

The Influence of Risk Assessment Evidence on Judicial Sentencing Decisions

by
Melissa Jonnson

B.A. (Hons.), University of British Columbia (Okanagan), 2015

Thesis Submitted in Partial Fulfillment of the
Requirements for the Degree of
Master of Arts

in the
Department of Psychology
Faculty of Arts and Social Sciences

© Melissa Jonnson 2018
SIMON FRASER UNIVERSITY
Fall 2018

Copyright in this work rests with the author. Please ensure that any reproduction or re-use is done in accordance with the relevant national copyright legislation.

Approval

Name: **Melissa Jonnson**

Degree: **Master of Arts (Clinical Psychology)**

Title: **The Influence of Risk Assessment Evidence on
Judicial Sentencing Decisions**

Examining Committee: **Chair:** V. Gordon Rose
Senior Lecturer

Jodi Viljoen
Senior Supervisor
Associate Professor

Deborah Connolly
Supervisor
Professor

David MacAlister
External Examiner
Associate Professor
School of Criminology

Date Defended/Approved: December 11, 2018

Ethics Statement

The author, whose name appears on the title page of this work, has obtained, for the research described in this work, either:

- a. human research ethics approval from the Simon Fraser University Office of Research Ethics

or

- b. advance approval of the animal care protocol from the University Animal Care Committee of Simon Fraser University

or has conducted the research

- c. as a co-investigator, collaborator, or research assistant in a research project approved in advance.

A copy of the approval letter has been filed with the Theses Office of the University Library at the time of submission of this thesis or project.

The original application for approval and letter of approval are filed with the relevant offices. Inquiries may be directed to those authorities.

Simon Fraser University Library
Burnaby, British Columbia, Canada

Update Spring 2016

Abstract

Despite the increased presence of risk assessment evidence in sentencing proceedings, its role has been contentious, and judges' use of this evidence is unclear. This study examines judges' opinions about risk assessment tools and assesses the influence of tools on judicial sentencing decisions. Specifically, it uses an experimental vignette design to evaluate whether judges' impressions of a hypothetical offender or their sentencing recommendations are affected by the presence of an empirically-supported tool and accompanying risk rating. Participants were 170 judges from 34 American states and two Canadian provinces. The most common advantage of tools highlighted by judges was that they provide information that is helpful for making decisions, while the most common concern was their potential to obscure bias. Judges' impressions and placement recommendations were not significantly influenced by tool presence. Judges' program recommendations, however, were more intensive and consistent for the high-risk offender when a tool was present.

Keywords: risk assessment tools; judicial opinions; restrictive sanctions; evidence-based sentencing; mixed methods design

Dedication

I dedicate this thesis to my mother, Pamela Jonnson.

Mom, you have supported me in so many ways, from our daily conversations to being by my side in times of need. Over the years, you have taught me not only how to face my fears and reach for the stars, but also how to stop and nurture myself along the way. I am forever grateful for your wisdom and compassion, which continue to be a guiding light in my life.

“In your light I learn how to love. In your beauty, how to make poems. You dance inside my chest where no-one sees you, but sometimes I do, and that sight becomes this art.”

— Rumi

Acknowledgements

First, to my senior supervisor, Dr. Jodi Viljoen, thank you for your insight, guidance, and support over the past two years. The time and care you devote to your students is truly remarkable, and I have grown so much throughout the course of this program because of you. I am extremely grateful for your supervision and look forward to pursuing more goals with you in the years to come!

Thank you also to my committee members, Drs. Deborah Connolly and David MacAlister for helping to shape this project. I genuinely appreciate your perceptive feedback and encouragement. To Dr. Gina Vincent and Robert Batting, Q.C., thank you for your constructive suggestions during the pilot-testing phase of this project.

Thank you to the Social Sciences and Humanities Research Council of Canada, the American Psychology–Law Association, the International Association of Forensic and Mental Health Services, and Simon Fraser University for helping to fund this research. I am grateful for the opportunities you have facilitated.

Of course, this project would not have been possible without the judges who participated in it, so thank you all for your valuable contributions. I have enjoyed reading your comments and learning more about your views on forensic risk assessment.

I would also like to thank my cohort and lab mates, who continue to make this journey so memorable. You are so thoughtful and inspiring, and I am grateful for all of you!

To my parents, Pam and Owen, who have sacrificed so much for me, and my brother, Levi Jonnson, who has always been my Mjöltnir, I would not be who or where I am today without your steady love and support. Thank you for all that you have done throughout my life and the life of this project—you mean the world to me!

Last, but certainly not least, thank you to my partner, Ryan Phillips. You are so generous and hard-working in ways I could never be, and I appreciate the many late nights and weekends you have sacrificed to help me realize this goal. Thank you for all your conversations, hugs, meals, and pep talks. In such a vast universe, I am so happy to share this speck of space and time with you!!

Table of Contents

Approval.....	ii
Ethics Statement.....	iii
Abstract.....	iv
Dedication.....	v
Acknowledgements.....	vi
Table of Contents.....	vii
List of Tables.....	ix
List of Figures.....	x
Chapter 1. Introduction.....	1
Advantages of Using Risk Assessment Tools in Sentencing.....	2
Concerns with Using Risk Assessment Tools in Sentencing.....	4
Judges' Opinions and Use of Risk Assessment Tools in Sentencing.....	6
The Present Study.....	9
Chapter 2. Methods.....	12
Sample Requirements.....	12
Sample Characteristics.....	12
Procedure.....	13
Materials.....	14
Opinions about the purposes of sentencing.....	14
Opinions about risk assessment tools.....	15
Vignette.....	15
Impressions of the offender.....	16
Sentencing propensities.....	17
Demographics questionnaire.....	17
Analytic Plan.....	17
Descriptive.....	17
Research Question 1.....	18
Research Question 2.....	19
Research Question 3.....	20
Research Question 4.....	20
Research Question 5.....	21
Chapter 3. Results.....	22
Descriptive Statistics.....	22
Manipulation checks.....	23
Research Question 1: What are judges' opinions about using risk assessment tools to make sentencing decisions?.....	23
Benefits of using tools.....	27
Concerns with using tools.....	29
Circumstances that influence the usefulness of tools.....	32
Research Question 2: Do risk assessment tools moderate associations between an offender's risk level and judges' impressions of the offender?.....	33

Research Question 3: Do risk assessment tools influence judges' sentencing recommendations?.....	36
Placement restrictiveness.....	38
Program intensity.....	40
Research Question 4: Does the effect of a tool on judges' sentencing recommendations depend on how much emphasis judges think tools should be given?	42
Research Question 5: Do risk assessment tools increase the consistency of judges' sentencing recommendations?	43
Placement recommendations	43
Program recommendations	44
Chapter 4. Discussion	45
Judges had diverse opinions about using risk assessment tools to make sentencing decisions	45
Risk assessment tools did not amplify associations between offender risk level and judges' impressions of the offender	47
Risk assessment tools influenced program, but not placement, recommendations	48
Risk assessment tools increased the consistency of program, but not placement, recommendations.....	50
Strengths and limitations.....	51
Implications	53
Conclusion	54
References.....	56
Appendix A. Survey.....	64
Appendix B. Pilot testing questions	74

List of Tables

Table 1.	Sample Sizes Required to Detect Specific Effect Sizes	12
Table 2.	Summary of 2 x 2 Vignette Design.....	14
Table 3.	Psychometric Properties of Variables Related to Judges' Opinions about Risk Assessment Tools	25
Table 4.	Ordinal Logistic Regression Results for the Main Effect of Risk Level on Judges' Impressions of the Offender.....	35
Table 5.	Frequency Table Displaying the Number of Judges in Each Condition Who Recommended Each Placement Type.....	37
Table 6.	Frequency Table Displaying the Number of Judges in Each Condition Who Recommended Each Program Intensity	37
Table 7.	Consistency in Judges' Placement Recommendations According to Vignette Condition	44
Table 8.	Consistency in Judges' Program Recommendations According to Vignette Condition	44

List of Figures

Figure 1.	Tree diagram depicting the categories and subcategories that emerged from the conventional content analysis.	28
Figure 2.	Bar graph showing the percentage of judges in each group who recommended a restrictive placement.....	39
Figure 3.	Mosaic plots showing the effect of risk level on judges' placement recommendations when a risk assessment tool was present or absent.	39
Figure 4.	Bar graph showing the percentage of judges in each group who recommended intensive programming.	41
Figure 5.	Mosaic plots showing the effect of risk level on judges' program recommendations when a risk assessment tool was present or absent.	41
Figure 6.	Bar graph showing the percentage of judges in each group who recommended a restrictive sentence for the high-risk offender.....	43

Chapter 1.

Introduction

Using risk assessment tools to guide decisions about criminal sanctioning has become common practice in both the adult and juvenile justice systems over the past several decades (DeMatteo, Wolbransky, & LaDuke, 2016; Monahan & Skeem, 2016; Singh et al., 2014). Recent estimates suggest that courts across the United States use risk assessment tools to inform sentencing decisions “in some or all cases” (Starr, 2014, p. 809) and, in some states, sentencing guidelines require judges to consider risk assessment evidence in their decisions (Monahan & Skeem, 2016). Likewise, risk assessment instruments have permeated correctional realms across Canada, Europe, and other parts of the world (Bonta, 2007). Despite the increased presence of risk assessment evidence in sentencing proceedings, its role in the sentencing process has been contentious and research on judges’ use of this evidence has been scarce. Consequently, the present study aims to explore how risk assessment tools influence judges’ sentencing practices.

Risk assessment tools include a set of risk factors that have been linked to recidivism (e.g., substance use issues, employment problems, history of violence) and are typically used to summarize an offender’s overall risk level (i.e., risk to reoffend; Skeem & Monahan, 2011). Some tools involve combining item scores algorithmically to produce a total score (i.e., actuarial prediction), while others involve weighing the relevance and importance of each item in a structured manner to produce a summary risk statement (i.e., structured professional judgement). Prior to the development of such tools, risk levels were estimated subjectively using unstructured clinical judgement. Although certain tools estimate risk more precisely than others, research repeatedly demonstrates that empirically-supported tools outperform unstructured clinical judgement in predicting future antisocial behaviour (Ægisdóttir et al., 2006; Monahan & Skeem, 2016). Having established the predictive value of risk assessment tools, experts are now urging researchers to shift their focus to the implications of using these tools for risk management and reduction (Peterson-Badali, Skilling, & Haqanee, 2015; Skeem & Monahan, 2011).

Since judges play a prominent role in determining supervision levels and other risk management strategies for offenders, understanding how risk assessment tools affect their decisions is pertinent. Surprisingly, however, research on judges' use of risk assessment evidence is rare and many mental health professionals who use risk assessment tools to make legal recommendations do not know whether these recommendations are implemented into practice (Singh et al., 2014). Most existing studies have focused on system-wide changes following the implementation of a tool (e.g., changes in incarceration rates) and have been conducted by non-academic organizations involved in the design or implementation of the tool (Viljoen, Jonnson, Cochrane, & Vargen, 2018). Moreover, these studies have been fraught with confounds, such that it is difficult to gauge whether effects should be attributed to the tool or to other factors such as historical trends (Viljoen et al., 2018). As such, the impact of risk assessment tools on sentencing practices is still unclear.

Although risk assessment evidence is generally recognized as being theoretically and empirically useful for assisting judges in making evidence-based sentencing decisions (Bonta, 2007; Elek, Warren, & Casey, 2015; Imrey & Dawid 2015; Monahan & Skeem, 2014, 2016), concerns have been raised about whether pre-sentence risk estimates might introduce a more insidious form of bias and exacerbate economic disparities in sentencing (Chanenson & Hyatt, 2016; Cole, 2007; Hannah-Moffat, 2011; Maurutto & Hannah-Moffat, 2007; Monahan & Skeem, 2014). In the following sections, I will summarize the primary arguments for and against the inclusion of risk assessment evidence in sentencing proceedings.

Advantages of Using Risk Assessment Tools in Sentencing

Due to their increased predictive accuracy and objectivity over unstructured clinical opinion, some scholars have posited that risk assessment tools are essential to evidence-based penalty (Bonta, 2007; Thompson, 2017). Indeed, standardized risk assessment tools could conceivably benefit criminal sentencing practices by increasing consistency and transparency in decision-making, since they require clinicians and legal decision-makers to consider the same inventory of risk-related factors for each offender being adjudicated (Skeem, 2013). For these reasons, using empirically-supported tools to inform sentencing decisions appears more appropriate than using unstructured clinical judgement.

In addition to improving predictions and providing more structure to decisions, tools may assist with resource allocation. Given that resources to manage risk (e.g., secure facilities, supervision, treatment) are limited, it is important to effectively identify offenders who would most benefit from them. Research consistently demonstrates that higher-risk offenders are the ones most likely to benefit from intensive risk management strategies (Andrews & Dowden, 2006; Bonta, 2007; Nolan & Stewart, 2017). Further, lower-risk offenders who are given excessive treatment and supervision are more likely to be exposed to high-risk offenders, which may increase their likelihood of reoffending (Latessa, Lovins, & Smith, 2010; Nolan & Stewart, 2017). Consequently, the risk-need-responsivity (RNR) model—a widely-utilized framework for managing and reducing recidivism risk—suggests that offenders who represent the highest risk should receive higher levels of supervision and more intensive programming, while those who represent a lesser risk should receive lower levels of supervision and less intensive programming (Andrews, Bonta, & Hoge, 1990).

Structured risk assessment tools are presumed to improve adherence to the RNR model by identifying risk levels for case planning. In the context of sentencing, risk assessment tools might assist judges in determining the intensity of supervision and programming that is most appropriate, based on the RNR model. Starr (2014) designed an experiment to assess the influence of tool-based risk ratings (i.e., “low” versus “moderate-to-high” risk) on criminal law students’ sentence recommendations for a hypothetical offender. Students who read a vignette with a risk rating recommended longer sentences for the higher-risk defendant and shorter sentences for the lower-risk defendant, than those who read a vignette without a risk rating. These findings provide some evidence that tool-based risk ratings may improve adherence to the RNR model, although studies with judges are needed.

Although their impact on RNR adherence has not been determined, tools have been heralded as a solution to overincarceration because of their potential to help divert lower risk offenders away from custodial channels and into community alternatives (e.g., supervised release, diversion programs; Laura & John Arnold Foundation, 2016; Vincent, Guy, Perrault, & Gershenson, 2016). As noted previously, some studies have examined large-scale changes in adjudication outcomes following justice reform initiatives that included the implementation of a structured risk assessment tool (e.g., Annie E. Casey Foundation, 2017; Maloney & Miller, 2015). A recent systematic review

of such studies found that, overall, incarceration rates were lower in the years following the implementation of a risk assessment tool (Viljoen et al., 2018). However, since the reforms involved substantial changes in addition to the use of risk assessment tools (e.g., staff training and data monitoring), the extent to which risk assessment practices were responsible for changes in sentencing outcomes is unclear.

Concerns with Using Risk Assessment Tools in Sentencing

Although risk assessment tools may reduce incarceration and improve adherence to the RNR model, those who are wary of pre-sentence risk assessment highlight that tools may unduly amplify differences in sanctions between higher- and lower- risk offenders and exacerbate demographic-based disparities in sentencing (Cole, 2007; Hannah-Moffat, 2011; Harcourt, 2015; Large, Ryan, & Nielssen, 2012; Scurich & Monahan, 2016; Starr, 2014). Disparate sentencing refers to systematic differences in sentencing outcomes across groups and can contribute to systemic biases such as the overrepresentation of minority groups in prison (Starr, 2014). Although descriptors of risk do not prescribe specific sentences, they are designed to influence judges' sentencing decisions and, in doing so, might excessively influence the harshness or lenience of penalties. As such, concerns have been raised about whether "high-risk" descriptors may contribute to sentences that are disproportionate to the gravity of the offence, especially for disadvantaged groups who may score higher on tools due to risk factors that tend to co-occur with social disadvantage (e.g., poor educational attainment, early caregiver disruption, history of trauma; Cole, 2007; Maurutto & Hannah-Moffat, 2007; Tonry, 2014).

While limited research has assessed how risk descriptions and risk ratings influence judges' sentencing decisions, several studies have explored the impact of psychopathy descriptions and labels (e.g., Blais & Forth, 2014; Murrie, Boccaccini, McCoy, & Cornell, 2007; Sörman et al., 2014; Viljoen, MacDougall, Gagnon, & Douglas, 2010). For example, Murrie et al. (2007) designed a series of vignettes of a hypothetical offender whose diagnosis (e.g., psychopathy) and personality traits (e.g., psychopathic characteristics) were varied methodically. Results included that judges were less likely to recommend deferred prosecution—the least restrictive legal sanction—for offenders described as having psychopathic personality features compared to those described as not having these features; a diagnostic label of psychopathy, however, did not exert any

influence on judges' sentencing propensities. These findings suggest that descriptions of offenders, including their past behaviours and attitudes, may have a greater influence on judges' sentencing decisions than clinician-evaluated labels. However, gaps remain in our understanding of how risk assessment evidence impacts judicial outcomes.

A further concern with pre-sentence risk assessment is that legal practitioners may misconstrue the meaning of risk categories (Hannah-Moffat, 2011). An offender classified as "high-risk" shares characteristics with a subgroup of offenders that tends to recidivate at a higher rate, on average, than other subgroups; however, courts may misinterpret this risk label as a reflection of the offender's individual proclivity for crime (Hannah-Moffat, 2011; Hart & Cooke, 2013), which may (re)produce negative views of the offender and elicit more punitive sentencing than is warranted based on the gravity of the offense (Evans & Salekin, 2014). For instance, high-risk offenders may be perceived as more culpable and less amenable to treatment than low-risk offenders, despite the lack of research supporting these assumptions. Meanwhile, descriptors of "low-risk" may encourage a denial of resources and unreasonably lenient sentencing.

Concerns about risk assessment tools appear to be heightened in the context of juvenile offenders (Cole, 2007; Maurutto & Hannah-Moffat, 2007). Notably, descriptors of an offender's risk remain in their files throughout their sentence and may affect future correctional decisions such as suitability for parole and intensity of surveillance (Hannah-Moffat, 2011). This is especially problematic for juvenile offenders, many of whom tend to desist from crime as they mature (Sweeten, Piquero, & Steinberg, 2013). In addition, incarceration has been shown to interfere with developmental processes in adolescents, such as acquiring psychosocial maturity and impulse control skills, which may exacerbate their risk for future offending (Dmitrieva, Monahan, Cauffman, & Steinberg, 2012; Lambie & Randell, 2013). As such, excessively harsh decisions may negatively impact adolescent development and impede rehabilitation. Moreover, some scholars argue that focusing on a juvenile's risk might violate proportional sentencing—a governing principle in some juvenile justice laws (e.g., the Youth Criminal Justice Act in Canada)—that requires an offender's sentence to be commensurate with the seriousness of the offence and the accountability of the offender (Cole, 2007; Maurutto & Hannah-Moffat, 2007). Despite the heightened concerns related to labelling young offenders as "high-risk", risk assessment tools are widely used in juvenile justice settings (Wachter, 2015).

