What are Judges' Views of Risk Assessments, and How Do Tools Affect Adolescent

Dispositions?

Melissa R. Jonnson and Jodi L. Viljoen

Simon Fraser University

Author Note

Melissa R. Jonnson, Department of Psychology, Simon Fraser University; Jodi L. Viljoen, Department of Psychology, Simon Fraser University

Correspondence concerning this article should be addressed to Melissa Jonnson, Department of Psychology, Simon Fraser University, 8888 University Drive, Burnaby BC, V5A 1S6. Email: mjonnson@sfu.ca

Abstract

Despite the increased presence of risk assessment tools in pre-sentence hearings, their role has been contentious, and the extent to which they influence dispositions is unclear. This study uses a self-report questionnaire to examine judges' opinions about risk assessment tools, as well as an experimental vignette design to evaluate whether judges' placement and program recommendations for a high- and low-risk adolescent are affected by the presence of an empirically supported tool and accompanying risk rating. It also assesses the influence of tool presence on judges' impressions of the adolescent's treatability and culpability. Participants included 170 judges from 34 American states and two Canadian provinces. The most common advantage of tools highlighted by judges was the additional information they provide, while the most common concern was their potential to obscure bias. Judges' placement recommendations and impressions of the adolescent were not significantly influenced by tool presence, but their program recommendations were more intensive and consistent for the high-risk adolescent when a tool was present than when it was absent.

Keywords: recidivism risk, judicial opinions, restrictive sanctions, evidence-based sentencing, mixed methods

What are Judges' Views of Risk Assessments, and How Do Tools Affect Adolescent

Dispositions?

Risk assessment tools are commonly used to guide judges' placement and program decisions (DeMatteo et al., 2016; Monahan & Skeem, 2016; Singh et al., 2014; Wachter, 2015). Indeed, recent estimates suggest that courts across the United States use tools to inform their dispositions "in some or all cases" (Starr, 2014, p. 809). Despite the increased presence of tools in pre-sentence hearings, their role has been contentious and research on judges' use of these tools has been scarce (Garrett & Monahan, 2019). Consequently, the present study examines (a) judges' opinions about the pros and cons of using tools to inform their dispositions, (b) the impact of tool-based risk ratings on judges' impressions of a hypothetical adolescent found guilty of a crime, and (c) the extent to which risk ratings influence the intensity and consistency of judges' placement and program recommendations for the adolescent. Below, we summarize prominent arguments for and against using tools in sentencing. We then review research on judges' opinions about tools and the impact of tools on judges' dispositions.

Arguments For Using Tools

Arguments in favor of pre-sentence risk assessment are typically premised on the idea that public protection is the main purpose of sentencing (Monahan & Skeem, 2016). In this context, risk assessment is necessary to ensure that high-risk adults and adolescents are identified and appropriately supervised or incapacitated. Since standardized tools outperform unstructured clinical judgement in predicting antisocial behaviour (Ægisdóttir et al., 2006; Hanson & Morton-Bourgon, 2009), advocates argue that tools are essential to evidence-based sentencing (Bonta, 2007; Elek et al., 2015; Thompson, 2017). In addition to their predictive superiority, tools could conceivably increase consistency, objectivity, and transparency in decision-making since they

require clinicians and legal decision-makers to consider the same inventory of risk-related factors for each defendant (Hoge, 2002; Skeem, 2013).

Tools may also assist with resource allocation by identifying the most appropriate type of supervision and programming for each person based on the risk-need-responsivity (RNR) model, a widely utilized framework for managing and reducing recidivism risk (Andrews & Dowden, 2006; Bonta & Andrews, 2017). The RNR model states that high-risk defendants are the ones most likely to benefit from intensive risk management strategies, whereas low-risk defendants should be diverted away from custodial channels into community alternatives (Andrews & Dowden, 2006; Nolan & Stewart, 2017). Adhering to the RNR model is especially beneficial for adolescents since exposing low-risk youth to higher-risk youth appears to have an iatrogenic effect (Andrews & Dowden, 2006). Given their potential to flag low-risk adolescents for diversion, tools have been heralded as a method to reduce overincarceration while protecting public safety (Laura & John Arnold Foundation, n.d.; Vincent et al., 2016).

Arguments Against Using Tools

Critics of pre-sentence risk assessment, on the other hand, tend to view sentencing as serving a retributive function (Monahan & Skeem, 2016). In this context, tools appear to violate the principle of proportionality, which states that punishments should be based on a defendant's culpability and the gravity of their offence, not their likelihood of committing future crimes (Cole, 2007; Hannah-Moffat, 2011). Some have argued that risk and culpability can become conflated with tools because courts might misinterpret a 'high-risk' label to mean that a person is inherently "bad" or untreatable (Cole, 2007; Hannah-Moffat, 2011; Hart & Cooke, 2013; Maurutto & Hannah-Moffat, 2007). Consequently, tools might elicit more punitive sentencing than is warranted (Evans & Salekin, 2014). Punitive dispositions are especially problematic

when applied to adolescents. Since incarceration has been shown to interfere with developmental

processes such as acquiring impulse control skills (Dmitrieva et al., 2012; Lambie & Randell,

2013), excessively harsh decisions may increase adolescents' risk for future offending.

Critics of pre-sentence risk assessment also worry that tools may exacerbate

demographic-based disparities in sentencing (Hannah-Moffat, 2011; Harcourt, 2015; Large et al.,

2014; Scurich & Monahan, 2016; Starr, 2014). Even if tools do not include demographic

variables such as race or socioeconomic status, critics highlight that other factors may act as

proxies. Specifically, disadvantaged groups may score higher on tools due to risk factors that

tend to co-occur with social disadvantage, such as poor educational attainment, employment

instability, and history of trauma (Cole, 2007; Maurutto & Hannah-Moffat, 2007; Shepherd &

Anthony, 2018). To the extent that risk assessment reports are influential in the courtroom, these

inflated scores may contribute to systemic biases such as the overrepresentation of minority

groups in prison (Starr, 2014).

