*(04), 222.

Parsons, T. (1943). The kinship system of the contemporary United States. *American anthropologist*, *45*(1), 22-38.

Peck, J. H., & Jennings, W. G. (2016). A critical examination of "being Black" in the juvenile justice system. *Law and Human Behavior*, *40*(3), 219.

Pilnik, L., & Kendall, J. R. (2012). Victimization and trauma experienced by children and youth: implications for legal advocates.

Piquero, A. R. (2008). Disproportionate minority contact. *The future of children*, 59-79.

Popovici, I., Homer, J. F., Fang, H., & French, M. T. (2012). Alcohol use and crime: findings from a longitudinal sample of US adolescents and young adults. *Alcoholism: Clinical and Experimental Research*, *36*(3), 532-543.

Powell, J. A. (2007). Structural racism: Building upon the insights of John Calmore- A tribute to John O. Calmore's work. *North Carolina Law Review, 86*(3), 791–816.

Pullmann, M. D., Kerbs, J., Koroloff, N., Veach-White, E., Gaylor, R., & Sieler, D. (2006). Juvenile offenders with mental health needs: Reducing recidivism using wraparound. *Crime & Delinquency*, *52*(3), 375-397.

Rennie, C. E., & Dolan, M. C. (2010). The significance of protective factors in the assessment of risk. *Criminal Behaviour and Mental Health*, *20*(1), 8-22.

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*(5), 615-620.

Roesch, R. (1988). Community psychology and the law. *American Journal of Community Psychology*, *16*(4), 451-463.

Rosenberg, L. (2018) Community services for mental illnesses and substance use disorders: The moral test of our time. *The Journal of Behavioral Health Services & Research*, 45(2), 157-159.

Rucker, J. M., & Richeson, J. A. (2021). Toward an understanding of structural racism: Implications for criminal justice. *Science*, *374*(6565), 286-290.

Schwalbe, C. S. (2007). Risk assessment for juvenile justice: A meta-analysis. *Law and human behavior*, *31*(5), 449.

Sebastian, C., Viding, E. Williams, K.D., & Blakemore, S.J. (2010). Social brain development and the affective consequences of ostracism in adolescence. Brain and Cognition, 72, 134–145.

Shields, I. W., & Simourd, D. J. (1991). Predicting predatory behavior in a population of incarcerated young offenders. *Criminal Justice and Behavior*, *18*(2), 180-194.

Silver, E., Smith, W. R., & Banks, S. (2000). Constructing actuarial devices for predicting recidivism: A comparison of methods. *Criminal Justice and Behavior, 27*(6), 733-764.

Simourd, D. J., Hoge, R. D., Andrews, D. A., & Leschied, A. W. (1994). An empirically-based typology of male young offenders. *Canadian Journal of Criminology*, *36*(4), 447-461.

Singh, P. S. J., & Azman, A. (2020). Dealing with Juvenile Delinquency: Integrated Social Work Approach. *Asian Social Work Journal*, *5*(2), 32-43.

Singh, J. P., Desmarais, S. L., Hurducas, C., Arbach-Lucioni, K., Condemarin, C., Dean, K., Otto, R. K. (2014). International Perspectives on the practical application of violence risk assessment: A global survey of 44 countries. International Journal of Forensic Mental Health, 13, 193– 206. http://dx.doi.org/10.1080/14999013.2014.922141

Skeem, J. L., & Lowenkamp, C. T. (2016). Risk, race, and recidivism: Predictive bias and disparate impact. *Criminology*, *54*(4), 680-712.

Spivak, A. L., Wagner, B. M., Whitmer, J. M., & Charish, C. L. (2014). Gender and status offending: Judicial paternalism in juvenile justice processing. *Feminist Criminology*, *9*(3), 224-248.

Spohn, C. (1999). Gender and sentencing of drug offenders: Is chivalry dead? *Criminal Justice Policy Review, 9*(3-4), 365-399.

Schwartz, J., & Steffensmeier, D. (2012). Stability and change in girls' delinquency and the gender gap: Trends in violence and alcohol offending across multiple sources of evidence. In *Delinquent Girls* (pp. 3-23). Springer, New York, NY.

Simons, R. L., Simons, L. G., Burt, C. H., Brody, G. H., & Cutrona, C. (2005). Collective efficacy, authoritative parenting and delinquency: A longitudinal test of a model integrating community-and family-level processes. *Criminology*, *43*(4), 989-1029.