Judges' Opinions and Use of Risk Assessment Tools in Sentencing

The extent to which harms or benefits of risk assessment tools are realized likely depends on how much weight judges afford to the tools. As outlined by Monahan and Skeem (2016), jurisprudential theories have been developed to help explain the main objectives of sentencing. Theories based on retribution tend to emphasize the culpability of the offender, while those based on utilitarianism tend to emphasize public protection. The weight judges place on risk assessment evidence will likely vary depending on whether they view adjudication as serving a retributive or protective function. For example, risk is irrelevant if one believes that sentences should be based solely on an offender's culpability for past crimes, whereas it is crucial if one believes that sentences should be based on protecting society from future crimes (Monahan & Skeem, 2016). Although these theoretical objectives of sentencing have been used to rationalize why risk assessment tools should be required for or precluded from sentencing proceedings, few studies have examined judges' opinions about, or use of, risk assessment tools in the sentencing context.

Of the studies that have examined judges' opinions on these matters, most used a survey method to query judges' perceptions and use of risk assessment tools. Bonta, Bourgon, Jesseman, and Yessine (2005) asked 104 Canadian judges to rate the importance of various objectives of sentencing; on average, rehabilitation was considered the most important purpose of sentencing and general deterrence was considered the least important. The same study asked judges to rate the importance of 34 factors commonly included in pre-sentence reports; risk for violence was rated the 10th most important factor, while risk for general recidivism was rated the 16th most important factor. Shook and Sarri (2007) surveyed 80 American judges about their perceptions and use of risk assessment tools; judges rated the usefulness of tools as 3.24 out of 5 for making pretrial detention decisions and 3.71 out of 5 for making post-adjudication placement decisions. Almost half of these judges (48%) reported that they routinely used tools to reach a disposition. Neither of these studies asked judges to identify specific concerns they have about risk assessment tools or advantages they feel such tools provide.

A recent survey study of judges in Virginia (Monahan, Metz, & Garrett, 2018), however, did include an open-ended component querying judges' views about the role of nonviolent risk assessment in sentencing. Overall, judges expressed agreement with using a risk assessment tool to guide sentencing decisions; however, they explained that having a lack of alternative interventions limited their ability to adhere to the tool's recommendations to divert low-risk offenders. The Virginia Criminal Sentencing Commission (2012) also surveyed judges in Virginia about a previously-implemented risk assessment instrument. Although judges generally supported statewide sentencing guidelines that included a risk assessment tool, they also shared concerns about demographic factors being included in the tool (e.g., age, gender, employment status, and marital status). Finally, the U.S. Department of Justice recently published a white paper (i.e., an authoritative document describing an organization's position on a complex issue) on the implications of using risk assessments during the sentencing process (Chanenson & Hyatt, 2016). They reviewed pertinent literature and surveyed a modest, interjurisdictional sample of judges. Judges articulated some benefits of risk assessment including increased transparency and information, as well as some concerns including the potential for misusing the tool and the perpetuation of sentencing inequities.

In addition to studies querying judicial opinions, a few studies have examined the effect of tools on judicial rulings. Urquhart and Viljoen (2014) reviewed transcripts for 50 Canadian, American, and international juvenile offender cases and concluded that, overall, risk assessment evidence appeared somewhat influential in judges' sentencing decisions; however, the extent to which judges emphasized risk assessment evidence in relation to other factors varied extensively. van Wingerden, van Wilsem, and Moerings (2014) examined the effect of tool-based risk statements on sentencing outcomes. Using a sample of 6,118 Dutch offenders who were matched on several relevant characteristics, they compared sentencing outcomes for offenders with and without a tool-based pre-sentence report. They anticipated that the risk level stated in the pre-sentence reports would create a *framing effect* and cause judges to overweight information consistent with the risk level (Baron, 2008; Isaacs, 2011); contrary to their expectations, both low- and high-risk offenders with a report received less restrictive sentences than those without a report. The authors highlighted the need to replicate their findings in other countries.

Jung, Brown, Ennis, and Ledi (2015) examined associations between judicial sentencing outcomes and scores from three commonly used risk assessment tools (the Level of Service Inventory [Andrews & Bonta, 1995], Level of Service/Case Management Inventory [Andrews, Bonta, & Wormith, 2004], and Historical Clinical Risk Management-20 [Webster, Douglas, Eaves, & Hart, 1997]) that were provided to the judges; total scores were moderately predictive of sentencing outcomes ($\eta^2 = .10 - .11$). Likewise, Hilton and Simmons (2001) analysed whether associations between risk level and tribunal decisions were stronger when decision-makers had access to an actuarial tool (the Violence Risk Appraisal Guide; Quinsey, Harris, Rice, & Cormier, 1998) report; there was a small improvement in the effect of the actuarial risk score on the tribunal decision when a report was present (R^2 change = .01, $p = .05$), but this effect disappeared when the outcome was dichotomized (i.e., transfer versus detain). The conflicting results and high potential for confounding bias in these two studies makes it difficult to gauge the extent to which tools were influential.

In a large-scale study of criminal cases in Kentucky, Stevenson (in press) evaluated the impact of a pre-trial risk assessment tool on bail practices and racial disparities after a new law had mandated use of the tool. Although tools were initially associated with a small increase in release rates, these changes did not last as judges returned to their previous practices over time. Racial disparities remained unchanged after accounting for regional trends. These findings suggest that tools may not be as influential as advocates or critics suggest. As such, the author highlighted the importance of on-going evaluation and adaptation of tools. Likewise, Goldkamp and Gottfriedson (1985) argued decades ago that careful evaluation of tools is essential to ensuring that sentencing reform goals are realized.

While the studies above speak to the effect of tools on sentencing outcomes generally, some research has assessed whether certain formats of risk communication are more influential than others. Clinicians can communicate risk assessment findings in three main ways: (1) by describing an offender's relevant risk and protective factors, (2) by predicting an offender's likelihood of reoffending, and/or (3) by recommending various risk management strategies aimed at reducing risk (Evans & Salekin, 2014). Evans and Salekin (2014) queried judges' opinions about the probative value of these three formats and tested the impact of the formats on hypothetical verdicts. All formats were rated as equally probative; however, the risk prediction format resulted in more restrictive rulings

than the risk description or risk management formats. As such, the authors reasoned that risk prediction statements may be more “fear-inducing” than other messages about risk and may prompt judges to exercise greater caution in high-risk cases.

The Present Study

In sum, the proliferation of risk assessment tools in sentencing proceedings has fueled a vigorous debate about how such tools might alter sentencing outcomes for better or worse. Amidst this debate is a shortage of information about how judges themselves view or use tools in their sentencing practices. Although evidence suggests that judges consider risk assessment tools as somewhat useful for making sentencing decisions, contextual information about how judges’ think tools may be helpful or harmful is lacking. Further, little is known about how tool-based risk ratings influence judges’ impressions of offenders or the consistency of judicial decisions. Finally, it is unclear whether certain judges (e.g., ones who believe substantial weight should be placed on tools) are more persuaded by risk assessment evidence than others.

The present study was designed to address such questions. Specifically, it used a mixed-methods approach to examine (1) judges’ opinions about the risks and benefits of using risk assessment tools to make sentencing decisions, (2) the influence of a risk assessment tool on judges’ impressions of a hypothetical adolescent offender, (3) the influence of a tool on judges’ sentencing recommendations for that offender, (4) differences between judges based on how much emphasis they think should be placed on tools, and (5) the influence of a tool on the consistency of judges’ sentencing recommendations.

First, a survey was used to query judges’ opinions about the role of risk assessment tools in sentencing. The survey included a series of Likert-type questions, as well as an open-ended question about the pros and cons of using tools to guide sentencing decisions. Second, a 2 × 2 between-subjects vignette design was used to assess whether the presence of a risk assessment tool influenced judges’ impressions of a hypothetical adolescent offender (e.g., treatability and culpability) and/or their likelihood of recommending specific sentences for that offender (e.g., diversion or incarceration). For this part of the study, judges were randomly assigned to one of four vignette conditions in which risk level (low or high) and risk assessment tool (present or

absent) were systematically manipulated. Judges were asked to rate their agreement with various statements about the offender's character, as well as their likelihood of recommending certain sentences. The survey was designed to be applicable to both adult and juvenile court judges, while the vignette focused on an adolescent offender. An adolescent was chosen for the vignette because offending peaks in adolescence (Sweeten et al., 2013) and because of the increased potential for apprehension about using risk assessment tools to sentence adolescent offenders, as discussed above.

My research questions and hypotheses were as follows:

- Q1.** What are judges' opinions about using risk assessment tools to make sentencing decisions?
- H1.** No predictions are made for this research question given that the question is exploratory in nature and there is relatively little extant research on which to base hypotheses.
- Q2.** Do risk assessment tools moderate the effect of an offender's risk level on judge's impressions of the offender?
- H2.** Yes—when a risk assessment tool is present, the effect of risk level on judges' impressions of the offender will be amplified such that judges will rate the high-risk offender as less treatable and more culpable and the low-risk offender as more treatable and less culpable.
- Q3.** Do risk assessment tools influence judges' sentencing recommendations?
- H3.** Yes—for the high-risk offender, a greater number of judges will recommend a restrictive placement and intensive programming when a tool is present than when it is absent. For the low-risk offender, more judges will recommend a nonrestrictive placement and nonintensive programming when a tool is present than when it is absent.
- Q4.** Does the association between tool presence and sentencing recommendation depend on how much emphasis judges think tools should be given?
- H4.** Yes—tools will have a greater effect on recommendations for judges who say that a moderate or substantial emphasis should be placed on tools compared to those who endorse minimal or no emphasis.

Q5. Do risk assessment tools increase consistency between judges about sentencing?

H5. Yes—judges' placement and program recommendations will be more consistent when a tool is present than when it is absent.

Chapter 2.

Methods

Sample Requirements

An a-priori power analysis was conducted using G*Power 3 (Faul, Erdfelder, Lang, & Buchner, 2007) to help determine the minimum sample size that would be required for this study. I evaluated a series of effect sizes for a power of .80, using Cohen's guidelines for f s of small (.1), medium (.25), and large (.4). Table 1 displays the number of respondents that would be needed to detect an effect for small, medium, and large effects. While hypotheses about effect sizes were precluded by the limited research on the questions under scrutiny, I was most interested in detecting moderate-to-large effect sizes, as these would be the most relevant for clinicians, criminal justice professionals, and policy-makers. Based on my calculations using a Type I error rate of .05, a sample size of 158 was identified as the target necessary to detect effect sizes of $f = .25$ and greater with reasonable certainty. After delivering survey requests to 2,450 judges, I recruited 170 judges to participate in the study, thereby achieving my target sample size.

Table 1. Sample Sizes Required to Detect Specific Effect Sizes

Anticipated effect size	Required sample size
Cohen's $f = 0.1$	967
Cohen's $f = 0.25$	158
Cohen's $f = 0.4$	64

Sample Characteristics

Judges from the National Council of Juvenile and Family Court Judges (NCJFCJ; $n = 1,378$) and the American Judges' Association (AJA; $n = 1,072$) were invited to participate in this study through email. After removing cases due to missing data, 170 judges (62 women, 95 men, two who identified outside of these categories, and 11 who did not report their gender) were included in the study. As such, the response rate for

this study was 6.9%, which is slightly lower than response rates for other studies that invited judges to participate by email (e.g., 14.3% in Evans & Salekin, 2016; 8.3% in Viljoen & McLachlan, 2016; 8 to 18.8% in Wise & Safer, 2004). Of the judges who participated, 76.5% identified as Caucasian or White, 5.9% as African-American or Black, 4.7% as Latino/Latina or Hispanic, 1.8% as Indigenous or Native American, and 4.7% as 'Other'. Eleven judges did not report their ethnicity (6.5%). Judges practiced in 34 different states, including Michigan ($n = 64$), Ohio ($n = 10$), and Virginia ($n = 8$). Two practiced in British Columbia, Canada and one practiced in Quebec, Canada. Most were between the ages of 40 to 59 years old (47.6%) or 60 to 79 years old (40.6%), while the rest were under the age of 39 years old (4.7%) or did not indicate their age (7.1%). Judges had been practicing for an average of 28.3 years ($SD = 9.76$; range = 0 to 46).

When asked which age group judges had the most experience with, responses were fairly evenly distributed across adults (40.0%), juveniles (20.6%), and both (33.5%). Ten judges did not respond to this question (5.9%). When asked how often risk assessment tools are involved in the pre-sentence reports they read, 26.5% of judges responded 'always' (i.e., 99-100% of the time), 16.5% responded 'almost always' (i.e., 81-98% of the time), 14.7% responded 'frequently' (i.e., 41-80% of the time), 13.5% responded 'sometimes' (i.e., 11-40% of the time), 13.5% responded 'rarely' (i.e., 1-10% of the time), 7.6% responded 'never', and 7.6% did not respond. Finally, 40.0% of judges reported that they preside over a criminal court, 30.6% over a juvenile court, 12.4% over a family court, and 10.6% over another type of court. Eleven judges did not identify the type of court they preside over. Regarding political orientation, judges fell in the middle of the political spectrum on average (i.e., 4.71 on a scale ranging from 0 to 10), but political views ranges from far left (i.e., 0 to 2; 9.7%) to far right (i.e., 8 to 10; 9.0%).

Procedure

This study used a mixed-methods approach that combined self-report questionnaires with a 2 x 2 experimental vignette design. Judges participated in the study online using SurveyMonkey software. This design was chosen because web-based surveys are touted as being more accessible and economical than file review studies, and vignette designs can elicit stronger causal conclusions about how interrelated variables affect complex decisions (Brauer et al., 2009; Dillman, Smyth, & Christian, 2009; Muheizen, 2014).

Following approval from SFU's Research Ethics Board, judges were invited to participate in the study by email. First contact and follow-up emails were sent in accordance with Tailored-Design Methodology (Dillman et al., 2009), an evidence-based survey implementation strategy. After consenting to participate, judges were randomly assigned to one of four vignette conditions that varied based on whether the hypothetical offender was presented as *low-* or *high-risk* (i.e., described as having few or many risk factors) and whether a risk assessment tool was *present* or *absent* (see Table 2 for a tabular summary of the vignette design).

Table 2. Summary of 2 × 2 Vignette Design

		Risk assessment tool	
		<i>Present</i>	<i>Absent</i>
Risk level	<i>Low</i>	1	2
	<i>High</i>	3	4

Regardless of which condition they were assigned to, each judge responded to questions regarding their (1) legal attitudes (e.g., whether sentencing should serve a retributive or rehabilitative purpose), (2) opinions about risk assessment tools (e.g., their usefulness for informing sentencing decisions), (3) impressions of the hypothetical offender presented in the vignette (e.g., treatability, culpability), and (4) likelihood of recommending certain sentences for the offender (e.g., diversion, incarceration). The questions were presented in this order to maximize the effects that individual factors might have on judges' sentencing decisions. After responding to the above questions, judges completed a brief demographics questionnaire, including a question about their political orientation, and a few follow-up validation questions (e.g., manipulation checks). In total, the survey was estimated to take approximately 15-20 minutes to complete (see Appendix A for full survey).

Materials

Opinions about the purposes of sentencing

Judges were asked to identify, on a six-point Likert-type scale, the extent to which they agreed with five theoretical objectives of sentencing: punishment/retribution,

individual deterrence, societal deterrence, rehabilitation, and protection of society. These objectives were identified from literature that discusses how risk assessment goals compliment or conflict with jurisprudential theories of sentencing (Bonta et al., 2005; Monahan & Skeem, 2016). Response options ranged from 0 (strongly disagree) to 5 (strongly agree).

Opinions about risk assessment tools

Judges were asked to indicate, on a six-point Likert-type scale, the extent to which they agreed with various statements about risk assessment tools. Response options ranged from 0 (strongly disagree) to 5 (strongly agree). Opinions queried included the usefulness of risk assessment tools for making certain kinds of decisions (e.g., “Risk assessment tools are useful for making decisions about where to place an offender”) and the appropriateness of including certain variables in tools (e.g., “History of trauma should be included in risk assessment tools”). These questions were generated based on consultations with four experts (three psychologists and one lawyer) who have relevant research and/or practical expertise, as well as from extant literature (Monahan & Skeem, 2014). Judges also responded to three categorical questions about how much emphasis should be placed on tools, how much weight tool results should be given compared to clinical opinion, and which point(s) in the sanctioning process tools should be used. Finally, judges were asked one open-ended question: “What do you think are the pros and cons of using risk assessment tools to help make decisions about supervision, placements, or services?”.

Vignette

A vignette was developed based on prior vignette research (e.g., Murrie et al., 2007) and a casebook with sample risk assessment reports (Heilbrun, DeMatteo, Holliday, & LaDuke, 2014). The vignette provided a brief account of an adolescent offender, including his current charge and history of offending (see Appendix A). The current offence and criminal history were consistent across all four vignette conditions because research shows that these variables can influence the length of sentences imposed by judges (Hilinski-Rosick, Freiburger, & Verheek, 2014). ‘Theft Under \$5000’ was selected as the current offence because theft is one of the most common offences committed by juveniles in the United States (Sickmund & Puzanchera, 2014) and in

Canada (National Crime Prevention Centre, 2012). It was also chosen because it is of moderate severity, which should hypothetically optimize the influence of the experimental manipulations. An adolescent offender was chosen because concerns about risk assessment tools appear to be heightened in the context of juvenile offenders (Cole, 2007; Maurutto & Hannah-Moffat, 2007).

The four vignette conditions were as follows: (1) *low-risk with tool present*, (2) *low-risk with tool absent*, (3) *high-risk with tool present*, and (4) *high-risk with tool absent*. In conditions 1 and 3, the vignette included a passage explaining that a qualified clinician evaluated the offender using an evidence-based risk assessment tool, as well as an accompanying risk rating (i.e., risk to reoffend) that was consistent with the number of risk factors described. The risk rating was presented categorically because research indicates that judges tend to prefer this format and it is the most commonly used format in clinical practice (Kwartner, Lyons, & Boccaccini, 2006; Evans & Salekin, 2014, 2016). The risk assessment tool was unspecified to increase the generalizability of results. The vignettes were reviewed by a panel of four experts (the same ones as described above) and were modified based on their feedback (see Appendix B for pilot testing questions).