Judges' Opinions About Using Tools in Sentencing

Given judges' prominent role in determining supervision levels and other risk

management strategies for those found guilty of an offence, understanding how tools affect their

decisions is pertinent. However, research on judges' opinions about tools has been limited, and

results have been mixed. For instance, one study asked Canadian judges to rate the importance of

34 factors commonly included in pre-sentence reports; risk for violence and risk for general

recidivism were rated as the 10th and 16th most important factors, respectively (Bonta et al.,

2005). However, only 21% of judges indicated that they would like pre-sentence reports to

include results from a standardized risk tool. In a study of American judges, about half (48%)

reported that they routinely used tools to reach a disposition (Shook & Sarri, 2007). These judges

rated the usefulness of tools as 3.2 out of 5 for making pretrial detention decisions and 3.7 out of 5 for making post-adjudication placement decisions.

These studies provide a reasonable starting point for understanding judges' general perceptions of tools. However, quantitative survey items are limited because they do not allow respondents to raise additional points or explain their reasoning. In contrast, qualitative techniques (e.g., open-ended questions) can provide richer data and facilitate strategies to improve the usefulness of tools. To our knowledge, only three studies have used a qualitative or mixed methods approach (Chanenson & Hyatt, 2016; Metz et al., 2019; Monahan et al., 2018). Judges in those studies noted that tools increase objectivity in sentencing, provide more data than would otherwise be available, and help reduce prison populations. Some judges also raised concerns that tools fail to recognize individual circumstances, may lack empirical support, and may be misused. Although they included qualitative data, only one of the studies used a systematic method for analyzing the data (Metz et al., 2019), and two focused on a specific tool used in Virginia (Metz et al., 2019; Monahan et al., 2018) rather than opinions about tools more generally. In addition, studies have not yet tested if judges' view adolescent risk assessments more or less positively than adult risk assessments or explored judicial views about how tools might contribute to racial and ethnic biases, despite these being key areas of controversy.

Influence of Tools on Judicial Decisions

In addition to considering judges' opinions of evidence-based sentencing, it is important to explore whether tools have a tangible impact on their decisions. A recent systematic review found that, overall, pretrial detention rates tended to decrease slightly after tools were implemented, particularly for low-risk defendants (Viljoen et al., 2019). Relatedly, some studies show that when judges are provided with risk assessment results, their decisions are consistent

with the RNR model such that high-risk defendants receive more program referrals and restrictive placements than low-risk defendants (e.g., Jung et al., 2015; Vincent et al., 2016). However, these studies have some important gaps.

First, rates of heterogeneity were high in the systematic review (Viljoen et al., 2019), indicating substantial variability in findings. Second, many studies failed to control for confounds. For instance, studies that showed large-scale decreases in detention rates following tool implementation did not examine the relative impact of other initiatives that were implemented simultaneously (e.g., alternatives to detention programs). Likewise, studies that found strong adherence to the RNR model when a tool was used (e.g., Jung et al., 2015; Vincent et al., 2016) did not include a comparison group where a tool was not used. Thus, it is impossible to tease apart whether results are attributable to the presence of a tool or to other factors.

Of the few studies that have directly compared detention decisions when tools are and are not used, all have relied on naturalistic field designs (e.g., Hilton & Simmons, 2001) and only one has included judges (van Wingerden et al., 2014). Although naturalistic studies have good ecological validity, they often have difficulty controlling for confounds. For example, although Hilton and Simmons (2001) found a small improvement in the effect of risk scores on tribunal decisions when decision-makers had access to an actuarial tool report, longer stay patients were more likely to have a report on file, so report presence was conflated with length of stay.

In addition to the lack of tightly controlled studies with judges, insufficient attention has been devoted to the mechanisms by which tools might influence decisions. As noted above, there are concerns that tools lead to harsher dispositions by exacerbating stigma towards high-risk defendants. van Wingerden et al. (2014) proposed that risk levels stated in pre-sentence reports might create a framing effect (Isaacs, 2011) causing judges to overweight information consistent

with the risk level. However, contrary to their expectations, both low- and high-risk defendants

with a risk assessment report received less restrictive sentences than those without a report.

Researchers have not tested whether tools exacerbate stigma towards high-risk adolescents.

A final gap in the literature is the absence of studies looking at how tools effect the

consistency and reliability of judicial decisions (Hoge, 2002; Skeem, 2013). This question is

important because consistent sentencing (i.e., the extent to which similar cases are treated

similarly) is a fundamental principle of justice across several countries (e.g., *Criminal Code of*

Canada, 1985; National Center for State Courts et al., 2009; Pina-Sánchez, 2015; Tonry, 2018),

and it can increase public confidence that the justice system is operating fairly and predictably

(Pina-Sánchez, 2015). If judges are following this principle, one would expect different judges to

recommend a comparable disposition for the same defendant. Since risk assessment tools

provide standardization and structure to the information provided to judges, they could

conceivably decrease variability in dispositions.

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 the direct impact of tools on placement and

program decisions, the mechanisms by which tools influence decisions, the effect of tools on

decision consistency, and judges' own opinions about the benefits and drawbacks of tools. The

present mixed methods study was designed to address such gaps. Specifically, we used a detailed

self-report questionnaire to examine judges' opinions about tools, as well as an experimental

vignette design to test the influence of tool presence on the (a) perceived treatability and

culpability of an adolescent defendant, (b) intensity of placement and program recommendations

for the adolescent, and (c) consistency of recommendations.

For the experimental component, we manipulated the presence of a tool (i.e., statement that a tool was used and an explicit risk rating versus no mention of a tool or risk rating) and the number of risk factors described in the vignette (few risk factors versus several). Both manipulations were included because we anticipated that tools may have a different impact on high- and low-risk defendants. In addition, even when risk is not formally rated with a tool, many experts describe risk factors in their reports (Storey et al., 2015; Wilson et al., 2015). As such, we wanted to differentiate the effect of the tool-based risk rating from the description of risk factors, which is consistent with studies examining the influence of psychopathy labels on legal decisions (e.g., Murrie et al., 2007).