Stack, C. (1974). *All our kin: Strategies for survivor in a Black community*. New York: Harper and Row.

Steinberg, L., Cauffman, E., & Monahan, K. C. (2015). Psychosocial maturity and desistance from crime in a sample of serious juvenile offenders. *OJJDP Juvenile Justice Bulletin.*

Stern, S. B., & Smith, C. A. (1995). Family processes and delinquency in an ecological context. *Social Service Review*, *69*(4), 703-731.

St. John, V., Murphy, K., & Liberman, A. (2020). Recommendations for addressing racial bias in risk and needs assessment in the juvenile justice system. *Child Trends.*

Swets, J. A., Dawes, R. M., & Monahan, J. (2000). Psychological science can improve diagnostic decisions. *Psychological Science in the Public Interest*, *1*(1), 1-26.

Takane, Y., & De Leeuw, J. (1987). On the relationship between item response theory and factor analysis of discretized variables. *Psychometrika*, *52*(3), 393-408.

Tarter, R. E., Kirisci, L., Vanyukov, M., Cornelius, J., Pajer, K., Shoal, G. D., & Giancola, P. R. (2002). Predicting adolescent violence: impact of family history, substance use, psychiatric history, and social adjustment. *American journal of psychiatry*, *159*(9), 1541-1547.

Thomas, C. W., & Sieverdes, C. M. (1975) Juvenile court intake: An analysis of discretionary decision-making. *Criminology, 12*(4), 413-43.

Tucker Sr, R. B. (2014). The color of mass incarceration. *Ethnic Studies Review*, *37*(1), 135-149.

Van Voorhis, P., Wright, E. M., Salisbury, E., & Bauman, A. (2010). Women's risk factors and their contributions to existing risk/needs assessment: The current status of a gender-responsive supplement. *Criminal Justice and Behavior*, *37*(3), 261-288.

Viljoen, J. L., Jonnson, M. R., Cochrane, D. M., Vargen, L. M., & Vincent, G. M. (2019). Impact of risk assessment instruments on rates of pretrial detention, postconviction placements, and release: A systematic review and meta-analysis. Law and Human Behavior, 43(5), 397–420. https://doi.org/10.1037/lhb0000344

Vincent, G. M., Guy, L. S., & Grisso, T. (2012). Risk assessment in juvenile justice: A guidebook for implementation.

Vincent, G. M., & Viljoen, J. L. (2020). Racist Algorithms or Systemic Problems? Risk Assessments and Racial Disparities. *Criminal Justice and Behavior*, 0093854820954501.

Ward, T., & Brown, M. (2004). The good lives model and conceptual issues in offender rehabilitation. Psychology, Crime & Law, 10(3), 243–257.

Wells, L. E., & Rankin, J. H. (1991). Families and delinquency: A meta-analysis of the impact of broken homes. *Social Problems, 38*(1), 71-93.

Williams, D. R., & Mohammed, S. A. (2013). Racism and health I: Pathways and scientific evidence. *American Behavioral Scientist, 57*(8), 1152–1173. https://doi.org/10.1177/0002764213487340

Windle, M. (2000). A latent growth curve model of delinquent activity among adolescents. *Applied Developmental Science*, *4*(4), 193-207.

Woolfenden, S., Williams, K. J., & Peat, J. (2001). Family and parenting interventions in children and adolescents with conduct disorder and delinquency aged 10-17. *Cochrane Database of Systematic Reviews*, (2).

Woolfenden, S. R., Williams, K., & Peat, J. K. (2002). Family and parenting interventions for conduct disorder and delinquency: a meta-analysis of randomized controlled trials. *Archives of disease in childhood*, *86*(4), 251-256.

Wordes, M., Bynum, T. S., & Corley, C. J. (1994). Locking up youth: The impact of race on detention decisions. *Journal of research in crime and delinquency*, *31*(2), 149-165.

Wormith, J. S. (2017). Automated offender risk assessment. *Criminology & Public Policy*, *16*, 281.

Wright, K. N., Clear, T. R., & Dickson, P. (1984). Universal Applicability of Probation Risk-Assessment Instruments: A Critique. *Criminology*, *22*(1), 113-134.

Zane, S. N., & Pupo, J. A. (2021). Disproportionate Minority Contact in the Juvenile Justice System: A Systematic Review and Meta-Analysis. *Justice Quarterly*, 1-26.