

What Do Juvenile Probation Officers Think of Using the SAVRY and YLS/CMI for Case Management, and Do They Use the Instruments Properly?

Laura S. Guy and Rebecca J. Nelson

Psychiatry, University of Massachusetts Medical School, Worcester, Massachusetts, USA

Samantha L. Fusco-Morin

Psychology, Fordham University, Astoria, New York, USA

Gina M. Vincent

Psychiatry, University of Massachusetts Medical School, Worcester, Massachusetts, USA

Juvenile probation officers (JPOs; $n = 71$) in the United States were interviewed three and ten months after the SAVRY or YLS/CMI was implemented in their office. Quantitative and qualitative analyses were used to explore their experiences using the instruments and adherence to practice guidelines. JPOs typically perceived the instruments as being 'somewhat' or 'very' helpful for guiding their case planning decisions. A frequently cited barrier to using both instruments in practice related to the increased length of time it took to complete reports; yet, at the same time, some JPOs also acknowledged that use of the measures forced them to gather important information about the youth's background and current situation that proved useful. Most JPOs (77%, $n = 33$ of 43) using the SAVRY expressed preference for a risk assessment model that emphasized use of appropriate professional discretion rather than a score-based approach. "Buy-in" for the instruments and the reported difficulties varied across sites. The present findings may inform recommendations specifically for delivering training on the SAVRY and YLS/CMI and, more broadly, strategies to promote their effective implementation in juvenile justice settings.

Keywords: juvenile probation, risk assessment, implementation, SAVRY, YLS/CMI

Probation is by far the most likely disposition for adjudicated juveniles in the United States, although the number of youth placed on probation has declined steadily during the past decade and a half. In addition, there is some evidence that the offenses for which youth are placed on probation are becoming more serious. In 2009, just over one quarter of all juveniles placed on formal probation had been arrested for person-related offenses, a sharp increase from the corresponding rate of 16% in 1985 (Livsey, 2012).

In many jurisdictions in the United States, juvenile probation officers (JPOs) can play a crucial role in shaping the path of a case through the judicial system, and may have

considerable influence on the dispositional process and outcome. For example, they often have responsibility to decide whether a case should be handled formally or informally, to make detention decisions, and to submit pre-disposition investigations with recommendations to the judge (Bilchik, 1999). Their decision-making regarding the appropriate programming for youth can lead to significant reductions in re-offense rates particularly with respect to various intervention programs (e.g., individual counseling, behavioral programs, interpersonal skills) and appropriate matching to services (Lipsey, 2009; Lipsey & Wilson, 1998; Vieira, Skilling, & Peterson-Badali, 2009). As such, the relative dearth in knowledge about JPOs' decision-making when forming their recommendations is an important domain that warrants additional investigation. As Mulvey and Iselin (2008) noted, despite the availability of instruments in

Address correspondence to Laura S. Guy, University of Massachusetts Medical School, Department of Psychiatry, Chang Building, 222 Maple Ave., Shrewsbury, MA 01545, USA. E-mail: Laura.Guy@umassmed.edu

juvenile justice to structure decisions about issues highly relevant to disposition, such as risk for future offending, juvenile justice professionals “make decisions based mainly on their intuition” (p. 38). The authors observed that although heavy caseloads and tight deadlines contribute to the lack of reliance on such decision support aids, “the ethos of the court also reinforces a reliance on unstructured professional judgment” (Mulvey & Iselin, 2008, p. 39).

The use of unstructured judgment for assessing risk for future offending and violence is highly problematic; meta-analytic data demonstrate this approach to be substantially less accurate than structured approaches (Guy, 2008; Hanson & Morton-Bourgon, 2009). Therefore, it is noteworthy that, despite the pervasiveness of the use of unstructured professional judgment for assessing risk and formulating intervention and case management plans, jurisdictions that follow the recommendations in the Juvenile Justice and Delinquency Prevention Act (JJJPA, 2002) now have in place some form of structured instrument for assessing risk, many of which are locally developed and not validated (Vincent, Terry, & Maney, 2009). Specifically, the JJJPA states that juvenile justice programs should assist states “. . . in the design and utilization of risk assessment mechanisms to aid juvenile justice personnel in determining appropriate sanctions for delinquent behavior” (p. 18). The Act also states delinquency should be addressed by programs “designed to reduce risks and develop competencies in at-risk juveniles that will prevent, and reduce the rate of, violent delinquent behavior” (p. 1). According to Shook and Sarri (2007), another factor that may have led to the increased use of structured decision-making procedures in juvenile justice is federal legislation to incentivize compliance with accountability-based sanctions (Juvenile Accountability Incentive Block Grant, P.L. 105–109).