We predicted that, overall, judges would express moderate support for using tools to inform placement and program recommendations, although they would agree that more caution is warranted when using tools with adolescents. We also predicted that tool presence would exacerbate pessimistic views about the high-risk adolescent (e.g., that he is difficult to treat and prone to lifetime criminality). Furthermore, we anticipated that tool presence would increase the intensity of their placement and program recommendations for the high-risk adolescent, whereas it would decrease the intensity of these recommendations for the low-risk adolescent. Finally, we expected that tool presence would increase the consistency of judges' recommendations.

Methods

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. Our actual response rate was 9.2%, which is slightly lower than rates reported by other studies using

online recruitment (e.g., 14.3% in Evans & Salekin, 2016; 8 to 18.8% in Wise & Safer, 2004) and much lower than rates obtained by mail recruitment (e.g., 25% in Murrie et al., 2007; 52.8% in Monahan et al., 2018) . As is common in survey research (Dillman et al., 2014), some respondents ($n = 24$) answered only the first few questions and an additional 32 provided no responses. After removing these incomplete cases, 170 judges (39.0% women, 59.7% men, and 1.3% another gender) were included, which resulted in an adjusted response rate of 6.9%.

Most judges were Caucasian (81.8%), 6.3% were African American, 5.0% were Latino/Latina or Hispanic, 1.9% were Indigenous or Native American, and 5.0% identified as 'Other'. Judges practiced in 34 states, such as Michigan ($n = 64$), Ohio ($n = 10$), and Virginia ($n = 8$), and three practiced in Canada. Most were between the ages of 40 to 59 years old (51.3%) or 60 to 79 years old (43.7%), while the rest were below 40 years old (5.1%). Judges had been practicing for an average of 28.3 years ($SD = 9.76$; range = 0 to 46). Since demographic information was collected at the end of the survey, we were unable to test whether the 56 excluded judges differed from the 170 retained judges.

Procedure

After obtaining institutional review board approval, judges were invited to participate in the online study by email. In accordance with Tailored-Design Methodology (Dillman et al., 2014), an evidence-based survey implementation strategy, we emphasized how the study results would be useful, highlighted that judges are in a unique position to provide the requested advice, and situated more interesting questions at the beginning of the survey. We also sent a follow-up email to judges approximately four weeks after the first contact email. The study took 15-20 minutes to complete, and judges did not receive external incentives.

Given the complementary strengths of qualitative and quantitative approaches (Creswell,

2013; Johnson et al., 2007), we combined a self-report survey with a 2 x 2 experimental vignette design in which risk level (high or low) and tool presence (present or absent) were varied systematically. To achieve a balance of breadth and depth, we included a range of categorical questions about tools in our survey in addition to an open-ended question. Judges' responses to the open-ended question helped us contextualize their responses to the categorical items and informed our interpretation of the experimental findings. Judges were randomly assigned to one of the four vignette conditions. We included two manipulation checks to ensure judges attended appropriately to the vignette. After removing those who failed at least one manipulation check (*n* = 31), 139 judges were included in the experimental portion of the study. Judges were reasonably dispersed across groups, with 34 in the low-risk, tool-present group, 32 in the low-risk, tool-absent group, 28 in the high-risk, tool-present group, and 45 in the high-risk, tool-absent group.

Materials

Opinions About Risk Assessment Tools

A 26-item questionnaire was developed to assess judges' opinions about pre-sentence risk assessment. Consistent with the stepwise procedure for measure development (Holmbeck & Devine, 2009), content was generated from relevant theories, extant literature (e.g., Monahan & Skeem, 2014), and consultation with four experts (three forensic psychologists and one lawyer). Judges indicated, on a six-point Likert-type scale (0 = *strongly disagree*; 5 = *strongly agree*), the extent to which they agreed with 22 statements about tools (e.g., "Tools are useful for making decisions about where to place an offender"; "History of trauma should be included in risk assessment tools"). 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 at. Finally,

judges were asked one open-ended question about the pros and cons of using tools to help make decisions about supervision, placements, or services.

Vignette

The vignette provided a brief account of an adolescent who had pled guilty to a 'Theft Under $5000' charge. This offence was chosen because of its moderate severity and because theft is one of the most common offences committed by juveniles in the United States (Sickmund & Puzzanchera, 2014) and Canada (National Crime Prevention Centre, 2012). An adolescent was chosen because of the heightened concerns about labelling juveniles as high-risk (Cole, 2007; Maurutto & Hannah-Moffat, 2007). The adolescent's current charge and offence history were consistent across conditions.

The 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 (see Supplemental Materials). The high-risk conditions described the adolescent as having several risk factors (i.e., history of violence, school truancy and suspensions, caregiver criminality, precarious housing, substance use, criminogenic attitudes, suspected gang affiliation), whereas the low-risk ones described him as having few risk factors and several protective factors (i.e., high commitment to school, prosocial attitudes, strong attachment to caregivers, steady employment, remorse for current offence). The tool present conditions stated that an "empirically supported risk assessment tool was used" and provided an accompanying risk rating (i.e., high or low), while the tool absent conditions did not mention a tool or the adolescent's risk level. The tool in the vignette was unspecified to increase generalizability since most research on judges' use of tools has focused on a specific tool (e.g., Metz et al., 2019; Monahan et al., 2018). Using an unnamed tool also helped account for the fact that some judges might be more familiar with a specific tool

(e.g., one that is used in their district) compared to other judges.

Impressions of the Adolescent

Using the stepwise procedure noted above, a 10-item questionnaire was developed to

assess judges' impressions of the hypothetical adolescent. Items involved treatment amenability

(e.g., "Steve has the potential to be rehabilitated), long-term criminality (e.g., "Steve will

probably be a lifelong criminal), criminal intent (e.g., "Steve planned the current offense ahead

of time") and criminal responsibility (e.g., "Steve deserves a harsher penalty than other juveniles

who committed similar crimes"). Judges rated items on a 6-point Likert-type scale ranging from

0 (*strongly disagree*) to 5 (*strongly agree*).