Impressions of the offender

A 10-item questionnaire was developed to assess judges' impressions of the treatability and culpability of the hypothetical offender. Consistent with the stepwise procedure for measure development specified by Holmbeck and Devine (2009), content was generated based on relevant theories, empirical literature, consultation with four experts (same as above), and related instruments. For example, items involving treatment amenability and long-term criminality were derived from research involving perceptions of psychopaths (Smith, Edens, Clark, & Rulseh, 2014) and juvenile delinquents (Salekin, Yff, Neumann, Leistico, & Zalot, 2002). Items involving criminal intent and responsibility were based on attribution theory (Weiner & Kukla, 1970). Judges rated items on a 6-point Likert-type scale ranging from 0 (strongly disagree) to 5 (strongly agree).

Sentencing propensities

Judges' sentencing propensities were assessed by asking judges to rate their likelihood of recommending each of the following placements: diversion, probation, non-secure residential placement (e.g., group home), and secure custodial placement (e.g., incarceration). Judges were also asked to rate their likelihood of recommending a transfer to adult court and various programming options. The programming options varied in intensity and ranged from no programs/services to a residential treatment program. Items were selected based on a review of previous studies of judicial decision-making (e.g., Murrie et al., 2007; Penner, 2009), as well as a consideration of relevant sentencing and programming options. Once again, items were rated on a six-point Likert-type scale ranging from 0 (highly unlikely) to 5 (highly likely).

Demographics questionnaire

A brief demographics questionnaire was created to collect information about participants' age, gender, ethnicity, geographical location, and judicial experience. In addition, an 11-point, left–right discrete visual analog scale was used to measure judges' political orientations. This type of scale has demonstrated stronger validity than other commonly used response formats, such as 10-point and 101-point response scales (Kroh, 2007). While people must have a certain degree of social cognition to describe their political orientation in relation to others', research indicates that, in general, people are adept at locating their political views on left–right spectrums. Indeed, reliability and validity estimates for left–right political scales are higher than those for other common survey items (Kroh, 2007). The wording for this item was consistent with that used in other studies of political orientation (Kroh, 2007; Mowle, Edens, Clark, & Sörman, 2016).

Analytic Plan

Descriptive

Central tendency, variability, skewness, and kurtosis were examined for each of the study variables. Skewness and kurtosis were assessed visually by inspecting histograms and Q–Q plots of questionnaire responses across groups and the total sample. Outliers were not expected in this dataset given the small range of the response

scales. However, responses to the manipulation checks were examined and an incorrect response to both checks was considered justification for removing a participant from analyses pertaining to the vignette. I also examined the dataset for missing data, and judges who failed to respond to over 30% of the items were excluded from subsequent analyses.

Research Question 1

To help understand judges' opinions about risk assessment tools, I examined their mean responses to the questionnaire pertaining to their opinions about tools. I also tested whether these responses differed according to judges' responses to the demographics questions. For dichotomous items (e.g., gender), I conducted Mann-Whitney U tests using Bonferroni corrections. For items with more than three responses options (e.g., whether judges had more experience with juveniles, adults, or both), I conducted Kruskal-Wallis H tests using Bonferroni corrections.

I also explored judges' opinions about risk assessment tools using conventional content analysis (CCA). Specifically, I analysed judges' responses to the question, "What do you think are the pros and cons of using risk assessment tools to help make decisions about supervision, placements, or services?". As outlined by Hsieh and Shannon (2005), CCA is an inductive approach to qualitative content analysis and involves approaching the data without a preconceived categorical structure. Instead, the researcher allows categories/themes to emerge spontaneously as they read and re-read participant responses.

Following the steps recommended by Hsieh and Shannon (2005), I first reviewed participants' responses several times to achieve a broad sense of the data. Next, I read the data word-by-word and highlighted important concepts, while making notes about my first impressions and thoughts of the data. Through this process, I generated several "codes" (i.e., labels that captured key ideas in the data). I sorted these codes into meaningful subcategories, or themes, which were subsequently grouped into a smaller number of broad categories to help organize the data. Finally, I developed a definition for each category and subcategory and selected excerpts from the data that exemplified each of them.

Hsieh and Shannon (2005) point out that the main weakness of CCA is that it can increase the likelihood of missing important themes in the data since relevant theories are not considered. Consistent with recommendations (Creswell, 2013), I compensated for this weakness by using intramethod triangulation (i.e., using more than one research method to better understand the phenomenon of interest). Specifically, the analysis of responses to the open-ended item was complemented by the analysis of responses to the Likert-type and categorical items. Moreover, I provided rich descriptions of the data such that readers can draw their own conclusions about the appropriateness of the categories and subcategories that were identified using CCA.

Research Question 2

To test whether a risk assessment tool moderates the association between risk level and judges' impressions of an offender, I conducted a series of ordinal logistic regression (OLR) models. In total, I ran ten OLR analyses to examine the main effects and interaction effects of the predictor variables (i.e., risk level and tool presence) on each of the outcome variables (i.e., the 10 items used to evaluate judges' impressions of the offender). I used Nagelkerke's pseudo R-square estimates to describe the proportion of the total variance that was accounted for by each OLR model and odds ratios (ORs) to describe the magnitudes of the effects. Since familywise Type 1 error rates can be inflated when conducting multiple tests of significance, I applied Bonferroni corrections by adjusting the alpha level for each regression model based on the number of tests being performed.

OLR has four assumptions: (1) the outcome variable must be measured at the ordinal level, (2) the predictor variables must be treated as either categorical or continuous, (3) there must be no multicollinearity between the predictor variables, and (4) there must be proportional odds across the cumulative splits of the outcome variable. The first two assumptions were automatically met due to the nature of my data. However, to ensure that the assumption for no multicollinearity was met (i.e., to ensure that the predictors were not excessively correlated), I evaluated the phi-coefficient between the two predictor variables (i.e., risk level and tool presence). To confirm that the assumption for proportional odds was met (i.e., to ensure that a single effect size could be used to accurately summarize the effects at different levels of the outcome variable), I ran a test of parallel lines for each model. Since this test is known for being

overly sensitive (O'Connell, 2006), I followed up significant tests (i.e., violations) with a series of binary logistic regressions to help decide whether the ordinal approach was justified. Specifically, I compared the ORs across each level of the outcome variable to determine whether they were reasonably similar. I followed recommendations by O'Connell (2006) that if the ORs are reasonably similar, the cumulative OR produced by the ordinal model can be used to describe the results; if the ORs are not similar, separate logistic regression results should be used to help explain the pattern of effects.

Research Question 3

To assess whether risk assessment tools influence judges' sentencing recommendations, I used loglinear analyses. Specifically, I examined the main effects and interaction effects of the predictor variables (i.e., risk level and tool presence) on the outcome variables (i.e., placement and program recommendations). For the outcomes, judges were given four placement options to choose from (diversion, probation, nonsecure residential, or secure custodial) and four program options to choose from (none, weekly, intensive, or residential). As such, I planned to conduct two 2 x 2 x 4 loglinear analyses. Loglinear analysis is a more sophisticated version of Pearson's chi-square test and has similar assumptions: cells in the contingency table must be independent and expected frequencies must exceed five in at least 80% of the cells. Further, expected frequencies must exceed one in all cells. The first assumption was automatically met due to the nature of my data and the second was verified by analyzing contingency tables for my variables. For tests in which the second assumption was violated, I followed guidelines recommended by Field, Miles, and Field (2012) to explore whether data could be collapsed either across one of the variables or across the levels of one of the variables.

Research Question 4

To evaluate whether tools had a greater effect on judges who reported that more emphasis should be placed on tools, I divided judges into two groups: those who endorsed a moderate or substantial emphasis on tools and those who endorsed minimal or no emphasis on tools. I planned to conduct loglinear analyses to evaluate the three-way interactions between Tool Presence x Judicial Emphasis x Sentencing Recommendation across high-risk and low-risk vignettes. However, given the power

reduction caused by separating groups based on tool presence, offender risk level, *and* judicial emphasis, assumptions were only met for the analysis pertaining to placement recommendations for the high-risk offender. As such, I narrowed my research question and performed a single loglinear analysis to assess the Tool Presence x Judicial Emphasis x Placement Recommendation interaction for the high-risk vignettes.

Research Question 5

To examine whether the presence of a risk assessment tool increases consistency in judges' sentencing recommendations, I tested whether the standard deviations for placement and program recommendations were significantly different across groups using Krishnamoorthy and Lee's (2014) modified signed-likelihood ratio test (MSLRT). If the omnibus MSLRT was significant, I planned to conduct follow-up pairwise comparisons to assess which pairs of standard deviations were significantly different from one another. I used the R package 'cvequality' to conduct these tests (Version 0.1.3; Marwick & Krishnamoorthy, 2018).

Chapter 3.

Results

Descriptive Statistics

Judges were reasonably dispersed among the four vignette groups, with 38 judges in the low-risk, tool-present group, 34 in the low-risk, tool-absent group, 41 in the high-risk, tool-present group, and 49 in the high-risk, tool-absent group. Judges' scores on items related to their legal attitudes and opinions about risk assessment tools were comparable across the four vignette groups, as indicated by a series of nonsignificant ANOVAs. Age, gender, ethnicity, judicial experience, and political orientation were also similar across vignette groups. As such, the groups were considered comparable across key study variables. As expected, no outliers were present, and the actual range obtained for each variable was the same as its theoretical range. Fifty-six cases were removed due to missing data. Thirty-two of these cases provided no responses at all, which suggests that they either had technical issues completing the survey or decided not to participate after reviewing the consent form. A further 24 judges only answered the first five questions.

Ninety-four judges (55.3%) provided qualitative responses to the open-ended question, "What do you think are the pros and cons of using risk assessment tools to help make decisions about supervision, placements, or services?" To determine whether certain kinds of judges were more likely to respond to this question than others, I conducted a series of t-tests and chi-squared analyses using Bonferroni corrections. Judges who did and did not respond were similar in gender, ethnicity, and political orientation, as well as on the 27 items related to legal attitudes and opinions about risk assessment tools. However, judges who responded to this item were more likely to be from the older age group (i.e., 60-79 years) than the younger age groups (i.e., 20-59), $\chi^2(1) = 10.74, p < .01$. Similarly, they reported more years of experience ($M = 31.03, SD = 9.17$) than those who did not respond to this question ($M = 25.00, SD = 9.48$), $t(135) = -3.77, p < .001$. Finally, judges who responded to this question disagreed more strongly that race should be included in risk assessment tools ($M = 0.80, SD = 1.26$) compared to those who did not respond ($M = 1.72, SD = 1.68$), $t(136) = 3.99, p < .001$.

Manipulation checks

The first manipulation check asked judges which crime the offender committed. In response to this check, 95% of judges correctly identified the crime as 'Theft under \$5000'. Five judges were unsure which crime he had committed, three thought he had committed a drug offense, and none thought he had committed a common assault. Interestingly, the three who identified the crime as a drug offense were all in the high-risk, tool-present vignette condition and one reported in the feedback section that he was looking for a 'Theft over \$5000' option. As such, these errors could have been related to the framing effect of the vignette rather than a lack of attention and, consequently, those judges were not eliminated from the analyses. The second manipulation check involved identifying which risk level the offender was identified as in the vignette, so it pertained only to those vignettes in which a tool was present. Of the judges who received a low-risk, tool-present vignette, 92.1% correctly stated that the offender's risk level was identified as low. One thought the risk level was identified as moderate, one thought it was identified as high, and one thought that it was not identified. Of the judges who received a high-risk, tool-present vignette, 75.6% correctly stated that the offender's risk level was identified as high. Five judges thought it was identified as moderate and five thought it was not identified. Since none of the judges answered both manipulation items incorrectly, none were removed from the analyses.

Research Question 1: What are judges' opinions about using risk assessment tools to make sentencing decisions?

Judges mean responses to the Likert-type items querying their opinions about risk assessment tools are presented in Table 3. Overall, judges agreed only slightly that tools predict reoffending with reasonable accuracy. They expressed slightly stronger agreement that tools should be used primarily to identify high-risk offenders for increased sanctions, rather than low-risk offenders for reduced sanctions. Regarding which types of decisions tools are most useful for, they agreed most strongly that tools are useful for making program/service decisions and disagreed slightly that tools are useful for informing the length of an offender's sentence. On average, judges disagreed that tools lead to more restrictive sentences but were undecided about whether tools lead to less restrictive sentences. There was stronger agreement that tools increase, rather than decrease, fairness and consistency in sentencing. However, judges, as a

group, disagreed that tools reduce the overrepresentation of minorities in prison. They also disagreed that tools contribute to such overrepresentation.

On average, judges moderately agreed that it is important to consider future risk when making sentencing decisions and slightly disagreed that it is inappropriate to sentence someone based on what they might do in the future. They expressed moderate agreement with the idea that more caution is warranted when using risk assessment tools to inform juvenile dispositions compared to adult sentences. Finally, regarding potentially contentious items that might be included in risk assessment tools (i.e., gender, race, age, trauma history, educational attainment, employment status), judges agreed most strongly that an offender's history of trauma should be included. They also agreed that age, educational attainment, and employment status should be included but were undecided about whether gender should be included. Race was the only factor that they disagreed should be included in tools. After applying Bonferroni corrections, judges' responses to these items did not differ according to demographic variables (i.e., age, gender, ethnicity, or political views) or the age group they have the most experience with (i.e., adults, juveniles, or both).

In addition to the Likert-type items above, judges responded to three categorical items. Consistent with past research, most judges (55.3%) said that tool results should be given moderate emphasis when making placement decisions (e.g., custodial versus non-custodial placement), while 26.5% said they should be given minimal emphasis. Of the remaining judges, 12.4% said tool results should be given substantial emphasis, 5.3% said they should be given no emphasis and one judge did not respond. Compared to clinical opinion, most judges said risk assessment tool results should be given less emphasis (49.4%), 40.0% said they should be given equal emphasis, and 8.2% said they should be given more emphasis; four judges (2.4%) did not respond. Finally, when asked which point(s) in the sanctioning process risk assessment tools should be used at, 71.8% endorsed using them at the pre-trial detention stage, 29.4% at the adjudication stage, 81.2% at the disposition/sentencing stage, and 65.3% at the parole/release stage. For this question, judges could select more than one response.

Table 3. Psychometric Properties of Variables Related to Judges' Opinions about Risk Assessment Tools

Variable	<i>N</i>	<i>M</i>	<i>SD</i>	Skew	Kurtosis
To what extent to you agree with the following statements about risk assessment tools?					
Tools predict an offender's risk for reoffending with reasonable accuracy	170	2.96	1.27	-0.80	0.01
Tools should be used primarily to identify low-risk offenders (i.e., to reduce their sanction)	169	2.59	1.42	-0.28	-0.70
Tools should be used primarily to identify high-risk offenders (i.e., to increase their sanction)	168	2.86	1.52	-0.49	-0.75
Tools are useful for making decisions about the length of an offender's sentence	167	2.41	1.49	-0.21	-1.06
Tools are useful for making decisions about where to place an offender (e.g., in custody or in the community)	169	3.40	1.32	-0.93	0.32
Tools are useful for making decisions about an offender's supervision level	168	3.66	1.23	-1.06	0.89
Tools are useful for making decisions about program/service referrals	170	3.91	1.14	-1.45	2.43
Tools lead to more restrictive sentences (i.e., more offenders being incarcerated)	164	1.57	1.22	0.70	0.17
Tools lead to less restrictive sentences (i.e., fewer offenders being incarcerated)	166	2.71	1.31	-0.39	-0.52
Tools increase fairness and consistency in sentencing	168	3.03	1.40	-0.63	-0.25
Tools decrease fairness and consistency in sentencing	164	1.60	1.33	0.91	0.42
Tools help reduce the overrepresentation of minorities in jail/prison	163	2.09	1.40	0.14	-0.85
Tools contribute to the overrepresentation of minorities in jail/prison	164	1.76	1.30	0.68	0.13
It is inappropriate to sentence someone based on what they might do in the future	167	2.14	1.53	0.33	-0.84
It is important to consider future risk when making sentencing decisions because it is relevant to public protection and safety	169	3.82	1.18	-1.32	2.04
We should be more cautious when using RAs to inform juvenile dispositions than adult sentences	161	2.88	1.53	-0.32	-0.92

Variable	<i>N</i>	<i>M</i>	<i>SD</i>	Skew	Kurtosis
To what extent do you agree that the following risk factors should be included in tools?					
Gender	168	2.45	1.72	-0.17	-1.25
Race	170	1.21	1.53	1.04	-0.05
Age	169	3.63	1.37	-1.17	0.95
History of trauma	168	4.14	1.15	-1.82	3.87
Educational attainment	169	3.54	1.35	-0.99	0.51
Employment status	168	3.62	1.23	-1.06	1.12

Note. Response options ranged from 0 (*strongly disagree*) to 5 (*strongly agree*). As such, 2.50 reflects the midpoint of the scale.

In response to the open-ended question, “What do you think are the pros and cons of using risk assessment tools to help make decisions about supervision, placements, or services?”, judges expressed diverse opinions ranging from definitive support to strong skepticism. Using CCA, I identified 35 codes and organized them into a total of 14 subcategories. These subcategories seemed to fit best into three broad categories: (1) benefits of using tools ($n = 59$ comments made by 46 judges), (2) concerns with using tools ($n = 108$ comments made by 61 judges), and (3) circumstances that influence the usefulness of tools ($n = 32$ comments made by 28 judges; see Figure 1 for tree diagram). Within the first category, judges identified six main benefits to using tools: (a) they are helpful for making decisions, (b) they increase objectivity and consistency, (c) they help divert low-risk offenders, (d) they may reduce overincarceration, (e) they facilitate transparency, and (f) they may reduce racial disparity. Within the second category, judges identified six main concerns about using tools: (g) concerns about bias, (h) concerns about tool validity, (i) concerns about using a “one-size-fits-all” approach, (j) concerns about shifting discretion away from judges, (k) concerns that tools are too costly or time-consuming, and (l) concerns that tools are unnecessary. Finally, within the third category, two subcategories emerged reflecting the sentiment that the usefulness of tools is contingent on certain factors: (m) the usefulness of tools depends on the characteristics of the tool, the tool user, and the context the tool is being used in and (n) tools are most useful when they are treated as one piece of evidence among many. In the following sections, each category and subcategory will be discussed in turn. Excerpts from the data will be used to help illustrate the subcategories.