Sentencing Propensities

Judges were asked to rate their likelihood of recommending the following placements:

diversion, probation, non-secure residential placement (e.g., group home), and secure custodial

placement (e.g., incarceration). They were also asked to rate their likelihood of recommending

the following program options: none, weekly, intensive (e.g., intensive support and supervision

program), or residential (e.g., full-time treatment program). Items were selected based on

previous studies of judicial decision-making (e.g., Murrie et al., 2007; Penner, 2009) and a

consideration of relevant placement and program options. Once again, response options ranged

from 0 (*highly unlikely*) to 5 (*highly likely*).

Analytic Plan

Consistent with recommendations (Creswell, 2013), we used intramethod triangulation to

assess judges' opinions about tools. Specifically, we examined their mean responses to the

Likert-type items, then analyzed their qualitative responses using conventional content analysis

(CCA)**.** CCA is an inductive approach that allows themes to emerge spontaneously rather than

using a preconceived categorical structure (Hsieh & Shannon, 2005). To help ensure the coding

process was reflexive and rigorous, a second trained coder blindly reviewed judges' comments

and categorized them using the coding frame identified by the first coder, as suggested by

O'Connor and Joffe (2020). Intercoder reliability was assessed using Cohen's kappa (κ).

To test whether tool presence moderated the association between adolescent risk level

and judicial impressions of the adolescent, we conducted 10 ordinal logistic regressions, one for

each item in the impressions questionnaire. We applied a Bonferroni correction (.05/10), such

that an alpha level of $p < .005$ was considered the threshold for significance.

To assess whether tool presence influenced placement or program recommendations, we

used chi-square and loglinear analyses and compared standardized residuals and odds ratios

(ORs) across cells. Since ORs are most interpretable in 2×2 contingency tables (Field et al.,

2012), we divided placement recommendations into restrictive (nonsecure residential and secure

custodial) and nonrestrictive (diversion and probation) categories and program recommendations

into intensive (fulltime and residential) and nonintensive (none or weekly) categories.

Finally, to examine whether tool presence increased the consistency of judges'

recommendations, we used Krishnamoorthy and Lee's (2014) modified signed-likelihood ratio

test (MSLRT). If the omnibus MSLRT was significant, we conducted follow-up pairwise

comparisons to assess which standard deviations significantly differed. We used the R package

'cvequality' to conduct these tests (Version 0.1.3; Marwick & Krishnamoorthy, 2018).

Results

Judges' Opinions About Risk Assessment Tools

For the purpose of making placement decisions, most judges (55.6%) said tool results

should be given moderate emphasis, 26.6% said minimal emphasis, 12.4% said substantial

emphasis, and 5.3% said no emphasis. Compared to clinical opinion, most judges said tool results should be given less emphasis (50.6%), 41.0% said equal emphasis, and 8.4% said more emphasis. Almost three-quarters of judges (71.8%) endorsed using tools for pre-trial detention, 29.4% for adjudication, 81.2% for dispositions/sentencing, and 65.3% for parole/release. The statements judges agreed most strongly with were that trauma history should be included as a risk factor in tools, tools are useful for making decisions about programs/services, and it is important to consider future risk when making sentencing decisions (Table 1). The statements they disagreed most strongly with were that race should be included as a risk factor in tools, tools lead to more restrictive sentences, and tools decrease fairness and consistency. As a group, judges agreed slightly that we should be more cautious when using tools with juveniles.

Although opinions were generally favourable to tools at the aggregate level, judges' responses to the open-ended question ranged from definitive support to strong skepticism. Ninety-four judges (55.3%) provided qualitative responses to the open-ended question. Judges who responded to this item were similar in gender and ethnicity to those who did not (ps = ns). However, they were more likely to be 60-79 years old than from the younger age groups, χ^2 (1) = 10.74, p = .001, and they reported more years of experience (M = 31.03, SD = 9.17 vs. M = 25.00, SD = 9.48), t (135) = -3.77, p < .001. Using CCA, we identified 35 codes and organized them into 12 subcategories (Figure 1). The subcategories fit best into three broad categories: benefits of using tools (n = 57 comments made by 51 judges), concerns (n = 98 comments made by 68 judges), and circumstances that influence tool usefulness (n = 33 comments made by 29 judges). Intercoder reliability for sorting all comments into the 12 subcategories fell in the "nearly perfect" range (κ = .91; Landis & Koch, 1977; O'Connor & Joffe, 2020).

Benefits of Using Tools

The most common reason judges articulated for using tools (*n* = 33) was that *they are helpful for making decisions* simply because they provide more data than would otherwise be available. 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." Several judges (*n* = 17) also emphasised that *tools can help increase objectivity and consistency*, as illustrated by the following comment:

> Absent an evidence-based risk assessment tool, 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 recognize and set aside implicit biases. An objective risk assessment tool can balance those factors.

Three smaller subcategories captured judges' views that tools can *help divert low-risk adolescents* from programs that are geared towards medium- and high-risk adolescents (*n* = 3), *help reduce overincarceration* (n = 2), and *facilitate transparency* (n = 2), such as by helping judges "frame the conversation with [an] offender at the sentence hearing."

Concerns with Using Tools

The most common criticism of tools involved *concerns about validity* (*n* = 31). One worried that judges "must blindly rely on the scientific accuracy of the assessor without understanding the basis of the assessment tool." Several judges also raised *concerns about bias* (*n* = 27), such as biases that can arise from relying on defendants' self-report, implicit and explicit biases of the assessor, and systemic biases that can "adversely impact people of color." Another prevalent concern was that *tools use a 'one-size-fits-all' approach* (*n* = 20). One judge, for example, highlighted that "juveniles are individual persons, not shoes. A one size fits all approach to juvenile justice is not workable." Twelve judges expressed concern that *tools limit*

judges' discretion. For instance, one judge stated, "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." Finally, eight judges indicated that *tools are unnecessary, too costly, or too time-consuming*.