Benefits of using tools

Overall, 46 judges made 59 comments relating to benefits of using tools. The most common reason judges articulated for using risk assessment tools ($n = 31$) was that *they are helpful for making decisions*. Most comments within this subcategory noted that tools are helpful in general, simply because they provide relevant information. For example, one judge wrote, “I look at assessment tools as a way to get more information. In making my decisions, more info is always better than less”. Some judges also highlighted specific decisions that tools can assist with such as those concerning pretrial release, placement, supervision, services, and release from parole/probation. Regarding

service decisions, one judge noted that “being able to identify areas of need for types of services for juvenile offenders (to address inadequate social supports, deficiencies in education, mental health needs, substance abuse, problems at home, etc.)” was an advantage of using a tool.

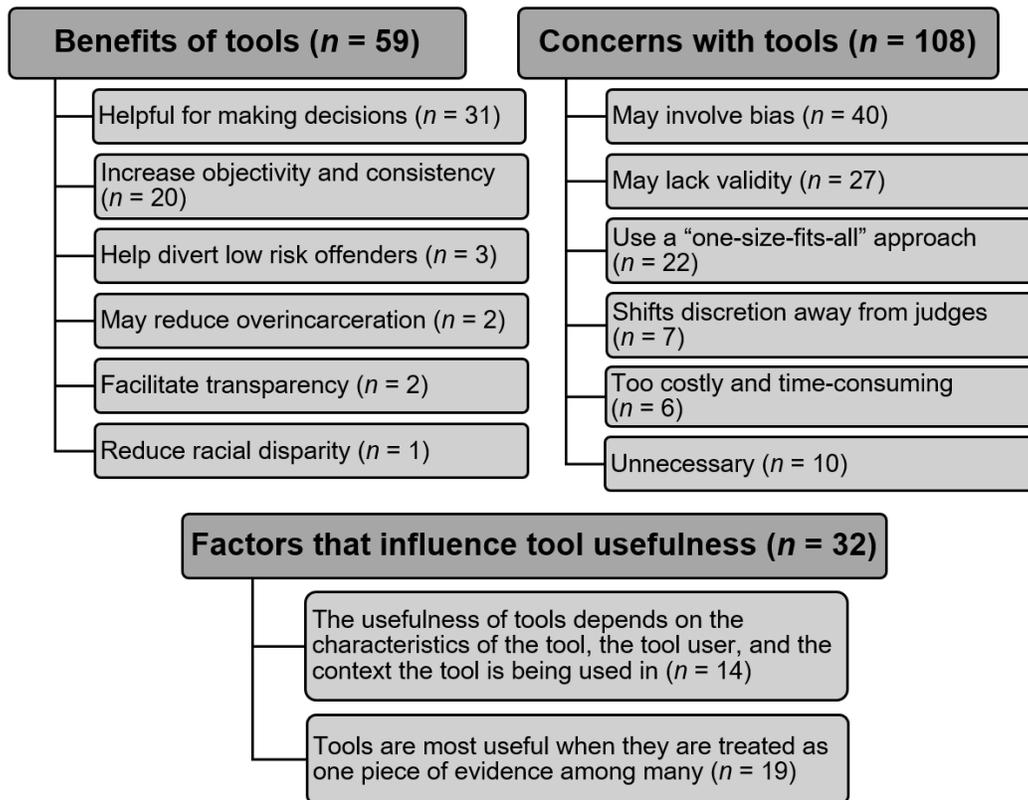


Figure 1. Tree diagram depicting the categories and subcategories that emerged from the conventional content analysis.

Several judges ($n = 20$) also emphasised that *risk assessment tools can help increase objectivity and consistency* by reducing bias, as illustrated by the following comments by two different judges:

I can think of no "cons." The pros are that, as individual judges, absent an evidence-based risk assessment tools [*sic*], we are shooting in the dark, relying only on the past criminal history, allocution, and our gut instincts. Not to say that these factors are unimportant, but it is difficult for individual judges to always recognized and set aside implicit biases. An objective risk assessment tool, can balance those factors.

Facts and circumstances before us come first however bias or emotion can interfere with justice. As such, risk assessment tools are extremely valuable tools for thought and participate in bringing us to consider factors that may have been overlooked.

Other judges seemed to resonate with these sentiments. Some expressed support for using “evidence-based practices” generally, while others highlighted that tools can help courts produce more reliable and consistent dispositions.

Another subcategory captured judges’ views that *risk assessment tools can help divert low-risk offenders* from programs that are geared towards medium- and high-risk offenders ($n = 3$). Exemplifying this subcategory, one judge wrote that “a risk assessment can be effective in identifying low-risk individuals who can be diverted from programming which avoids increasing recidivism by placing them with high-risk offenders.” A similar subcategory captured two judges’ views that tools *may reduce overincarceration*. One of these judges referred to meta-analytic research showing that petitions and recidivism rates in their court have decreased substantially in the 15 years since they started using an empirically supported risk assessment tool.

Two judges suggested that *risk assessment tools can facilitate transparency*. One of these judges explained that a risk assessment tool “helps the court frame the conversation with [an] offender at the sentence hearing,” while the other noted that it can help “articulate justification for dispositional orders.” These comments seem to capture ways that individual judges use tools in their practice. Finally, one judge noted that tools “should *reduce racial disparity*”.

While the subcategories above reveal that judges see several potential advantages of using risk assessment tools, the following subcategories highlight that judges also have some important concerns about using tools.

Concerns with using tools

In total, 61 judges made 108 comments relating to concerns with using tools. The most common criticism of tools articulated by judges related to *concerns about bias* ($n = 40$). Judges pointed out a variety of ways that bias can arise during the risk assessment process. Several judges indicated that the data used to conduct a risk assessment may be inaccurate, especially if it is reliant on an offender’s self-report. For instance, one judge noted that “in Michigan the self-reporting of information by defendants often results in non-sensical risk-assessments when compared to actual, factual recorded history regarding the defendant.” Other judges echoed concerns about the accuracy of self-

reported data and one suggested that tools are vulnerable to manipulation by astute offenders.

Beyond concerns about the data collection process, judges expressed concern that tools may be inherently biased due to the weight given to certain factors. For example, one judge wrote that “the ‘objectivity’ of many tools are illusory [*sic*] because the values assigned to certain risk factors are subjective, rendering the supposedly objective score invalid.” Skepticism about the neutrality of tools was accompanied by concerns about the competency and biases of those using the tools, such as assessors and organizations. One judge pointed out that “the bias, explicit and implicit, of the person administering the tool, together with the general culture of the organization that person belongs to impacts the objectivity of the tool.” Another judge wrote, “Garbage in = Garbage out. Too many times have seen the tools being manipulated by policy to encourage a certain result. Usually the result is tied into funding. This creates huge credibility issues for the court.” Finally, a few judges raised concerns that tools may be racially biased, as illustrated in the following excerpt:

In my experience as both a judge and practitioner I have become convinced the disproportionate outcomes based upon race and economic factors are not well reflected in risk tools. Too much emphasis is placed upon subjective factors which seem to adversely impact people of color.

As the above examples demonstrate, concerns about various types of bias were prominent in judges’ responses.

The second most common issue raised by judges related to *concerns about the validity of risk assessment tools* ($n = 27$). Several judges expressed a desire for more research concerning the predictive accuracy and error rate of tools. One indicated a discomfort with using tools because judges “must blindly rely on the scientific accuracy of the assessor without understanding the basis of the assessment tool.” Another reasoned that “there should be long-term studies done with respect to the accuracy of risk-assessment tools and adjustments made for socioeconomic status, race, cognitive impairments/special ed, and parental criminal histories.” Some judges expressed a general distrust towards tools, such as one who wrote, “I remain sceptical of non-human calculations and continue to

rely on human intuition and judgment for my decision making.” Another shared the following anecdote:

I find it almost laughable that the risk assessment tool comes up with conclusions that are demonstrably untrue. For example, I'm told that the assessment says that the offender is unlikely to engage in future criminal behavior, but after the report was prepared and before the sentencing took place, the offender committed a new crime.

Other judges shared concerns about the current lack of empirical support for tools but expressed more optimism regarding the future of risk assessment tools:

I think the field of risk assessment tools, while it has developed over the last several years, still has a ways to go in determining risk and appropriate decisions by the court. I anticipate that in the next decade there will be more developments so that someday the tools we are using today will be outdated and not relied on. The tools will be used in more jurisdictions which will help to norm the tools for different types of communities.

Collectively, these comments reflect judges' views that there is insufficient evidence available to support the use of risk assessment tools.

Another prevalent concern among judges was that *risk assessment tools may obscure the individuality of offenders by using a “one-size-fits-all” approach (n = 22)*. One judge, for example, explained that tools “are generalized and can't take into consideration all the possibilities on a particular individual.” Another wrote the following:

Each juvenile is different. The court should use its judgment to make its decisions, not some formula based on so-called objective criteria. Each child is different. Juveniles are individual persons, not shoes. A one size fits all approach to juvenile justice is not workable.

Relatedly, some judges felt that some important variables are not captured by tools. Factors that judges noted were missing from tools included the credibility of the offender, the offender's history of trauma, the severity of the crime, and the “length of time between charges or out of custody time between charges.”

Another subcategory related to judges' concerns that *tools shift discretion away from judges (n = 7)*. For instance, judges expressed concern that tools shift too much discretion to police or to an “often nameless evaluator”. One judge argued that “judges should be free to make decisions based upon their own considerations,” while another proclaimed, “Let judges be judges.” Likewise, one

judge said, "I don't think a risk assessment tool can substitute for the experience and judgment of the decision maker - otherwise we could simply have a computer program do our job." This subcategory appears to capture some worries related to the ways that risk assessment tools may be changing the role of judges. One judge alluded to another judge's prediction that "we may someday live in a society where people are sent to prison by an algorithm rather than by a judge."

Another subcategory highlighted concerns that *tools are too costly and time-consuming* ($n = 6$). One judge noted that "most courts outside of metropolitan areas do not have pretrial services so often [have] inadequate staff and resources" to implement tools. Another stated that "risk assessment tools are a waste of time and resources for the taxpayer." These comments illustrate that some judges view tools as impractical due to the time and resources required to implement them.

A final subcategory captured the position of some judges that *risk assessment tools are unnecessary* ($n = 4$). One judge wrote, "I find risk assessment tools completely unnecessary for my sentencing. I won't use them." Another suggested that "much of it is common sense as to a personal history." These judges seemed to feel that tools are unhelpful in general.

Circumstances that influence the usefulness of tools

Overall, 28 judges made 32 comments about circumstances that influence the usefulness of tools. Although some judges were clearly for or against using risk assessment tools to make judicial decisions, many indicated that *the usefulness of tools depends on certain factors*, such as the type of tool, the user of the tool, and the purpose for which the tool is used ($n = 14$). One judge pointed out that "there are so many different risk assessment tools that answering a question about a tool is quite difficult. It really does matter which one." Similarly, another indicated that it "depends on the tool, the expertise of the person using the tool and the knowledge of the person relying on it." Another judge asserted that tools are "useful to inform risk needs assessments only. They should NEVER be relied on to inform length of sentence."

Furthermore, several responses ($n = 19$) suggested that *tools are "just tools" and should be treated as only one piece of evidence among many*. For instance, one judge

suggested that tools should be used as “an aide to judgement not a substitute for it.” Another noted that tools “must be a part of any decision but [are] not the only factor.” The following quotation expresses a similar idea:

Risk assessments are tools and should be treated as such. They should be tested for their statistical reliability and validity. The person administering the assessment should be properly trained in order to glean useful information. Other information about the youth should supplement the risk assessment and common sense should be used. If circumstances warrant using going [*sic*] a different direction than a risk assessment, the [*sic*] judicial discretion should be used.

Overall, this subcategory reflects the view that tools are most valuable when they are empirically tested, used by appropriately trained individuals, and considered in the context of other pertinent information.

Research Question 2: Do risk assessment tools moderate associations between an offender’s risk level and judges’ impressions of the offender?

Prior to testing whether risk assessment tools influenced judges’ sentencing recommendations for the hypothetical offender, I assessed whether impressions of that offender differed between the tool-present and tool-absent conditions. A series of 10 OLRs were computed to test the hypothesis that the presence of a risk assessment tool (and accompanying risk rating) would exacerbate the effect of offender risk level on judges’ impressions of the offender. Using a Bonferroni adjustment of $.05/10$, an alpha level of $p < .005$ was considered the threshold for significance.

A nonsignificant phi-coefficient ($\phi = .072$, $p = .361$) indicated that no multicollinearity existed between the predictor variables. However, the test of parallel lines was significant in four of the 10 planned analyses, indicating that the assumption of proportional odds was questionable for those outcomes. In those cases, I followed up the OLR models with binary logistic regressions corresponding to each cumulative split of the outcome. Upon examining the pattern of effects for each of these four variables, I determined that the violation of proportionality was likely arising from the low number of responses in the extreme response categories (i.e., strongly agree or strongly disagree). Otherwise, the effects were reasonably consistent in direction and magnitude with the

overall effect sizes produced by the omnibus OLR models. As such, I used overall effect sizes to describe the results from all 10 OLR models.

The main effect of risk level (i.e., whether the offender was described as having several or few risk factors) was significant in six out of the 10 analyses. Details of these main effects are presented in Table 4. When the offender was presented as having several risk factors, judges believed more strongly that he would be a lifelong criminal ($M = 2.53$ vs. $M = 0.60$), that he deserved a harsher penalty than other offenders who committed a similar crime ($M = 2.05$ vs. $M = 0.96$), that personality factors contributed to his behaviour ($M = 3.68$ vs. $M = 2.19$), and that situational factors contributed to his behaviour ($M = 4.11$ vs. $M = 2.76$). They believed less strongly and that he could be rehabilitated ($M = 3.79$ vs. $M = 4.60$) but agreed more strongly that they would refer him to a treatment program ($M = 3.85$ vs. $M = 2.19$). Recall that scores below 2.50 reflect disagreement and scores above 2.50 reflect agreement. So, judges in the high-risk and low-risk conditions both tended to disagree, on average, that the offender deserved a harsher punishment, but those in the low-risk condition disagreed with that statement more strongly.

There were no differences between high- and low-risk vignettes in judges' impressions of whether the offender's actions were spontaneous or planned or whether his risk could be reduced with appropriate interventions. Neither the main effect of tool presence nor the interaction between risk level and tool presence were significant in any of the analyses, even without the Bonferroni correction, indicating that the risk assessment tool did not moderate the associations outlined above between offender risk level and judges' impressions of the offender. To maintain clarity and conciseness, details for these nonsignificant effects are not presented.

Table 4. Ordinal Logistic Regression Results for the Main Effect of Risk Level on Judges' Impressions of the Offender

Variable	Estimate (SE)	95% CI for odds ratio			<i>R</i> ²
		LL	OR	UL	
I would refer Steve to a treatment program	-2.246 (0.434)*	4.03	9.45	22.11	0.251
Steve will probably be a lifelong criminal	-2.895 (0.487)*	6.95	18.08	46.99	0.428
Steve stole the pellet gun purposefully, with the intention to harm someone	-1.056 (0.426)	1.25	2.87	6.63	0.185
Steve stole the pellet gun spontaneously, without intending to harm anyone	1.075 (0.421)	0.15	0.34	0.78	0.064
Steve deserves a harsher penalty than other juveniles who committed similar crimes	-1.381 (0.428)*	1.72	3.98	9.21	0.129
Steve has the potential to be rehabilitated	2.537 (0.560)*	0.03	0.08	0.24	0.245
Steve planned the current offense ahead of time	-0.134 (0.414)	0.51	1.14	2.57	0.004
Personality factors contributed to Steve's unlawful behaviour	-1.860 (0.436)*	2.73	6.42	15.10	0.287
Situational factors contributed to Steve's unlawful behaviour	-1.818 (0.438)*	2.61	6.16	14.53	0.302
With appropriate interventions, Steve's risk will likely be reduced	-1.204 (0.440)	0.13	0.30	0.71	0.105

Note. CI = confidence interval; LL = lower limit; OR = odds ratio; *R*² = Nagelkerke's pseudo R-square for overall ordinal logistic regression model; SE = standard error; UL = upper limit.

**p* < 0.005

Research Question 3: Do risk assessment tools influence judges' sentencing recommendations?

Initially, I had planned to conduct two loglinear analyses to test the hypothesis that a risk assessment tool would influence (a) the restrictiveness of judges' placement recommendations and (b) the intensity of their program recommendations. However, for the analysis relating to placement recommendations, 37.5% of cells had expected frequencies that were less than five and, for the analysis relating to program recommendations, 56.3% had expected frequencies that were less than five. Further, several cells had expected frequencies under one, which resulted in a severe assumption violation. Although expected cell counts are distinct from observed cell counts, expected counts were reduced because some cells had an observed count of zero. Specifically, judges were unlikely to recommend a restrictive sentence or intensive programming for the low-risk offender, regardless of tool presence (see Table 5). Conversely, judges were unlikely to recommend a lenient sentence or less intensive programming for the high-risk offender (see Table 6). Although this pattern makes intuitive sense, it violated the cell count assumption, which causes such a substantial reduction in power that it is considered not worthwhile to proceed with loglinear analysis (Field et al., 2012).

Field et al. (2012) summarizes possible solutions to this problem. One option involves collapsing the data across the levels of one of the variables. The placement recommendation variable had four levels, so I considered dichotomizing it into more restrictive and less restrictive placements. However, of the judges who received a low-risk vignette, only one recommended one of the two most restrictive placements (see Table 5). Consequently, combining the two highest and two lowest response categories did not rectify the problem (i.e., there were still 37.5% of cells with expected values over five). Although the frequency table for program recommendations was more promising than the one for placement recommendations (see Table 6), dichotomizing the program variable still resulted in 25% of cells having expected values under five.

A second option suggested by Field et al. (2012) involves collapsing the data across one of the variables (i.e., either risk level or tool presence) and computing Pearson's chi-square tests with the two remaining variables. However, this would entail

evaluating the main effect of risk level or tool presence separately while ignoring the interaction effect, which would not help in determining whether risk assessment tools moderate the effect of risk level on sentencing recommendations. Although loglinear analysis is the preferred method for testing a three-way interaction, it can also be tested by conducting a chi-square analysis at different levels of each variable (Field et al., 2012). Indeed, this is how a significant loglinear test is typically followed up. Therefore, as an alternative to using loglinear analysis, I split my dataset by tool presence (i.e., present or absent) and conducted parallel chi-square analyses between risk level and sentencing recommendation (i.e., placement or program recommendation). I then compared standardized residuals and odds ratios to assess whether associations between the offender’s risk level and judges’ sentencing recommendations were stronger when a risk assessment tool was present than when it was absent.