Circumstances that Influence the Usefulness of Tools

Although some judges were clearly for or against using tools to make judicial decisions, many highlighted 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." Furthermore, several responses suggested that tools *should be treated as only one piece of evidence among many* ($n = 19$) and should be used as "an aide to judgement not a substitute for it."

Influence of Tools on Judges' Impressions of the Adolescent

Next, we examined results from the experimental manipulation. Regarding judges' impressions of the adolescent, the main effect of tool presence was nonsignificant across analyses, whereas the main effect of risk level was significant in six of the 10 regressions (Table 2). When the adolescent was presented as high risk, judges believed more strongly that he would be a lifelong criminal ($M = 2.47$ vs. $M = 0.53$), that he deserved a harsher penalty than his peers ($M = 2.00$ vs. $M = 0.86$), and that personality ($M = 3.59$ vs. $M = 2.19$) and situational ($M = 4.10$ vs. $M = 2.79$) factors contributed to his behaviour. They believed less strongly that he could be rehabilitated ($M = 3.82$ vs. $M = 4.68$) but more strongly that they would refer him to a treatment program ($M = 3.86$ vs. $M = 2.21$). Although the high-risk adolescent was viewed more negatively than the low-risk adolescent, the interaction between risk level and tool presence was not

significant in any of the models indicating that tool presence did not exacerbate such pessimism.

Influence of Tools on Judges' Placement and Program Recommendations

Judges were unlikely to recommend a restrictive placement for the low-risk adolescent or a lenient placement for the high-risk adolescent. Although this pattern makes intuitive sense, it violated the cell count assumption. As such, rather than using loglinear analyses to test the three-way interactions, we split the dataset by tool presence and conducted parallel chi-square analyses (Field et al., 2012). We used the same procedure for program recommendations, but since the cell count violation was minor (i.e., 25% of cells violated the assumption), we ran the full model using loglinear analysis as a final step.

Although there were significant associations between risk level and placement recommendation for both the tool-absent, $\chi^2(1) = 19.60$, $p < .000$, and tool-present, $\chi^2(1) = 24.36$, $p < .000$, groups, the differences across groups were negligible. For the high-risk adolescent, 56.0% of judges recommended a restrictive placement when a tool was present, compared to 48.8% when a tool was absent, and ORs were comparable across tool-present (1.27) and tool-absent (0.95) conditions. For the low-risk adolescent, 0% of judges recommended a restrictive placement when a tool was present, compared to 2.9% when a tool was absent. An OR could not be defined due to the zero count in one cell, but the standardized residuals were similar in the tool-present and tool-absent conditions (Figure 2).

There were also significant associations between risk level and program recommendation for the tool-absent, $\chi^2(1) = 27.86$, $p < .000$, and tool-present, $\chi^2(1) = 40.08$, $p < .000$, groups, and these differences were more substantial than those for placement recommendations. For the high-risk adolescent, 96.0% of judges recommended intensive programming when a tool was present, compared to 78.6% when a tool was absent. ORs indicated that judges were 24.0 times more

likely to recommend intensive programming when a tool was present, but only 6.6 times more

likely when a tool was absent. For the low-risk adolescent, 12.1% of judges recommended

intensive programming when a tool was present, compared to 16.1% when a tool was absent.

Although the standardized residual was further from zero when a tool was present (Figure 2),

ORs were comparable across tool-present (0.14) and tool-absent (0.19) conditions. Results from

the three-way loglinear analysis were consistent with the chi-square analyses. Specifically, the

highest order interaction (i.e., Tool Presence x Risk Level x Program Intensity) verged on

significance, X^2 (1) = 3.50, p = 0.062.

Influence of Tools on Consistency

The standard deviations of judges' placement recommendations were similar in tool-

present and tool-absent conditions, MSLRT = 0.56, p = 0.906 (Table 3). However, standard

deviations of program recommendations differed, MSLRT = 20.48, $p < .000$ (Table 3). Follow-

up pairwise comparisons revealed that the presence of a tool increased consistency in judges'

program recommendations for the high-risk, but not the low-risk, adolescent. Specifically,

variability was lower in the high-risk, tool-present condition ($SD = 0.50$) than in its tool-absent

counterpart ($SD = 0.76$), MSLRT = 6.86, p = .009.

<div align="center">

Discussion

</div>

Consistent with extant research (Shook & Sarri, 2007; Monahan et al., 2018), most

judges in the current study agreed that tools are useful for making placement and program

decisions. In fact, more judges indicated that tools should be used at the disposition stage than at

any other point during the sanctioning process. However, they also agreed slightly that more

caution is warranted when using tools to inform adolescent dispositions. In addition, most judges

reported that tool results should be given less weight than clinical opinion. Since 'clinical

opinion' was not defined, judges may have thought it included expert opinions, which are often

guided by structured measures. Nevertheless, prior research has also found that judges prefer

subjective clinical judgement over tools (Chanenson & Hyatt, 2016), and responses to our open-

ended questions revealed that many judges were skeptical of tool validity and objectivity.

Relatedly, many found it important to clarify that tools should not replace their judgement, which

may reflect the concern identified in other research (Dunnavant & Levitt, 2015; Monahan et al.,

2018) that limiting judicial discretion reduces judges' ability to apply wisdom and flexibility to

their decisions. Despite these concerns, many judges highlighted that tools provide useful

information and lead to more objective and consistent decisions.

Regarding the impact of tools on stigma toward high-risk defendants, the presence of a

tool and risk rating did not exacerbate pessimistic views of the high-risk adolescent in this study.

Although inconsistent with our hypothesis, this finding is consistent with Murrie et al. (2007),

who reported that descriptions of adolescents may be more influential than diagnostic labels such

as 'psychopathy'. Given that this is the first study to assess whether tools influence judges'

impressions of defendants, 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 tools exacerbate prejudice toward high-risk defendants.

As noted previously, the primary question in the literature has been about the effect of

tools on dispositions. This study found that tool presence influenced program, but not placement,

recommendations for the high-risk adolescent. Specifically, judges recommended more intensive

and consistent programming for the high-risk adolescent when a tool was present than when it

was absent. However, tool presence did not affect placement or program recommendations for

the low-risk adolescent, which suggests mixed support for the notions that tools increase

adherence to the RNR model or reduce over-incarceration (Bonta, 2007; Laura & John Arnold Foundation, n.d.). Although tool presence had a lesser impact than expected, it is not surprising in retrospect given judges' comments that tools are just one piece of evidence that they consider when making decisions. It seems likely that high-risk cases are more complex and challenging than low-risk cases and that tools may be more helpful in such cases when appropriate levels of programming are less apparent. Moreover, since high-risk adolescents 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).

That said, the current findings conflict somewhat with van Wingerden et al. (2014), who found that tools indiscriminately led to less restrictive judicial sentencing decisions after controlling for a wide range of confounds. In that study, the authors proposed that tools may produce an 'information effect', whereby the information they provide allows decision-makers to consider a defendant'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 a person's risk to reoffend. Notably, the risk ratings provided in the current study were consistent with the descriptions of the adolescent in the vignette (i.e., the adolescent identified as "high-risk" was described as having several risk factors, while the adolescent identified as "low-risk" was described as having few risk factors). Judges may, therefore, have been able to ascertain the adolescent's risk level based on the description alone and consequently produced similar placement recommendations regardless of tool presence.

Limitations

Although this study includes the largest sample of judges that exists in the mixed-methods literature on pre-sentence risk assessment, our response rate was only 9.2%, which significantly limits the generalizability of our results. Future studies should likely recruit judges by mail or a combination of mail and email as these methods appear to achieve higher response rates (e.g., Monahan et al., 2018). In addition, like other studies on judges' use of risk assessment (e.g., Chanenson & Hyatt, 2016), the current sample consisted predominantly of American judges, so future research should sample judges from other countries.

Based on our a priori power analyses, the current sample was large enough to detect moderate-to-large effects. However, it was likely underpowered to detect small effects which may still 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 some design issues that may have reduced power. For example, the low-risk adolescent 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 compare placement recommendations between those groups. In addition, 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 tools. As such, the effect of the tool may have been suppressed. However, assessors who use unstructured judgement to assess recidivism risk generally do not state this approach in their reports, so we kept the absence of a tool vague to be more realistic.

Despite its limitations, this study is the first to our knowledge to experimentally manipulate the presence of a risk assessment tool using a sample of judges and, hence, it increases our ability to draw causal conclusions about the influence of tools on judges'

placement and program recommendations. It is also the first study to examine judges' opinions

about pre-sentence risk assessment using a mixed methods approach. Finally, it is novel in its

investigation of whether tools exacerbate stigma towards high-risk adolescents (e.g., perceived

treatability) or increase the consistency of judges' recommendations.

Future Directions

This study found that the presence of a tool and accompanying risk rating may help

judges determine appropriate levels of programming for adolescents who are deemed high risk.

Since judges often set conditions that require youth to attend specific programs rather than

deciding solely on the intensity of such programming, future research could explore whether

tools increase consistency when judges are asked to match programs to specific criminogenic

needs (e.g., substance use). In addition, future research could explore whether judges' opinions

or recommendations are moderated by type of tool (e.g., actuarial versus structured professional

judgement) since tools vary widely, and judges likely feel differently about adhering to evidence-

based sentencing in general versus using a specific tool in practice.

Although tool presence did not influence placement recommendations in this study, it

will be important for future research to test whether such associations are moderated by a

defendant's race or ethnicity. Although some people hypothesize that tools help minimize the

influence of implicit biases in sentencing decisions by providing a standard and consistent

structure (see Eaglin & Solomon, 2015, p. 28), others worry that tools may exacerbate

demographic-based discrimination (e.g., Starr, 2014). Indeed, one of the biggest concerns

expressed in judges' open-ended responses was that they may involve biases, such as those that

contribute to sentencing disparities for people of color. However, this concern has not yet been

tested through carefully controlled experimental designs. Future research could also use

experimental designs to explore whether risk assessments and dispositions are influenced by other types of biases, such as anchoring and confirmation biases (Neal & Grisso, 2014).

Finally, future research should continue examining the mechanisms by which tools influence judicial decisions. In this study, we manipulated tool presence by explicitly stating that the adolescent had been assessed with a tool and was determined to be either low- or high-risk for reoffending. We selected this approach because experts often refer to risk factors in court-ordered reports regardless of whether a tool is completed (Wilson et al., 2015). However, it is possible that the primary benefit of tools lies in their ability to provide more accurate risk-related background information (e.g., history of aggression, school attendance) rather than their ability to clearly link that information to risk predictions. Therefore, more research is needed to understand whether tools influence judicial practices by influencing the nature of the reports themselves or simply by providing a clear risk prediction.

Implications for Policy and Practice

Findings suggest several implications for policy and practice. Given judges' concerns about tool validity and objectivity, as well as the wide variation in the quality of tools available, it seems prudent to ensure that tools are supported by empirical research as well as court actors before being implemented. Judges might be more likely to buy-in to using tools if they are involved in tool development and implementation (Metz et al., 2020; Vincent et al., 2016). Judges may also have a greater confidence in tools if they are offered training on the RNR model and other research supporting tool use (Metz et al., 2019; Vincent et al., 2016). In addition to providing education on general research support for tools, jurisdictions should collect their own data on dispositional outcomes before and after tools are implemented to help evaluate the effects of the tool and ensure desired changes are achieved (e.g., reductions in racial disparities

and overincarceration). Finally, given judges' concerns about the inflexible, one-size-fits-all

approach inherent in some tools, tool developers should create instruments that allow judges to

consider additional factors not captured by the tool. Indeed, several risk assessment tools already

include sections for case-specific factors to be added (e.g., Borum et al., 2006).