Table 5. Frequency Table Displaying the Number of Judges in Each Condition Who Recommended Each Placement Type

Risk level	RA tool	Diversion	Probation	Nonsecure residential	Secure custodial
Low	Present	7	29	0	0
	Absent	5	26	0	1
High	Present	1	14	13	8
	Absent	3	19	14	10

Table 6. Frequency Table Displaying the Number of Judges in Each Condition Who Recommended Each Program Intensity

Risk level	RA tool	None	Weekly	Fulltime	Residential
Low	Present	4	28	2	2
	Absent	2	25	3	2
High	Present	0	2	25	9
	Absent	3	7	25	10

Standardized residuals correspond to z-scores such that a value of ± 1.96 is significant at $p < .05$, a value of ± 2.58 is significant at $p < .01$, and a value of ± 3.29 is significant at $p < .001$. As such, standardized residuals above or below these values can be considered evidence that the observed cell count is significantly higher (or lower)

than would be expected if the null hypothesis were true (i.e., if cell frequencies were determined by chance). Since odds ratios are most interpretable and useful in 2 x 2 contingency tables (Field et al., 2012), I dichotomized the two outcome variables such that placement recommendations were divided into restrictive (nonsecure residential and secure custodial placements) and nonrestrictive (diversion and probation) categories, and program recommendations were divided into intensive (fulltime and residential programming) and nonintensive (no programming or weekly programming) categories. The assumption that expected cell counts be ≥ 5 was met for all chi-square analyses. Since the analysis pertaining to program recommendations had a more minor assumption violation (i.e., 25% of cells violated the assumption compared to the 20% that is typically considered acceptable), I ran the full model using loglinear analysis as a final step. Results for each pair of analyses are presented below.

Placement restrictiveness

Chi-square tests revealed significant associations between risk level and placement recommendation for both the tool-absent, $\chi^2(1) = 20.85, p < .001$, and tool-present, $\chi^2(1) = 29.65, p < .001$, group. To assess whether these associations differed between the tool-absent and tool-present groups, I compared the size of the standardized residuals and odds ratios. For the high-risk vignettes, 58.3% of judges recommended a restrictive placement when a tool was present, compared to 52.2% of judges who did so when a tool was absent (see Figure 2). Standardized residuals were further from zero when a tool was present (see Figure 3). Specifically, the standardized residual for recommending a restrictive sentence was 3.2 when a tool was present compared to 2.4 when a tool was absent. However, odds ratios were comparable across tool-present (OR = 1.40) and tool-absent (OR = 1.09) conditions.

For low-risk vignettes, 0% of judges recommended a restrictive placement when a tool was present, compared to 3.1% when a tool was absent. Although an OR could not be defined due to a count of zero in one cell, this difference was considered negligible because the standardized residuals were similar in the tool-present (-3.2) and tool-absent (-2.9) groups.

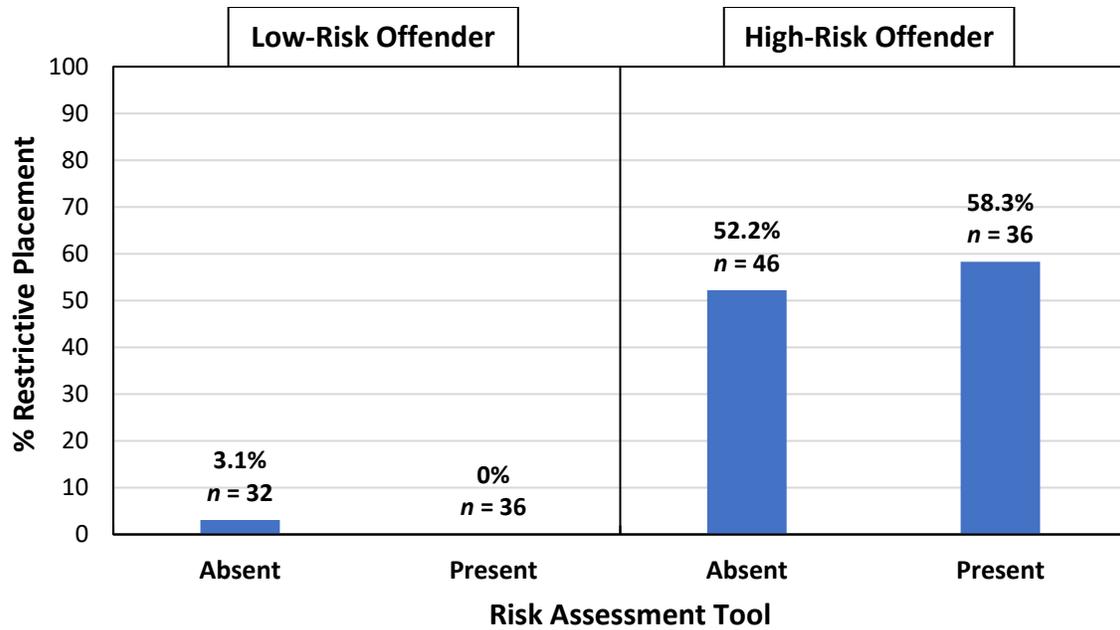


Figure 2. Bar graph showing the percentage of judges in each group who recommended a restrictive placement.

Note. The number of judges in each group are displayed below the percentages.

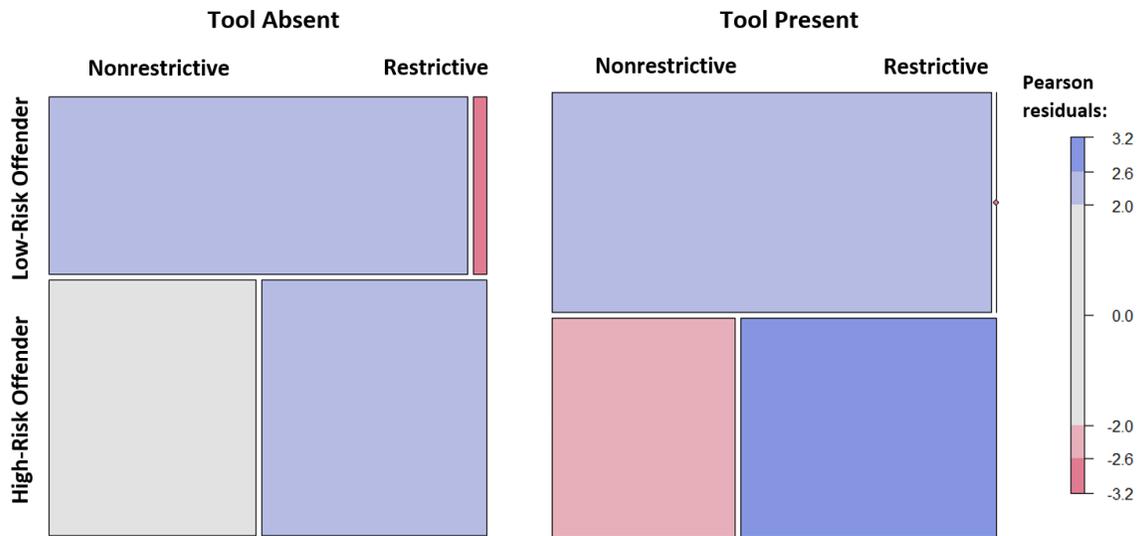


Figure 3. Mosaic plots showing the effect of risk level on judges' placement recommendations when a risk assessment tool was present or absent.

Note. Standardized residuals correspond to z-scores such that a value of ± 1.96 is significant at $p < .05$, a value of ± 2.58 is significant at $p < .01$, and a value of ± 3.29 is significant at $p < .001$.

Program intensity

Chi-square tests also revealed significant associations between risk level and program recommendation for both the tool-absent, $\chi^2(1) = 28.94, p < .001$, and tool-present, $\chi^2(1) = 50.16, p < .001$, group. Once again, I compared the size of the standardized residuals and odds ratios to assess whether there were meaningful differences between these associations. For the high-risk vignettes, 94.4% of judges recommended intensive programming (i.e., full-time attendance or residential treatment program) when a tool was present, compared to 77.8% of judges who did so when a tool was absent (see Figure 4). Standardized residuals were further from zero when a tool was present than when it was absent (see Figure 3). Specifically, the standardized residual for recommending intensive programming was 3.4 when a tool was present compared to 2.4 when a tool was absent. Odds ratios indicated that judges were 17 times more likely to recommend intensive programming than nonintensive programming when a tool was present, but only 3.5 times more likely when a tool was absent. Therefore, the analyses demonstrate that judges are more likely to recommend intensive programming for high-risk offenders when there was explicit reference to a risk assessment tool and an accompanying risk rating.

For low-risk vignettes, 11.1% of judges recommended more intensive programming when a tool was present, compared to 15.6% when a tool was absent (see Figure 4). Again, standardized residuals were further from zero when a tool was present than when it was absent (see Figure 5). The standardized residual for recommending intensive programming was -3.4 when a tool was present compared to -2.9 when a tool was absent. However, odds ratios were comparable across tool-present (OR = 0.13) and tool-absent (OR = 0.19) conditions. Results from the three-way loglinear analysis were consistent with the chi-square analyses. Specifically, the highest order interaction (i.e., the Tool Presence x Risk Level x Program Intensity interaction) was verging on significance, $\chi^2(1) = 3.63, p = 0.057$. Given the reduction in power caused by the assumption violation, it appears likely that there were legitimate differences in program recommendations between the tool-present and tool-absent conditions. Collectively, results from the separate chi-square analyses elucidate that the presence of a tool influenced judges program recommendations for the high-risk, but not the low-risk, offender, by making them more likely to recommend an intensive program.

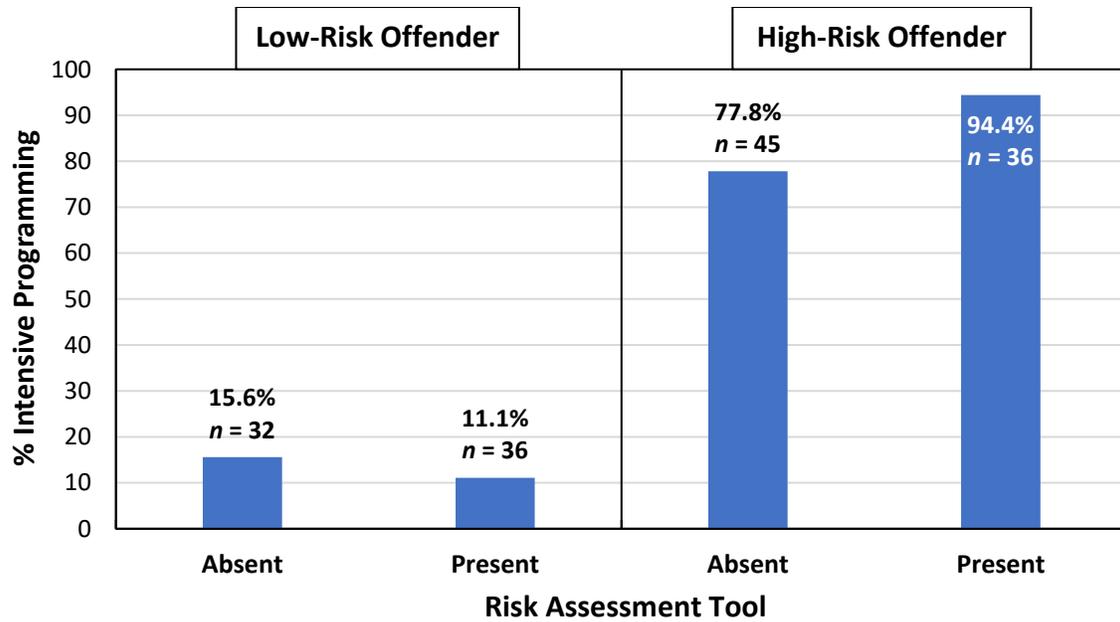


Figure 4. Bar graph showing the percentage of judges in each group who recommended intensive programming.

Note. The number of judges in each group are displayed below the percentages.

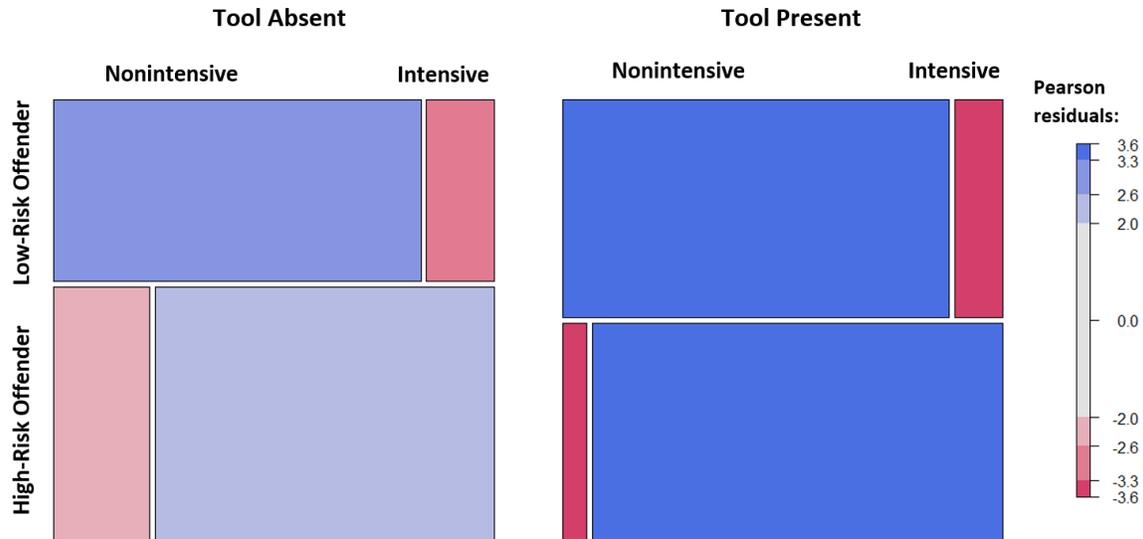


Figure 5. Mosaic plots showing the effect of risk level on judges' program recommendations when a risk assessment tool was present or absent.

Note. Standardized residuals correspond to z-scores such that a value of ± 1.96 is significant at $p < .05$, a value of ± 2.58 is significant at $p < .01$, and a value of ± 3.29 is significant at $p < .001$.

Research Question 4: Does the effect of a tool on judges' sentencing recommendations depend on how much emphasis judges think tools should be given?

Although the presence of a risk assessment tool did not influence placement recommendations, it is possible that the tool had a different effect for judges who believe tools should be given more weight when making sentencing decisions. Further, the effect of tool presence on program recommendations for the high-risk offender may be even stronger for judges who think tools should be given more weight. As such, judges were split into two groups: those who said tools should be given moderate or substantial emphasis (i.e., high-emphasis group) and those who said tools should be given minimal or no emphasis (i.e., low-emphasis group). I then prepared a set of loglinear analyses to determine whether patterns of placement or program recommendations differed as a function of judicial emphasis (high or low) and tool presence (present or absent). Analyses pertaining to placement and program recommendations for the low-risk offender were both underpowered, as indicated by the low expected cell counts in 50% of cells. The analysis pertaining to program recommendations for the high-risk offender was also underpowered (i.e., 37.5% of expected cell counts were under five). As such, those three models were not tested. However, all expected cell counts were above five for the analysis pertaining to placement recommendations for the high-risk offender, so the focus of this research question was narrowed and a loglinear analysis was performed for this model alone.

Although a visual inspection of the data suggested that the risk assessment tool had a stronger influence on judges who endorsed a greater emphasis on tools than on those who endorsed a lesser emphasis on tools (see Figure 6), results of the loglinear analysis indicated that neither the higher-order (i.e., three-way) nor lower-order (i.e., two-way) interactions were significant. Consequently, the effect of a tool on judges' sentencing recommendations did not appear to depend on how much emphasis judges thought tools should be given. Although the expected cell count assumption was met for this model, it is possible that the test was still underpowered given the small *ns*, especially in the low emphasis groups (*ns* were < 15).

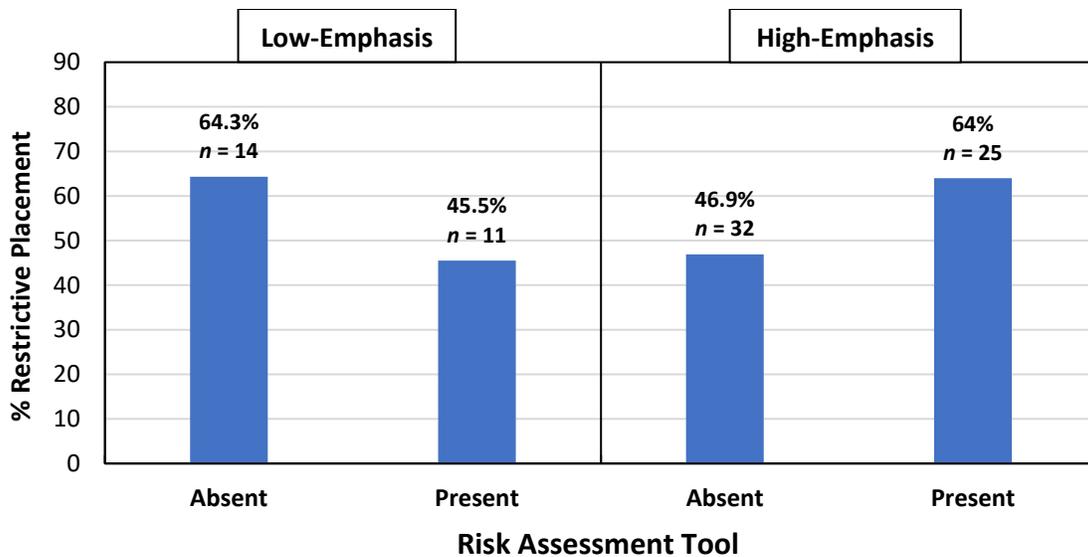


Figure 6. Bar graph showing the percentage of judges in each group who recommended a restrictive sentence for the high-risk offender.

Note. The number of judges in each group are displayed below the percentages.

Research Question 5: Do risk assessment tools increase the consistency of judges' sentencing recommendations?

Even if risk assessment tools do not influence the restrictiveness of judges' placement recommendations or the intensity of their program recommendations for most offenders, they may nevertheless increase the consistency of such recommendations. To test this hypothesis, I compared the standard deviations of judges' placement and program recommendations between conditions in which a tool was present and those in which it was absent.

Placement recommendations

Although the standard deviations of judges' placement recommendations appeared smaller in tool-present conditions than tool-absent conditions (see Table 7), the MSLRT test indicated that those differences were not significant, $MSLRT = 1.39, p = 0.707$. As such, the presence of a risk assessment tool (and accompanying risk rating) did not appear to increase consistency in judges' placement recommendations for the low-risk or high-risk offender.