Conclusion

Considerable discourse has arisen from the proliferation of risk assessment tools in the

courtroom. 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

dispositional decisions. Interestingly, judges' opinions about tools echoed those expressed in the

literature. They recognized that tools may provide more information about defendants and may

increase objectivity in decision-making. However, they also expressed concerns that tools may

simply shift bias from the judges to the professionals who develop and administer tools, thereby

making the bias more insidious.

These concerns are perhaps tempered by the fact that tools did not significantly influence

judges' impressions of the adolescent or the restrictiveness of their placement decisions. Instead,

the only apparent effects of the tool were that it increased the intensity and consistency of

program recommendations for the high-risk adolescent. As such, the benefits and risks of using

tools to guide placement decisions may be overestimated. 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., adolescents with multiple risk

factors) than with increasing or decreasing the restrictiveness of their sanctions. Therefore, to

maximize the potential utility of tools, researchers and tool developers should increase their

communication about tools with judges and rigorously evaluate the extent to which tools are

achieving their intended aims.

References

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. https://doi.org/10.1177/0306624X05282556

Bonta, J. (2007). Offender risk assessment and sentencing. *Canadian Journal of Criminology*

    and Criminal Justice, *49*, 519–529. https://doi.org/10.3138/cjccj.49.4.519

Bonta, J., & Andrews, D. A. (2017). *The psychology of criminal conduct* (6th ed.). Routledge.

Bonta, J., Bourgon, G., Jesseman, R., & Yessine, A. K. (2005). *Presentence reports in Canada.*

    (User Report 2005-03). Public Safety Canada.

Borum, R., Bartel, P., & Forth, A. (2006). *Manual for the structured assessment for violence risk*

    in youth (SAVRY). Odessa, FL: Psychological Assessment Resources.

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).

    https://digitalcommons.law.villanova.edu/cgi/viewcontent.cgi?article=1201&context=wp

    s

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.

    https://doi.org/10.3138/cjccj.49.4.493

Creswell, J.W. (2013). *Qualitative inquiry and research design: Choosing among five*

    approaches. SAGE Publications.

Criminal Code of Canada, R.S.C. 1985, c. C-46, s.718.2(a)(i).

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). American Psychological Association.

Dillman, D. A., Smyth, J. D., & Christian, L. M. (2014). *Internet, mail, and mixed-mode surveys: The tailored design method., 4th ed*. 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. https://doi.org/10.1017/S0954579412000545

Dunnavant, B., & Levitt, H. (2015). The development of wisdom in judicial decision-making. *The Humanistic Psychologist, 43*, 1–23. https://doi.org/10.1080/08873267.2014.993068

Eaglin, J., & Solomon, D. (2015). *Reducing racial and ethnic disparities in jails: Recommendations for local practice*. Brennan Center for Justice at New York University School of Law.

Ægisdóttir, S., White, M. J., Spengler, P. M., Maugherman, A. S., Anderson, L. A., Cook, R. S., Nichols, C. N., Lampropoulos, G. K., Walker, B. S., Cohen, G., & Rush, J. D. (2006). The Meta-Analysis of Clinical Judgment Project: Fifty-Six Years of Accumulated Research on Clinical Versus Statistical Prediction. *The Counseling Psychologist*, *34*, 341–382. https://doi.org/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.* National Center for State Courts.

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. https://doi.org/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. https://doi.org/10.1037/tam0000062

Field, A., Miles, J., & Field, Zoë. (2012). *Discovering statistics using R.* SAGE Publications.

Garrett, B., & Monahan, J. (2019). Judging risk: The use of risk assessment in sentencing.
Judicature, *103*(2), 42–48. https://heinonline-
org.proxy.lib.sfu.ca/HOL/Page?collection=journals&handle=hein.journals/judica103&id
=128&men_tab=srchresults

Hannah-Moffat, K. (2011). *Actuarial Sentencing: An "unsettled" proposition.* Unpublished
manuscript. http://www.albany.edu/scj/documents/Hannah-
Moffatt_RiskAssessment_000.pdf

Hanson, R. K., & Morton-Bourgon, K. E. (2009). The accuracy of recidivism risk assessments
for sexual offenders: A meta-analysis of 118 prediction studies. *Psychological
Assessment, 21*, 1–21. https://doi.org/10.1037/a0014421

Harcourt, B. E. (2015). Risk as a Proxy for Race: The Dangers of Risk Assessment. *Federal
Sentencing Reporter, 27*(4), 237–243. https://doi.org/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. https://doi.org/10.1002/bsl.2049

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.
https://doi.org/10.1023/A:1010607719239

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

https://doi.org/10.1177/0093854802029004003

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/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*, 691–696.

https://doi.org/10.1093/jpepsy/jsp046

Hsieh, H., & Shannon, S. (2005). Three Approaches to Qualitative Content Analysis. *Qualitative

Health Research, 15*, 1277–1288. https://doi.org/10.1177/1049732305276687

Isaacs, D. (2011). Baseline Framing in Sentencing. *The Yale Law Journal, 121*, 426–458.

http://www.jstor.org.proxy.lib.sfu.ca/stable/23079336

Johnson, R., Onwuegbuzie, A., & Turner, L. (2007). Toward a Definition of Mixed Methods

Research. *Journal of Mixed Methods Research, 1*, 112–133.

https://doi.org/10.1177/1558689806298224

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.