Program recommendations

The standard deviations of judges' program recommendations were identical between tool-absent and tool-present conditions for the vignettes involving a low-risk offender (see Table 8). However, for those involving a high-risk offender, the standard deviation was lower when a risk assessment tool was present than when it was absent. The MSLRT confirmed that the standard deviation in the high-risk, tool-present group was significantly lower than those in the other groups, $MSLRT = 23.62, p < .001$. Follow-up pairwise comparisons revealed that it was significantly different from all three of the other groups. Most importantly for Research Question 4, the standard deviation was lower in the high-risk, tool-present condition ($SD = 0.53$) than in its tool-absent counterpart ($SD = 0.81$), $MSLRT = 10.35, p < .01$. As such, the presence of a risk assessment tool did increase consistency in judges' program recommendations for the high-risk, but not the low-risk, offender.

Table 7. Consistency in Judges' Placement Recommendations According to Vignette Condition

Risk level	RA tool	<i>n</i>	M	SD
Low	Present	38	0.81	0.40
	Absent	34	0.91	0.53
High	Present	41	1.78	0.83
	Absent	49	1.67	0.90

Table 8. Consistency in Judges' Program Recommendations According to Vignette Condition

Risk level	RA tool	<i>n</i>	M	SD
Low	Present	38	1.06	0.63
	Absent	34	1.16	0.63
High	Present	41	2.19	0.53***
	Absent	49	1.93	0.81

Note. The standard deviation in the high-risk, tool-present group was significantly lower than those in the other three groups.

*** $p < .001$

Chapter 4.

Discussion

This study combined a survey and experimental vignette design to examine (1) judges' opinions about using risk assessment tools to make sentencing decisions, (2) whether a risk assessment tool amplifies the association between offender risk level and judges' impressions of the offender, (3) whether a tool influences judges' sentencing decisions, (4) whether the effects of a tool are different according to how much emphasis judges think tools should be given, and (5) whether a tool increases consistency in judges' sentencing recommendations. The survey queried judges' opinions about risk assessment tools generally, while the vignette queried judges' placement and program recommendations for a hypothetical adolescent offender who was being sentenced for one count of Theft Under \$5000. The vignettes differed according to (a) whether the offender was described as having few or many risk factors and (b) whether the offender was identified as having a "high" or "low" risk to reoffend, as determined by a qualified clinician using an empirically-supported risk assessment tool. Findings pertaining to my main research questions are discussed below.

Judges had diverse opinions about using risk assessment tools to make sentencing decisions

Findings suggest that, as a group, judges view risk assessment information as somewhat pertinent to the sentencing process. In fact, more judges indicated that risk assessment tools should be used at the disposition/sentencing stage than at other points during the sanctioning process (i.e., pre-trial detention, adjudication, or parole/release). Consistent with extant research showing that judges view risk assessment tools as somewhat useful for making sentencing decisions (Shook & Sarri, 2007), judges in the current study agreed that tools are useful for making placement- and program-related decisions. Further, they agreed that tools increase fairness and consistency in sentencing and, more so, that it is important to consider future risk when making sentencing decisions. However, most judges reported that tool results should be given less weight than clinical opinion. This finding is somewhat concerning given that unstructured clinical judgement is more susceptible to human errors than structured

assessment techniques (Ægisdóttir et al., 2006; Monahan & Skeem, 2016). As a group, judges were undecided about whether tools lead to more restrictive or more lenient sentencing or whether tools reduce or contribute to minority confinement. These responses are perhaps not surprising given that research has only recently begun to assess whether risk assessment tools influence restrictive sanctioning or minority confinement (Viljoen et al., 2018).

Out of the potentially contentious factors that are sometimes included in tools, race was the only factor that judges were opposed to including in risk assessment tools. Public opinions appear similar, with most people expressing opposition to using race as a risk factor for recidivism and fewer expressing opposition to using age or gender (Scurich & Monahan, 2016). Interestingly, judges agreed most strongly that trauma history should be included in tools. Although there is a robust link between trauma history and violence risk (Holloway, Cruise, Morin, Kaufman, & Steele, 2018), it is controversial to include such history in risk assessment tools as it is not clear whether past trauma and victimization should (or does) act as a mitigating or aggravating factor in sentencing (Grisso, 2002; Monahan & Skeem, 2016; Stevenson, 2009). For example, Grisso (2002) warns that using childhood maltreatment to justify a more punitive sentence in the absence of rehabilitative efforts could represent “the final victimization of the victims of childhood maltreatment.” Stevenson’s (2009) literature review confirmed that, although legal decision-makers may sympathize with juvenile offenders who have a history of abuse, they tend to punish those offenders more harshly than their nonabused counterparts, perhaps due to risk factors that often co-occur with abuse (e.g., chaotic upbringing, school problems, hostility). Future research is needed to clarify how judges use trauma history to inform their decisions about placement and program referrals.

Although judges’ opinions about risk assessment tools were positive overall, their qualitative responses revealed a diverse range of views. Mirroring opinions expressed in the literature, some judges viewed risk assessment tools as an essential part of the sentencing process, whereas others viewed their use as problematic. The primary advantages expressed by judges were that tools provide information that is helpful for making decisions and that tools may lead to more objective and consistent decisions. The primary concerns raised by judges were that tools may involve bias, may lack predictive validity, and may obscure the individuality of offenders by using a one-size-fits-all approach. These opinions are consistent with, and add support to, those

expressed by judges in other studies (Chanenson & Hyatt, 2016; Monahan et al., 2018; Virginia Criminal Sentencing Commission, 2012).

It seems paradoxical that some judges believe tools reduce the risk of bias, while others believe they increase that risk. Although research demonstrates that, overall, empirically-supported tools tend to provide more accurate risk estimates than unstructured clinical judgement (Ægisdóttir et al., 2006; Monahan & Skeem, 2016), these contradictory findings suggest that it may also be important to consider how tool users and developers might introduce bias into tools. In their review of appellate court decisions, Krauss and Scurich (2013) highlighted that judges may question the scientific legitimacy of risk assessment tools because some clinicians, while perhaps well-intentioned, adjust actuarial risk estimates in a manner that attenuates predictive accuracy and impedes transparency. As such, tools may create a more insidious form of bias than unstructured clinical judgement. Future studies should explore whether certain tools are more susceptible to user biases and, further, how such biases might be reduced.

In addition to the advantages and concerns highlighted by judges, an important theme that emerged was that judges lack information about risk assessment tools. Indeed, several judges expressed a desire for more information about the predictive accuracy of tools, as well as systemic outcomes related to their implementation. Finally, several judges expressed the sentiment that risk assessment tools should be used solely as tools, not decision-making authorities, and that they are most useful when taken as one piece of evidence among many. This sentiment may reflect judges' concern, identified in other research (Dunnivant & Levitt, 2015), that limiting judicial discretion (e.g., by using authoritative, tool-based decision-making matrices) might reduce their ability to apply wisdom and flexibility to their decisions. As revealed in prior research (Monahan et al., 2018), it seems important to judges to clarify that tools should supplement, rather than replace, their judgement.

Risk assessment tools did not amplify associations between offender risk level and judges' impressions of the offender

The debate about pre-sentence risk assessment has focused largely on whether tools impact sentencing outcomes, and less attention has been devoted to the

mechanisms by which tools might influence sentencing decisions. As such, prior to examining whether tools affect sentencing recommendations, this study investigated whether tools might influence judges' impressions of an offender. Findings suggest that, although judges viewed the high-risk offender more negatively than the low-risk offender on some criteria (e.g., likelihood of lifetime criminality, potential for rehabilitation), the presence of a risk assessment tool and risk rating did not exacerbate those views. This finding is consistent with Murrie et al. (2007), who reported that descriptions of offenders may be more influential than diagnostic labels such as 'psychopathy'. However, given that this is the first study to assess whether risk assessment tools influence judges' impressions of offenders, future research should attempt to replicate this finding with larger and more diverse samples of judges.

Pending such examinations, this finding should help quell fears that risk assessment tools might exacerbate negative impressions of high-risk offenders or create harmful, irrevocable labels. However, it may still be concerning that judges rated the high-risk offender as more likely to be a lifelong criminal and less likely to be rehabilitated given that the offender in the vignette was a juvenile and given the transiency of risk during adolescence (Sweeten et al., 2013). As such, judges may benefit from more information about how trajectories of offending differ for adolescents compared to adults. That said, although judges viewed the high-risk offender more pessimistically than the low-risk offender, their mean ratings suggested that they still expressed moderate agreement that the high-risk offender could be rehabilitated and were undecided about whether he would be a lifelong criminal (i.e., they neither agreed nor disagreed with the statement, on average). Future research should investigate whether these findings generalize to adult offenders.

Risk assessment tools influenced program, but not placement, recommendations

As noted above, the primary question in the literature has been about the impact of risk assessment evidence on sentencing outcomes. As predicted, judges recommended more intensive programming for the high-risk offender when a tool was present than when it was absent. This provides some support for the notion that tools may increase adherence to the RNR model. However, tool presence did not affect program recommendations for the low-risk offender. In addition, although judges'

placement recommendations appeared slightly more polarized when a risk assessment tool was present, these differences were found to be negligible. This finding is perhaps not surprising given judges' comments (outlined above) that risk assessment tools are just *one* piece of evidence among many that they consider when making sentencing decisions. Indeed, a study that reviewed transcripts from adolescent offender cases found that when judges referred to a risk assessment tool in their sentencing decisions, it was often in the context of a multitude of other important factors, suggesting that tools are given only some weight (Urquhart & Viljoen, 2014). It is possible, therefore, that tools have a more minor effect on judicial decisions than some proponents and critics of tools had predicted.

That said, the current findings conflict somewhat with Viljoen et al. (2018), a systematic review and meta-analysis that found that tools modestly decreased restrictive placements, on average. By aggregating results across several studies, the meta-analysis would have had more power to detect small effects, ones that may have gone undetected in the current study. However, many studies included in the meta-analysis were unpublished, organizational reports and, in most, placements were determined by decision-makers other than judges (e.g., detention staff). Therefore, discrepancies between those results and the current ones may be understandable.

The current findings also conflict with van Wingerden et al. (2014), included in the above meta-analysis, who found that risk assessment tools led to less restrictive judicial sentencing decisions after controlling for a wide range of confounds. In this study, the authors proposed that tools may produce an 'information effect', whereby the information they provide allows decision-makers to consider an offender's personal circumstances that may act as mitigating factors. From this perspective, the main benefit of tools may lie in their ability to comprehensively unearth and integrate factors relevant to sentencing, rather than to provide uncontextualized predictions about an individual's risk to reoffend. Notably, the risk ratings provided in the current study were consistent with the descriptions of the offenders in the vignette (i.e., the offender identified as "high-risk" was described as having several risk factors, while the offender identified as "low-risk" was described as having few risk factors). Judges may, therefore, have been able to ascertain the offender's risk level based on the description alone and consequently produced similar placement recommendations regardless of tool presence.

Considering this potential explanation, some readers may argue that the information provided in the offender description would realistically have been obtained by using a risk assessment tool and, as such, the tool-absent and tool-present conditions were conflated. However, research suggests that experts will often refer to risk factors in court-ordered reports regardless of whether a structured risk assessment tool is completed (Wilson, Crocker, Nichols, Charette, & Seto, 2015). Furthermore, even when a tool-based risk assessment is conducted, risk factors are often described without clarifying their relevance to future risk (Storey, Watt, & Hart, 2015). As such, this study preemptively included a description of risk factors to help disentangle the effects of ambiguous descriptions of an offender from risk prediction statements that are explicitly based on an empirically-supported risk assessment tool. Further, a primary hypothesis in this study was that risk assessment tools would have a different moderating effect depending on whether the offender was low-risk or high-risk and, as such, it seemed necessary to clearly distinguish the high-risk and low-risk vignettes using empirically-supported risk factors. Nevertheless, it is possible that the benefit of tools lies in their ability to provide more accurate risk-related background information rather than their ability to clearly link that information to risk predictions. As such, more research is needed to understand differences between reports that are based on risk assessment tools and those that are based on clinical opinion alone.

Finally, another explanation for why the tool did not have the predicted effect on judges' placement decisions is that grouping all judges together may have obscured the effect that the tool may have had on the subset of judges who place more emphasis on tools. Although I attempted to test whether the tool had a stronger effect on judges who advocated for a greater emphasis on tools, assumption violations precluded all but one statistical test from being performed. That test did not reveal significant differences in placement recommendations for high-risk offenders; however, it may have suffered from low power and, hence, further research is necessary to understand whether tools are more persuasive for certain types of judges or in certain types of sentencing regimes.

Risk assessment tools increased the consistency of program, but not placement, recommendations

Although the presence of a risk assessment tool did not appear to increase the consistency of judges' placement recommendations, judges' program recommendations

for the high-risk offender appeared more consistent when a risk assessment tool was present than when it was absent. Specifically, the standard deviation in the high-risk, tool-present group was significantly smaller than those in the other groups. This finding provides some preliminary evidence for the suggestion that tools may increase consistency in sentencing (Skeem, 2013) and is consistent with judges' opinions (outlined above) that tools are more useful for making program/service decisions compared to other types of decisions. It seems likely that high-risk cases may be more complex and challenging than low-risk cases and that risk assessment tools may be more helpful in such cases when appropriate levels of programming are less apparent. Further, since high-risk offenders typically receive a greater loss of liberty, judges may require greater justification for their decisions in these cases. Indeed, research has found that judges rate risk prediction statements as more probative in high-risk scenarios compared to low-risk scenarios (Evans & Salekin, 2014; Kwartner et al., 2006).

Future studies should continue to explore whether risk assessment tools help increase the consistency of judges' program recommendations in more complex cases. Further, judges often make referrals to specific types of services (or set conditions that require offenders to attend specific programs) rather than just deciding on the intensity of such programming. As such, future studies should extend the current findings by exploring whether risk assessment tools help increase consistency when judges are asked to match programs to specific criminogenic needs (e.g., substance use treatment for offenders with substance use needs).

Strengths and limitations

This study was the first to my knowledge to experimentally manipulate the presence of a risk assessment tool using a sample of judges and, hence, increased my ability to draw causal conclusions about the influence of risk assessment tools on judges' sentencing decisions. Indeed, one drawback of file review studies is that offenders in the experimental and control groups can differ on other important variables such as details of their index offence and background (van Wingerden et al., 2014). Although some of these differences can be controlled for using techniques like propensity score matching, it can be difficult to determine whether all important variables have been identified. The current experimental vignette design avoided these concerns by allowing me to keep all variables other than risk level and tool presence constant.

Further, the qualitative information obtained by the accompanying survey helped explain the experimental findings and highlighted avenues for future research.

Although the current sample was large enough to detect moderate-to-large effects, it was likely underpowered to detect small effects. Even small effects may be meaningful in the context of sentencing decisions as they can have large impacts on system-wide outcomes such as incarceration rates. Besides sample size, there are several design issues that may have reduced power in the current study. For example, the low-risk offender may have been portrayed as having too few risk factors, creating a floor effect. Indeed, judges were unlikely to recommend a restrictive placement in either the tool-present or tool-absent conditions, which made it difficult to test differences in placement recommendations between those groups. In addition, the only difference between the tool-present and tool-absent conditions was that the tool-present vignettes provided a risk rating that was reportedly based on an empirically-supported tool. Since the tool-absent conditions did not explicitly state that a tool was not used, judges may have inferred that a tool was used, especially since they may have been primed by the prior questions about risk assessment tools. As such, the effect of the tool may have been suppressed.

Another limitation was that the current sample size precluded me from thoroughly testing Research Question 3 to assess whether the effect of tools on sentencing outcomes depends on judges' opinions about how much emphasis should be placed on tools. As such, future research should attempt to replicate the current findings with a larger sample. Further, the current sample consisted predominantly of American judges and may not generalize outside of North America. It will be important to test the current research questions with judges from other countries, especially countries in which risk assessment tools might be used differently by the courts. For example, lawyers who operate in an inquisitorial legal system may be less prone to use risk assessment evidence to argue for a harsher sentence compared to lawyers operating in an adversarial system. Relatedly, given the low response rate in this study, it is unclear whether judges who responded differ in important ways from those who did not.

A final limitation of this study related to the lack of a pre-existing measure to assess impressions of the offender. Although I used a recommended approach to develop the questionnaire and was informed by similar measures used in other studies,

it would nevertheless have been preferable to use a measure that had undergone validity and reliability testing in prior studies. For example, if an empirically-supported tool had been available, I may have been able to compute a total score representing overall impressions of the offender, which would have increased my ability to detect significant effects by removing the need to use Bonferroni corrections for multiple tests. Of course, the benefit of using single items is that they are tied to concrete factors (e.g., rehabilitation potential) rather than latent variables (e.g., stigma).

Implications

To what extent should risk assessment tools be used to guide sentencing decisions? I began this thesis by reviewing arguments for and against the inclusion of risk assessment evidence in sentencing proceedings. Results from this study support findings from past research showing that judges view tools as useful and informative for making placement and program decisions. It also extends prior research by showing that judges' program recommendations for high-risk offenders are more intensive and more consistent when tool evidence is present than when it is absent. Further, the presence of a tool did not exacerbate negative perceptions of the high-risk offender or alter the restrictiveness of sentence recommendations. Collectively, then, these findings lend some support to the inclusion of risk assessment tool findings in sentencing proceedings.

That said, the lack of significant associations between tool presence and risk level on placement recommendations suggest that tools may not have the sizable impact on sanctioning practices that advocates might hope. As such, these findings suggest that organizations should temper their expectations about the ability of tools to reduce overincarceration on their own and should avoid viewing tools as the sole solution for such problems. Recent research suggests that implementing other strategies alongside tools may help increase their impact (Stevenson, in press; Viljoen et al., 2018). For instance, tools may be more effective if important stakeholders, such as judges, are involved in their implementation (Vincent et al., 2016). However, given that the judges in this study raised some legitimate concerns about tools that have not been sufficiently addressed by research (e.g., tools may obscure biases), it seems paramount to consider what degree of compliance should be expected from judges at this time (Monahan & Skeem, 2014). For now, involving judges in the development or selection of tools for

their jurisdiction may help address judges' concerns at the front end of implementation and ensure that tools are appropriate for the population served by each court. Legal organizations and researchers should also continue to collect information about sentencing and incarceration outcomes following the implementation of a tool (e.g., impact on racial disparity) to help respond to judges' concerns about tools.

Regarding implications for research, the current findings suggest that future studies should pay more attention to how risk assessment tools might influence the consistency of judicial decisions. Even if tools do not increase or decrease the restrictiveness of sanctions, they may benefit the sanctioning process if they can increase judicial agreement about cases. In addition, although only a few judges identified increased transparency as a benefit of tools, it would be interesting to explore offender and public perceptions about whether tools facilitate a greater sense of trust and confidence in the justice system. Finally, more research is needed to understand the implications of including trauma in risk assessment tools to ensure that victims of abuse are not re-victimized through the justice system. As research in this area progresses, efforts should be made to disseminate results to judges (e.g., by preparing fact sheets or brief reports for judicial agencies and presenting papers at judicial conferences), so they can not only stay informed about advances in the field but also influence future directions.

Conclusion

Considerable discourse has arisen from the proliferation of risk assessment tools in the courtroom that has occurred over the past few decades. This study was designed to better understand judges' views on this issue and, further, to test some of the main arguments presented by advocates and critics of using tools to guide sentencing decisions. Interestingly, judges' opinions about risk assessment tools echoed those expressed in the literature. They recognized that tools may provide more information about offenders and may reduce bias in decision-making. However, they also expressed concerns that tools may simply shift bias from judges to the professionals who develop and administer tools, thereby making the bias more insidious.

These concerns are perhaps tempered by the fact that, in this study, tools did *not* appear to significantly influence judges' impressions of the offender or the restrictiveness

of their placement recommendations. Instead, the only apparent effects of the tool were that it increased the intensity and consistency of program recommendations for the high-risk offender. As such, findings from this study suggest that proponents of risk assessment tools may overestimate their benefits for making placement decisions and critics may overestimate their risks. Most likely, tools play a subtler role in the sentencing process and may have more to do with increasing judicial agreement about appropriate levels of programming for complex cases (e.g., offenders with multiple risk factors) than with increasing or decreasing the restrictiveness of their sanctions.

As stated by a judge in this study, “the field of risk assessment tools, while it has developed over the last several years, still has a ways to go” and, indeed, researchers and judges alike still have much to learn about the implications of using tools in sentencing. Although, overall, judges shared similar hopes and qualms about risk assessment tools as those expressed in the research literature, their views ranged considerably. It will be important, going forward, to recognize the diversity in judicial opinions about risk assessment tools and to involve judges in the conversation about how the application of such instruments can be improved. Together, I believe we can make more informed decisions about our use of risk assessment evidence in sentencing and help ensure that tools are being used to equitably and responsibly reduce risk, rather than simply predict it.

References

- Andrews, D. A., & Bonta, J. (1995). *The Level of Service Inventory–Revised*. Toronto, Canada: Multi-Health Systems.
- Andrews, D. A., & Dowden, C. (2006). Risk principle of case classification in correctional treatment: A meta-analytic investigation. *International Journal of Offender Therapy and Comparative Criminology*, *50*, 88-100.
doi:10.1177/0306624X05282556
- Andrews, D. A., Bonta, J., & Hoge, R. D. (1990). Classification for effective rehabilitation: Rediscovering psychology. *Criminal Justice and Behavior*, *17*, 19-52.
doi:10.1177/0093854890017001004
- Andrews, D. A., Bonta, J. L., & Wormith, J. S. (2004). *Level of Service/Case Management Inventory (LS/CMI): User's manual*. Toronto, Canada: Multi-Health Systems.
- Annie E. Casey Foundation (2017). Juvenile Detention Alternatives Initiative (JDAI) Insights from the Annual Results Report. Annie E. Casey Foundation, Baltimore, Maryland.
- Baron, J. (2008). *Thinking and deciding* (4th ed.). New York, NY: Cambridge University Press.
- Blais, J., & Forth, A. E. (2014). Potential labeling effects: Influence of psychopathy diagnosis, defendant age, and defendant gender on mock jurors' decisions. *Psychology, Crime & Law*, *20*, 116-134.
doi:10.1080/1068316X.2012.749473
- Bonta, J. (2007). Offender risk assessment and sentencing. *Canadian Journal of Criminology and Criminal Justice*, *49*, 519-529. doi:10.3138/cjccj.49.4.519
- Bonta, J., Bourgon, G., Jesseman, R., & Yessine, A. K. (2005). *Presentence reports in Canada*. (User Report 2005-03). Ottawa, Canada: Public Canada. Retrieved from <http://www.publicsafety.gc.ca/cnt/rsrscs/pblctns/prsntnc-rprts-cnd/index-en.aspx>
- Brauer, P. M., Hanning, R. M., Arocha, J. F., Royall, D., Goy, R., Grant, A., Dietrich, L., Martino, R., & Horrocks, J. (2009). Creating case scenarios or vignettes using factorial study design methods. *Journal of Advanced Nursing*, *65*, 1937-1945.
doi:10.1111/j.13652648.2009.05055.x

- Chanenson, S. L. & Hyatt, J. M. (2016). *The use of risk assessment at sentencing: Implications for research and policy*. (U.S. Department of Justice White Paper). Retrieved from <https://digitalcommons.law.villanova.edu/cgi/viewcontent.cgi?article=1201&context=wps>
- Cole, D. P. (2007). The umpires strike back: Canadian judicial experience with risk-assessment instruments. *Canadian Journal of Criminology and Criminal Justice*, 49, 493-517. doi:10.3138/cjccj.49.4.493
- Creswell, J.W. (2013). *Qualitative inquiry and research design: Choosing among five approaches*. Los Angeles, CA: SAGE Publications.
- DeMatteo, D., Wolbransky, M., & LaDuke, C. (2016). Risk assessment with juveniles. In K. Heilbrun, D. DeMatteo, & N. S. Goldstein (Eds.), *APA handbook of psychology and juvenile justice* (pp. 365-384). Washington, DC: American Psychological Association. doi:10.1037/14643-017
- Dillman, D. A., Smyth, J. D., & Christian, L. M. (2009). *Internet, mail, and mixed-mode surveys: The tailored design method., 3rd ed.* Hoboken, NJ, US: John Wiley & Sons Inc.
- Dmitrieva, J., Monahan, K. C., Cauffman, E., & Steinberg, L. (2012). Arrested development: The effects of incarceration on the development of psychosocial maturity. *Development and Psychopathology*, 24, 1073-1090. doi:10.1017/S0954579412000545
- Dunnavant, B., & Levitt, H. (2015). The development of wisdom in judicial decision-making. *The Humanistic Psychologist*, 43, 1-23. doi:10.1080/08873267.2014.993068
- Ægisdóttir, S., White, M. J., Spengler, P. M., Maugherman, A. S., Anderson, L. A., Cook, R. S., & ... Rush, J. D. (2006). The Meta-Analysis of Clinical Judgment Project: Fifty-Six Years of Accumulated Research on Clinical Versus Statistical Prediction. *The Counseling Psychologist*, 34, 341-382. doi:10.1177/0011000005285875
- Elek, J. K., Warren, R. K., & Casey, P. M. (2015). *Using risk and needs assessment information at sentencing: Observations from ten jurisdictions*. Williamsburg, VA: National Center for State Courts. Retrieved from <https://www.ncsc.org/~media/Microsites/Files/CSI/RNA%202015/Final%20PEW%20Report%20updated%2010-5-15.ashx>
- Evans, S. A., & Salekin, K. L., (2014). Involuntary civil commitment: Communicating with the court regarding “danger to other”. *Law and Human Behavior*, 38, 325-336. doi:10.1037/lhb0000068

- Evans, S. A., & Salekin, K. L. (2016). Violence risk communication: What do judges and forensic clinicians prefer and understand? *Journal of Threat Assessment and Management*, 3(3-4), 143-164. doi:10.1037/tam0000062
- Faul, F., Erdfelder, E., Lang, A.-G., & Buchner, A. (2007). G*Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods*, 39, 175-191. doi:10.3758/BF03193146
- Field, A., Miles, J., & Field, Zoë. (2012). *Discovering statistics using R*. Thousand Oaks, CA: SAGE Publications.
- Goldkamp, J. S., & Gottfredson, M. R. (1985). *Policy guidelines for bail: An experiment in court reform*. Philadelphia: Temple University Press.
- Grisso, T. (2002). Using what we know about child maltreatment and delinquency. *Children's Services: Social Policy, Research, & Practice*, 5, 299–305. doi:10.1207/S15326918CS0504pass:[_]6
- Hannah-Moffat, K. (2011). *Actuarial Sentencing: An "unsettled" proposition*. Unpublished manuscript. Retrieved from http://www.albany.edu/scj/documents/Hannah-Moffatt_RiskAssessment_000.pdf
- Harcourt, B. E. (2015). Risk as a Proxy for Race: The Dangers of Risk Assessment. *Federal Sentencing Reporter*, 27(4), 237-243. doi:10.1525/fsr.2015.27.4.237
- Hart, S. D., & Cooke, D. J. (2013). Another look at the (im-)precision of individual risk estimates made using actuarial risk assessment instruments. *Behavioral Sciences & The Law*, 31, 81-102. doi:10.1002/bsl.2049
- Heilbrun, K., DeMatteo, D., Holliday, S., & LaDuke, C. (2014). *Forensic Mental Health Assessment: A Casebook* (Second Edition). Oxford: Oxford University Press.
- Hilinski-Rosick, C. M., Freiburger, T. L., & Verheek, A. (2014). The effects of legal and extralegal variables on the sentences of sex offenders. *Victims & Offenders*, 9, 334-351. doi:10.1080/15564886.2013.845637
- Hilton, N. Z., & Simmons, J. L. (2001). The influence of actuarial risk assessment in clinical judgments and tribunal decisions about mentally disordered offenders in maximum security. *Law and Human Behavior*, 25, 393–408. doi: 10.1023/A:1010607719239
- Holloway, E. D., Cruise, K. R., Morin, S. L., Kaufman, H., & Steele, R. D. (2018). Juvenile probation officers' evaluation of traumatic event exposures and traumatic stress symptoms as responsivity factors in risk assessment and case planning. *Law and Human Behavior*, 42, 369–384. <https://doi-org.proxy.lib.sfu.ca/10.1037/lhb0000283>

- Holmbeck, G. N., & Devine, K. A. (2009). Editorial: An author's checklist for measure development and validation manuscripts. *Journal of Pediatric Psychology, 34*(7), 691-696. doi:10.1093/jpepsy/jsp046
- Hsieh, H., & Shannon, S. (2005). Three Approaches to Qualitative Content Analysis. *Qualitative Health Research, 15*, 1277-1288. doi:10.1177/1049732305276687
- Imrey, P. B., & Dawid, A. P. (2015). A commentary on statistical assessment of violence recidivism risk. *Statistics and Public Policy, 2*, 1-18. doi:10.1080/2330443X.2015.1029338
- Isaacs, D. (2011). Baseline Framing in Sentencing. *The Yale Law Journal, 121*, 426-458. Retrieved from <http://www.jstor.org.proxy.lib.sfu.ca/stable/23079336>
- Jung, S., Brown, K., Ennis, L., & Ledi, D. (2015). The association between presentence risk evaluations and sentencing outcome. *Applied Psychology in Criminal Justice, 11*, 111–125. Retrieved from <http://proxy.lib.sfu.ca/login?url=https://search.ebscohost.com/login.aspx?direct=true&db=psyh&AN=2016-07720-004&site=ehost-live>
- Krauss, D. A., & Scurich, N. (2013). Risk assessment in the law: Legal admissibility, scientific validity, and some disparities between research and practice. *Behavioral Sciences & the Law, 31*, 215–229. doi:10.1002/bsl.2065
- Krishnamoorthy, K., & Lee, M. (2014). Improved tests for the equality of normal coefficients of variation. *Computational Statistics, 29*, 215-232. doi:10.1007/s00180-013-0445-2
- Kroh, M. (2007). Measuring Left–Right Political Orientation: The Choice of Response Format. *Public Opinion Quarterly, 71*, 204-220. doi:10.1093/poq/nfm009
- Kwartner, P. P., Lyons, P. J., & Boccaccini, M. T. (2006). Judges' risk communication preferences in risk for future violence cases. *The International Journal of Forensic Mental Health, 5*, 185-194. doi:10.1080/14999013.2006.10471242
- Lambie, I., & Randell, I. (2013). The impact of incarceration on juvenile offenders. *Clinical Psychology Review, 33*, 448-459. doi:10.1016/j.cpr.2013.01.007
- Large, M., Ryan, C., & Nielssen, O. (2012). Minority report on violence risk assessment. *The Psychiatrist, 36*(1), 36. doi:10.1192/pb.36.1.36
- Latessa, E. J., Lovins, L. B., & Smith, P. (2010). Follow-up evaluation of Ohio's community based correctional facility and halfway house programs – outcome study. Retrieved from https://www.uc.edu/content/dam/uc/ccjr/docs/reports/project_reports/2010%20HWH%20Executive%20Summary.pdf

- Laura and John Arnold Foundation (2016). New data: Pretrial risk assessment tool works to reduce crime, increase court appearances [Press Release]. Laura and John Arnold Foundation. Retrieved from <http://www.arnoldfoundation.org/new-data-pretrial-risk-assessment-tool-works-reduce-crime-increase-court-appearances/>
- Maurutto, P., & Hannah-Moffat, K. (2007). Understanding risk in the context of the Youth Criminal Justice Act. *Canadian Journal of Criminology and Criminal Justice*, 49, 465-491. doi:10.3138/cjccj.49.4.465
- Maloney, C., & Miller, J. (2015). The impact of a risk assessment instrument on juvenile detention decision-making: A check on "perceptual shorthand" and "going rates"? *Justice Quarterly*, 32, 900-927. doi:10.1080/07418825.2013.863961
- Marwick, B., & Krishnamoorthy, K. (2018). cvequality: tests for the equality of coefficients of variation from multiple groups [R software package version 0.1.3]. Retrieved from <https://github.com/benmarwick/cvequality>
- Monahan, J., Metz, A. L., & Garrett, B. L. (2018). Judicial appraisals of risk assessment in sentencing. *Behavioral Sciences & the Law*. doi:10.1002/bsl.2380
- Monahan, J., & Skeem, J. L. (2014). Risk redux: The resurgence of risk assessment in criminal sanctioning. *Federal Sentencing Reporter*, 26, 158-166. doi:10.1525/fsr.2014.26.3.158
- Monahan, J., & Skeem, J. L. (2016). Risk assessment in criminal sentencing. *Annual Review of Clinical Psychology*, 12, 489-513. doi:10.1146/annurev-clinpsy-021815-092945
- Mowle, E. N., Edens, J. F., Clark, J. W., & Sörman, K. (2016). Effects of mental health and neuroscience evidence on juror perceptions of a criminal defendant: The moderating role of political orientation. *Behavioral Sciences & the Law*, 34, 726-741. doi:10.1002/bsl.2251
- Muheizen, J. (2014). *The effect of risk assessment data and colleagues' consensus on clinical decisions to discharge insanity acquittees* (Doctoral Dissertation). Retrieved from <http://escholarship.org/uc/item/83m9n3hx>
- Murrie, D. C., Boccaccini, M. T., McCoy, W., & Cornell, D. G. (2007). Diagnostic labeling in juvenile court: How do descriptions of psychopathy and conduct disorder influence judges? *Journal of Clinical Child and Adolescent Psychology*, 36, 228-241. doi:10.1080/15374410701279602
- National Crime Prevention Centre. (2012). *A statistical snapshot of youth at risk and youth offending in Canada* (Catalogue No. PS4-126/2012E-PDF). Retrieved from <https://www.publicsafety.gc.ca/cnt/rsrscs/pblctns/ststclsnpst-yth/ssyr-eng.pdf>

- Nolan, A. & Stewart, L. (2017). *Low risk offenders: What does the research tell us?* (Correctional Service of Canada Research Report No. R-383). Retrieved from http://publications.gc.ca/collections/collection_2017/scc-csc/PS83-3-383-eng.pdf
- O'Connell, A. (2006). *Logistic Regression Models for Ordinal Response Variables*. Thousand Oaks, CA: SAGE Publications.
- Peterson-Badali, M., Skilling, T., & Haqanee, Z. (2015). Examining implementation of risk assessment in case management for youth in the justice system. *Criminal Justice and Behavior*, *42*, 304-320. doi:10.1177/0093854814549595
- Penner, E. K. (2009). *Judges' perceptions of juveniles who assert their right to silence* (Master's thesis). Retrieved from <http://summit.sfu.ca/item/9645>
- Quinsey, V. L., Harris, G. T., Rice, M.E., & Cormier, C. (1998). *Violent offenders: Appraising and managing risk*. Washington, DC: American Psychological Association.
- Salekin, R. T., Yff, R. A., Neumann, C. S., Leistico, A. R., & Zalot, A. A. (2002). Juvenile transfer to adult courts: A look at the prototypes for dangerousness sophistication-maturity and amenability to treatment through a legal lens. *Psychology, Public Policy, and Law*, *8*, 373-410. doi:10.1037/1076-8971.8.4.373
- Scurich, N., & Monahan, J. (2016). Evidence-based sentencing: Public openness and opposition to using gender, age, and race as risk factors for recidivism. *Law and Human Behavior*, *40*, 36-41. doi:10.1037/lhb0000161
- Shook, J. J., & Sarri, R. C. (2007). Structured decision making in juvenile justice: Judges' and probation officers' perceptions and use. *Children and Youth Services Review*, *29*, 1335–1351. doi:10.1016/j.childyouth.2007.05.008
- Sickmund, M., & Puzanchera, C. (2014). *Juvenile offenders and victims: 2014 national report*. Pittsburgh, VA: National Center for Juvenile Justice. Retrieved from <https://www.ojdp.gov/ojstatbb/nr2014/downloads/NR2014.pdf>
- 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. *The International Journal of Forensic Mental Health*, *13*, 193-206. doi:10.1080/14999013.2014.922141
- Skeem, J. (2013). Risk Technology in Sentencing: Testing the Promises and Perils (Commentary on Hannah-Moffat, 2011). *Justice Quarterly*, *30*, 297-303. doi:10.1080/07418825.2012.687513
- Skeem, J. L., & Monahan, J. (2011). Current directions in violence risk assessment. *Current Directions in Psychological Science*, *20*, 38-42. doi:10.1177/0963721410397271