Krishnamoorthy, K., & Lee, M. (2014). Improved tests for the equality of normal coefficients of

variation. *Computational Statistics*, 29, 215–232. https://doi.org/10.1007/s00180-013-

0445-2

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. https://doi.org/10.1080/14999013.2006.10471242

Lambie, I., & Randell, I. (2013). The impact of incarceration on juvenile offenders. *Clinical Psychology Review, 33*, 448–459. https://doi.org/10.1016/j.cpr.2013.01.007

Large, M., Ryan, C., Callaghan, S., Paton, M., & Singh, S. (2014). Can violence risk assessment really assist in clinical decision-making? *Australian & New Zealand Journal of Psychiatry, 48*, 286–288. https://doi.org/10.1177/0004867413498275

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. https://doi.org/10.3138/cjccj.49.4.465

Marwick, B., & Krishnamoorthy, K. (2018). cvequality: tests for the equality of coefficients of variation from multiple groups [R software package version 0.1.3]. https://github.com/benmarwick/cvequality

Metz, A., Monahan, J., Garrett, B., & Siebert, L. (2019). Risk and resources: A qualitative perspective on low-level sentencing in Virginia. *Journal of Community Psychology, 47*, 1476–1492. https://doi.org/10.1002/jcop.22199

Metz, A., Monahan, J., Siebert, L., & Garrett, B. (2020). Valid or voodoo? A qualitative study of attorney perceptions of risk assessment in sentencing and plea bargaining. *Journal of Community Psychology, 48*, 2053–2068. https://doi.org/10.1002/jcop.22404

Monahan, J., Metz, A. L., & Garrett, B. L. (2018). Judicial appraisals of risk assessment in sentencing. *Behavioral Sciences & the Law, 36*, 565–575. https://doi.org/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.

https://doi.org/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. https://doi.org/10.1146/annurev-clinpsy-021815-092945

Murrie, D. C., Boccaccini, M. T., Guarnera, L. A., & Rufino, K. A. (2013). Are forensic experts biased by the side that retained them? *Psychological Science*, *24*, 1889–1897. https://doi.org/10.1177/0956797613481812

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. https://doi.org/10.1080/15374410701279602

National Center for State Courts, Ostrom, B. J., Ostrom, C. W., Hanson, R. A., & Kleiman, M. (2009). *Assessing consistency and fairness in sentencing: A comparative study in three states* (National Institute of Justice Grant No. 2003-IJ-CX-1015). https://www.ncsc.org/~/media/Microsites/Files/CSI/Assessing%20Consistency.ashx.

National Crime Prevention Centre. (2012). *A statistical snapshot of youth at risk and youth offending in Canada* (Catalogue No. PS4-126/2012E-PDF). https://www.publicsafety.gc.ca/cnt/rsrcs/pblctns/ststclsnpsht-yth/ssyr-eng.pdf

Neal, T. M. S., & Grisso, T. (2014). The cognitive underpinnings of bias in forensic mental health evaluations. *Psychology, Public Policy, and Law*, *20*, 200–211 https://doi.org/10.1037/a0035824

Nolan, A. & Stewart, L. (2017). *Low risk offenders: What does the research tell us?* (Correctional Service of Canada Research Report No. R-383).

http://publications.gc.ca/collections/collection_2017/scc-csc/PS83-3-383-eng.pdf

Penner, E. K. (2009). *Judges' perceptions of juveniles who assert their right to silence* (Master's thesis). http://summit.sfu.ca/item/9645

Pina-Sánchez, J. (2015). Defining and measuring consistency in sentencing. In J.V. Roberts (Ed.), *Exploring Sentencing Practice in England and Wales* (pp. 76–92). Palgrave Macmillan.

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. https://doi.org/10.1037/lhb0000161

Shepherd, S. M., & Anthony, T. (2018). Popping the cultural bubble of violence risk assessment tools. *Journal of Forensic Psychiatry & Psychology*, *29*, 211–220. https://doi.org/10.1080/14789949.2017.1354055

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. https://doi.org/10.1016/j.childyouth.2007.05.008

Sickmund, M., & Puzzanchera, C. (2014). *Juvenile offenders and victims: 2014 national report.* National Center for Juvenile Justice. https://www.ojjdp.gov/ojstatbb/nr2014/downloads/NR2014.pdf

Singh, J. P., Desmarais, S. L., Hurducas, C., Arbach-Lucioni, K., Condemarin, C., Dean, K., Doyle, M., Folino, J. O., Godoy-Cervera, V., Grann, M., Ho, R. M. Y., Large, M. M., Nielsen, L. H., Pham, T. H., Rebocho, M. F., Reeves, K. A., Rettenberger, M., de Ruiter, C., Seewald, 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.

https://doi.org/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.

https://doi.org/10.1080/07418825.2012.687513

Starr, S. B. (2014). Evidence-based sentencing and the scientific rationalization of

discrimination. *Stanford Law Review, 66*, 803–872.

Stevenson, M. C. (2009). Perceptions of juvenile offenders who were abused as children. *Journal

of Aggression, Maltreatment & Trauma*, *18*, 331–349.

https://doi.org/10.1080/10926770902901428

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. https://doi.org/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. https://doi.org/10.1007/s10964-013-

9926-4

Thompson, C. (2017). *Myths and factors: Using risk and need assessments to enhance outcomes

and reduce disparities in the criminal justice system*. National Institute of Corrections.

https://info.nicic.gov/nicrp/system/files/032859.pdf

Tonry, M. (2018). Punishment and human dignity: Sentencing principles for twenty-first-century

America. *Crime and Justice, 47*, 119-157. https://doi.org/10.1086/696948

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. https://doi.org/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.

https://doi.org/10.1177/1477370814525937

Viljoen, J. L., Jonnson, M., Cochrane, D., & Vargen, L. (2019). Do Risk Assessment Tools

Decrease Incarceration and Restrictive Placements for Offenders? A Systematic Review

and Meta-Analysis. *Law and Human Behavior, 43*, 397–420.

https://doi.org/10.1037/lhb0000344

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. https://doi.org/10.1037/lhb0000214

Wachter, A. (2015). *Statewide risk assessment in juvenile probation: JJGPS StateScan*. National

Center for Juvenile Justice.

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.

https://doi.org/10.1002/bsl.2162

Wise, R. A., & Safer, M. A. (2004). What U.S. Judges Know and Believe About Eyewitness

Testimony. *Applied Cognitive Psychology*, *18*, 427–443. https://doi.org/10.1002/acp.993