- Smith, S. T., Edens, J. F., Clark, J., & Rulseh, A. (2014). 'So, what is a psychopath?' Venireperson perceptions, beliefs, and attitudes about psychopathic personality. *Law and Human Behavior*, 38, 490-500. doi:10.1037/lhb0000091
- Sörman, K., Edens, J. F., Smith, S. T., Svensson, O., Howner, K., Kristiansson, M., & Fischer, H. (2014). Forensic mental health professionals' perceptions of psychopathy: A prototypicality analysis of the Comprehensive Assessment of Psychopathic Personality in Sweden. *Law and Human Behavior*, 38, 405-417. doi:10.1037/lhb0000072
- Starr, S. B. (2014). Evidence-based sentencing and the scientific rationalization of discrimination. *Stanford Law Review*, 66, 803–872. Retrieved from <https://www.stanfordlawreview.org/print/article/evidence-based-sentencing-and-the-scientific-rationalization-of-discrimination/>
- Stevenson, M. C. (2009). Perceptions of juvenile offenders who were abused as children. *Journal of Aggression, Maltreatment & Trauma*, 18, 331–349. <https://doi-org.proxy.lib.sfu.ca/10.1080/10926770902901428>
- Stevenson, M. T. (in press). Assessing Risk Assessment in Action. *Minnesota Law Review*, 103. Retrieved from https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3016088
- Storey, J. E., Watt, K. A., & Hart, S. D. (2015). An examination of violence risk communication in practice using a structured professional judgment framework. *Behavioral Sciences & The Law*, 33, 39-55. doi:10.1002/bsl.2156
- Sweeten, G., Piquero, A. R., & Steinberg, L. (2013). Age and the explanation of crime, revisited. *Journal of Youth and Adolescence*, 42, 921-938.
- Thompson, C. (2017). *Myths and factors: Using risk and need assessments to enhance outcomes and reduce disparities in the criminal justice system*. Washington, DC: National Institute of Corrections, Community Corrections Collaborative Network. Retrieved from <https://info.nicic.gov/nicrp/system/files/032859.pdf>
- Tonry, M. (2014). Legal and Ethical Issues in the Prediction of Recidivism. *Federal Sentencing Reporter*, 26(3), 167-176. doi:10.1525/fsr.2014.26.3.167
- Urquhart, T. A., & Viljoen, J. L. (2014). The use of the SAVRY and YLS/CMI in adolescent court proceedings: A case law review. *The International Journal of Forensic Mental Health*, 13, 47-61. doi:10.1080/14999013.2014.885470
- van Wingerden, S., van Wilsem, J., & Moerings, M. (2014). Pre-sentence reports and punishment: A quasi-experiment assessing the effects of risk-based pre-sentence reports on sentencing. *European Journal of Criminology*, 11, 723-744. doi:10.1177/1477370814525937

- Virginia Sentencing Commission (2012). 2012 Annual Report. Richmond, Virginia: Virginia Sentencing Commission. Retrieved from <http://www.vcsc.virginia.gov/reports.html>
- Viljoen, J. L., Jonnson, M., Cochrane, D., & Vargen, L. (2018). *Do Risk Assessment Tools Decrease Incarceration and Restrictive Placements for Offenders? A Systematic Review and Meta-Analysis*. Manuscript in preparation.
- Viljoen, J. L., MacDougall, E. M., Gagnon, N. C., & Douglas, K. S. (2010). Psychopathy evidence in legal proceedings involving adolescent offenders. *Psychology, Public Policy, and Law*, 16, 254-283. doi:10.1037/a0019649
- Viljoen, J. L. & McLachlan, K. (2016). *The Perceived Value of Risk Assessments in Legal Proceedings: A Survey of Judges*. Unpublished manuscript.
- Vincent, G. M., Guy, L. S., Perrault, R. T., & Gershenson, B. (2016). Risk assessment matters, but only when implemented well: A multisite study in juvenile probation. *Law and Human Behavior*, 40, 683-696. doi:10.1037/lhb0000214
- Wachter, A. (2015). *Statewide risk assessment in juvenile probation: JJGPS StateScan*. Pittsburgh, PA: National Center for Juvenile Justice.
- Webster, C. D., Douglas, K. S., Eaves, D., & Hart, S. D. (1997). *The HCR-20: Assessing the risk for violence (Version 2)*. Burnaby, Canada: Mental Health, Law and Policy Institute, Simon Fraser University.
- Weiner, B., & Kukla, A. (1970). An attributional analysis of achievement motivation. *Journal of Personality and Social Psychology*, 15, 1-20.
- Wilson, C., Crocker, A., Nicholls, T., Charette, Y., & Seto, M. (2015). The use of risk and need factors in forensic mental health decision-making and the role of gender and index offense severity. *Behavioral Sciences & the Law*, 33, 19-38. doi:10.1002/bsl.2162
- Wise, R. A., & Safer, M. A. (2004). What US Judges Know and Believe About Eyewitness Testimony. *Applied Cognitive Psychology*, 18, 427-443. doi:10.1002/acp.993
- Youth Criminal Justice Act*, SC 2002. c. 1

Appendix A.

Survey

Legal Attitudes

In your opinion, what should the main purpose of imposing a sentence/disposition be?						
	Strongly Disagree	Somewhat Disagree	Slightly Disagree	Slightly Agree	Somewhat Agree	Strongly Agree
Punishment/Retribution	0	1	2	3	4	5
Individual deterrence	0	1	2	3	4	5
Societal deterrence	0	1	2	3	4	5
Rehabilitation	0	1	2	3	4	5
Protection of society	0	1	2	3	4	5

Opinions about Risk Assessment Tools

Risk assessment tools use a combination of risk-related factors (e.g., offense history, procriminal attitudes) to help evaluators estimate an offender's likelihood of engaging in future violence or other offending. To what extent do you agree with the following statements about risk assessment tools?						
	Strongly Disagree	Somewhat Disagree	Slightly Disagree	Slightly Agree	Somewhat Agree	Strongly Agree
Risk assessment tools predict an offender's risk for reoffending with reasonable accuracy.	0	1	2	3	4	5
Risk assessment tools should be used primarily to identify the lowest risk offenders (i.e., to reduce their sanction).	0	1	2	3	4	5
Risk assessment tools should be used primarily to identify the highest risk offenders (i.e., to increase their sanction).	0	1	2	3	4	5
Risk assessment tools are useful for making decisions about the length of an offender's sentence.	0	1	2	3	4	5

Risk assessment tools are useful for making decisions about where to place an offender (e.g., in custody or in the community).	0	1	2	3	4	5
Risk assessment tools are useful for making decisions about an offender's supervision level.	0	1	2	3	4	5
Risk assessment tools are useful for making decisions about program/service referrals.	0	1	2	3	4	5
Risk assessment tools lead to more restrictive sentencing (i.e., more offenders being incarcerated).	0	1	2	3	4	5
Risk assessment tools lead to less restrictive sentences (i.e., fewer offenders being incarcerated).	0	1	2	3	4	5
Risk assessment tools increase fairness and consistency in sentencing.	0	1	2	3	4	5
Risk assessment tools decrease fairness and consistency in sentencing.	0	1	2	3	4	5
Risk assessment tools help reduce the overrepresentation of minorities in jail/prison.	0	1	2	3	4	5
Risk assessment tools contribute to the overrepresentation of minorities in jail/prison.	0	1	2	3	4	5
It is inappropriate to sentence someone based on what they might do in the future.	0	1	2	3	4	5
It is important to consider future risk when making sentencing decisions because it is relevant to public protection and safety.	0	1	2	3	4	5
We should be more cautious when using risk assessments to inform juvenile dispositions (e.g., length of disposition) than adult sentences.	0	1	2	3	4	5

Risk assessment tools include a variety of factors. In addition to criminal history, some tools include demographic and socioeconomic factors such as age, education, and employment. To what extent do you agree that the following risk factors should be included in risk assessment tools?

	Strongly Disagree	Somewhat Disagree	Slightly Disagree	Slightly Agree	Somewhat Agree	Strongly Agree
Gender	0	1	2	3	4	5
Race	0	1	2	3	4	5
Age	0	1	2	3	4	5
History of trauma	0	1	2	3	4	5
Educational attainment	0	1	2	3	4	5
Employment status	0	1	2	3	4	5

When making placement decisions (e.g., custodial vs. non-custodial), how much emphasis should be placed on risk assessment tool results?

No emphasis	Minimal	Moderate	Substantial
-------------	---------	----------	-------------

Compared to clinical opinion, how much emphasis do you place on risk assessment tool results?

Less emphasis	Equal emphasis	More emphasis
---------------	----------------	---------------

At what point(s) during the sanctioning process should a risk assessment tool be used (check all that apply)?

Pre-trial detention	Adjudication	Disposition/Sentencing	Parole/Release
---------------------	--------------	------------------------	----------------

What do you think are the pros and cons of using risk assessment tools to help make decisions about supervision, placements, or services?

Vignette

(Sample A – Low-risk with tool present)

Steve is a 16-year-old male who has been charged with theft under \$5000 because he stole a pellet gun and night vision goggles valued at \$4,500 from an army surplus store. Police officers caught Steve with the stolen equipment shortly after the time of the offense. He has entered a guilty plea for the offense. The only offense on his record is one prior adjudication for shoplifting (theft under \$5000) that involved stealing a hunting knife from a sporting goods store. Steve pleaded guilty for that offense and was assigned a 6-month probation term.

Throughout his prior supervision period, Steve was polite and did not violate any of his conditions. According to family members and peers, he has a gentle nature and has never demonstrated aggressive behaviour even when provoked. Steve attends school regularly and is described by his teachers as a good student. He currently lives at home with his mother and older brother, whom he reports close relationships with. Steve looks up to his brother because he gets “straight As” and “has a lot of friends”. Further, he characterizes his home as nurturing and “very supportive”. Regarding his current offense, Steve appears genuinely remorseful and feels “really bad for stealing that stuff”. He reports that he uses alcohol infrequently (about four times per year) and has not tried any other substances. Steve has been employed as a line cook for the past two years and has never been fired from a job.

A qualified clinician evaluated Steve’s risk for future violence and reoffending with a widely-used risk assessment instrument that has accrued empirical support. The instrument consists of factors associated with recidivism, which the clinician uses to form an opinion about the offender’s risk. The results of the risk assessment suggest that Steve is at a low risk for reoffending.

(Sample B – Low-risk with tool absent)

Steve is a 16-year-old male who has been charged with theft under \$5000 because he stole a pellet gun and night vision goggles valued at \$4,500 from an army surplus store. Police officers caught Steve with the stolen equipment shortly after the time of the offense. He has entered a guilty plea for the offense. The only offense on his record is one prior adjudication for shoplifting (theft under \$5000) that involved stealing a hunting knife from a sporting goods store. Steve pleaded guilty for that offense and was assigned a 6-month probation term.

Throughout his prior supervision period, Steve was polite and did not violate any of his conditions. According to family members and peers, he has a gentle nature and has never demonstrated aggressive behaviour even when provoked. Steve attends school regularly and is described by his teachers as a good student. He currently lives at home with his mother and older brother, whom he reports close relationships with. Steve looks up to his brother because he gets “straight As” and “has a lot of friends”. Further, he characterizes his home as nurturing and “very supportive”. Regarding his current offense, Steve appears genuinely remorseful and feels “really bad for stealing that stuff”. He reports that he uses alcohol infrequently (about four times per year) and has not tried any other substances. Steve has been employed as a line cook for the past two years and has never been fired from a job.

(Sample C – High-risk with tool present)

Steve is a 16-year-old male who has been charged with theft under \$5000 because he stole a pellet gun and night vision goggles valued at \$4,500 from an army surplus store. Police officers caught Steve with the stolen equipment shortly after the time of the offense. He has entered a guilty plea for the offense. The only offense on his record is one prior conviction for shoplifting (theft under \$5000) that involved stealing a hunting knife from a sporting goods store. Steve pleaded guilty for that offense and was sentenced to a 6-month probation term.

According to his probation officer, Steve was oppositional and verbally aggressive throughout his prior supervision period. He has a history of truancy at school and has been suspended three times for knocking out peers and threatening teachers. His most recent suspension occurred because he was selling pepper spray to younger kids at school. Steve was kicked out of his mother’s home two years ago and does not know his biological father. He currently lives with his older brother who has been incarcerated in the past for violent and drug-related offenses. Steve says he has “mad respect” for his brother because he “doesn’t take crap from anyone.” Regarding his current offense, Steve states that he has “no regrets” and is willing to do “whatever it takes” to get what he wants. Steve was intoxicated at the time of his arrest and reports that he has used alcohol daily for the past three years. He has also experimented with cocaine, ecstasy, and crystal meth. Further, Steve is suspected of being involved in a well-known gang that engages in drug trafficking.

A qualified clinician evaluated Steve’s risk for future violence and reoffending

with a widely-used risk assessment instrument that has accrued empirical support. The instrument consists of factors associated with recidivism, which the clinician uses to form an opinion about the offender's risk. The results of the risk assessment suggest that Steve is at a high risk for reoffending.

(Sample D – High-risk with tool absent)

Steve is a 16-year-old male who has been charged with theft under \$5000 because he stole a pellet gun and night vision goggles valued at \$4,500 from an army surplus store. Police officers caught Steve with the stolen equipment shortly after the time of the offense. He has entered a guilty plea for the offense. The only offense on his record is one prior conviction for shoplifting (theft under \$5000) that involved stealing a hunting knife from a sporting goods store. Steve pleaded guilty for that offense and was sentenced to a 6-month probation term.

According to his probation officer, Steve was oppositional and verbally aggressive throughout his prior supervision period. He has a history of truancy at school and has been suspended three times for knocking out peers and threatening teachers. His most recent suspension occurred because he was selling pepper spray to younger kids at school. Steve was kicked out of his mother's home two years ago and does not know his biological father. He currently lives with his older brother who has been incarcerated in the past for violent and drug-related offenses. Steve says he has "mad respect" for his brother because he "doesn't take crap from anyone." Regarding his current offense, Steve states that he has "no regrets" and is willing to do "whatever it takes" to get what he wants. Steve was intoxicated at the time of his arrest and reports that he has used alcohol daily for the past three years. He has also experimented with cocaine, ecstasy, and crystal meth. Further, Steve is suspected of being involved in a well-known gang that engages in drug trafficking.

Perceptions of the Offender

To what extent do you agree with the following statements about Steve?						
	Strongly Disagree	Somewhat Disagree	Slightly Disagree	Slightly Agree	Somewhat Agree	Strongly Agree
I would refer Steve to a treatment program.	0	1	2	3	4	5
Steve will probably be a lifelong criminal.	0	1	2	3	4	5
Steve stole the pellet gun purposefully, with the intention to harm someone.	0	1	2	3	4	5
Steve stole the pellet gun spontaneously, without intending to harm anyone.	0	1	2	3	4	5
Steve deserves a harsher penalty than other juveniles who committed similar crimes.	0	1	2	3	4	5
Steve has the potential to be rehabilitated.	0	1	2	3	4	5
Steve planned the current offense ahead of time.	0	1	2	3	4	5
Personality factors contributed to Steve's unlawful behaviour.	0	1	2	3	4	5
Situational factors contributed to Steve's unlawful behaviour.	0	1	2	3	4	5
With appropriate interventions, Steve's risk will likely be reduced.	0	1	2	3	4	5

Sentencing Propensities

How likely would you be to recommend each of the following for Steve?						
<i>Note: Please base your responses on what you think would be an appropriate placement or program referral for Steve, regardless of any local guidelines that may apply.</i>						
	Very Unlikely	Somewhat Unlikely	Slightly Unlikely	Slightly Likely	Somewhat Likely	Very Likely
Diversion	0	1	2	3	4	5
Probation	0	1	2	3	4	5
Nonsecure residential placement (e.g., group home)	0	1	2	3	4	5
Secure custodial placement (e.g., incarceration)	0	1	2	3	4	5
Transfer/waiver to adult court	0	1	2	3	4	5
No programs/services	0	1	2	3	4	5
Weekly program/service (e.g., counselling)	0	1	2	3	4	5
Intensive support and supervision/full-time attendance/day program	0	1	2	3	4	5
Residential treatment program	0	1	2	3	4	5

Which of the above placements would you be MOST likely to recommend for Steve?

Which of the above programs/services would you be MOST likely to recommend for Steve?

Demographics Questionnaire

What is your age?

What gender do you identify as?

Man	Woman	Transgender	Other
-----	-------	-------------	-------

What racial/ethnic group do you identify as (check all that apply)?

African-American/Black	Caucasian/White	Indigenous/Native American	Latino/Latina/Hispanic	Other
------------------------	-----------------	----------------------------	------------------------	-------

What type of court do you preside over?

Criminal	Juvenile	Family	Other
----------	----------	--------	-------

What age group do you have the most experience with?

Adult	Juvenile	Both
-------	----------	------

How many years have you practiced law?

Which province or state do you currently practice in?

Politically-speaking, people often describe themselves as 'left' or 'right'. Regarding social and legal matters, where would you place yourself on the following political spectrum?										
Extreme left ←————→ Extreme right										
○	○	○	○	○	○	○	○	○	○	○

Follow-up Questions

What crime did Steve commit?

Common Assault	Theft under \$5000	Drug offense	Unsure
----------------	--------------------	--------------	--------

What level of risk was Steve identified as in the vignette?

Low	Moderate	High	Not identified
-----	----------	------	----------------

What level of risk do YOU think Steve presented?

Low	Moderate	High	Unsure
-----	----------	------	--------

How often are risk assessment tools included in the pre-sentence reports you read?

Always (99-100% of the time)	Almost always (81-98% of the time)	Frequently (41-80% of the time)	Sometimes (11-40% of the time)	Rarely (1-10% of the time)	Never
---------------------------------	---------------------------------------	------------------------------------	-----------------------------------	-------------------------------	-------

How did you learn about this survey?

Email forwarded by a colleague	Email forwarded by a professional organization	Received by mail	Other
--------------------------------	--	------------------	-------

Do you have any feedback for the researchers?

Appendix B.

Pilot testing questions

How credible do you think these vignettes are?

Not credible	Somewhat credible	Very credible
--------------	-------------------	---------------

Does each vignette provide a realistic portrayal of an offender? If not, how might their credibility be improved?

Was any terminology unclear? Are there any terms that judges may find confusing?

The vignettes are designed to control for factors that may influence judges' perceptions of offenders and sentencing propensities *other than those related to risk (e.g., delinquent peers and antisocial attitudes)*. As such, the offender demographics, current offence, and offence history have been kept constant across vignettes. Did you notice any aspects of the vignettes *not related to risk* that may influence judges' responses?

Did the description of each offender match the risk category identified by the risk assessment tool (i.e., high- or low-risk)? If not, what level of risk do you think the offender(s) presented and why?

→ Are there any factors that could be added to (or removed from) the vignettes to make the offender appear higher or lower risk, respectively?

Is there any information missing from the vignette that should be included?

Are there any other questions or items that should be included in the survey?

Do you have any other suggestions for improvement?