USE OF RISK ASSESSMENT INSTRUMENTS IN JUVENILE JUSTICE

Despite indications of the growing use of risk assessment instruments, little is known about the actual extent of use of these measures, or about the ways in which such measures may impact professionals’ decision making. It appears only three surveys have been conducted on this issue. Sampling 37 states, Barton and Gorsuch (1989, as cited in Shook & Sarri, 2007) reported that in almost half (47%) of the states, a formal risk assessment instrument was used to inform post-dispositional decisions and in a third (30%) of the states it was used for some other formal classification. Twenty-two percent of states indicated that no formal assessment of classification was used. Surveying all 50 states, Towberman (1992) reported that some form of structured decision making (SDM) support aid that assessed risk factors was used in most of the 48 states that responded, but in an inconsistent and irregular manner.

However, an empirically-derived assessment procedure was used in only a small minority of states.

In the most recent survey, Shook and Sarri (2007) provided rich data from judges and probation officers in three courts in each of four states (Illinois, Indiana, Michigan, and Ohio). They asked respondents questions concerning their use and opinion of SDM, focusing on three components: risk assessment, needs assessment, and security classification.¹ JPOs in this study used locally developed risk assessment instruments. Two-thirds of respondents reported using at least one of the components in case processing and 35% reported using all three components. Most (63%) judges and JPOs reported using risk assessment frequently. Those who used risk assessment more often were more likely to view the practice as being of value.

IMPORTANCE OF SOUND IMPLEMENTATION PROCEDURES

Literature focusing on implementation science applied within behavioral health domains has been accruing for several years (e.g., Powell et al., 2012; Proctor & Rosen, 2008; Proctor et al., 2011; Proctor et al., 2009). Within the narrower field of violence risk assessment, commentary about and research on implementation of risk assessment instruments has emerged recently among adult (e.g., Crocker et al., 2011; Desmarais, Van Dorn, Telford, Petrila, & Coffey, 2012; Doyle, Lewis, & Brisbane, 2008; Kroppan et al., 2011; Nicholls, Petersen, Brink, & Webster, 2011; Nonstad & Webster, 2011) and adolescent (e.g., Desmarais, Sellers, et al., 2012) populations. From the few implementation studies in probation, an emergent key theme is the importance of obtaining “buy-in” from evaluators and other stakeholders regarding the usefulness of conducting risk assessments in general as well as the specific instrument chosen (Savaya, Monnickendam, & Waysman, 2000; Schneider, Ervin, & Snyder-Joy, 1996; Shook & Sarri, 2007).

Additional obstacles to successful implementation include current and potential resistance from judges and front line users. Shook and Sarri (2007) reported that court professionals might have resisted the use of SDM in case processing because of concerns that it would usurp their professional judgment and be time consuming and/or difficult to use. In a juvenile probation study of the implementation of validated, comprehensive risk assessment instruments, Vincent, Paiva-Salisbury, Cook, Guy, and Perrault (2012)

¹Although Shook and Sarri (2007) examined risk assessment and needs assessment separately, contemporary risk assessment instruments such as the SAVRY and YLS/CMI contain criminogenic needs/dynamic risk factors. Although some argue for the separation of risk and need assessments (see Baird, 2009), such a distinction is moot when using the newer generation risk assessment instruments developed to directly inform risk management efforts, which contain criminogenic need factors as part of the risk assessment.

reported that JPOs anticipated the following barriers to the impending introduction of a risk instrument to their system: (1) the length of time it would take to complete it, (2) resistance to change, and (3) devaluing JPOs' role and judgment.

Other obstacles can compromise the integrity of the implementation process (for a thorough review, see Vincent, Guy, & Grisso, 2012). For example, in one study, unless JPOs had clear office policies regarding how to apply "results" from the risk assessment instrument in case management decisions, and unless they had training on these policies, they tended not to make decisions commensurate with youths' risk levels, regardless of whether they had received training on the instrument (Vincent, Guy, Gershenson, & McCabe, 2012).

JPOs' UNDERSTANDING OF RISK ASSESSMENT MODELS

An important aspect of promoting ongoing fidelity following implementation of a risk assessment instrument is ensuring that all users continue to adhere to the same methods for completing the measure. An issue that has not been investigated is the degree to which users of these instruments retain a thorough understanding of the foundational aspects of the models of risk assessment on which these tools are based. In brief, risk assessment instruments differ in the extent to which they apply structure to the assessment process. The two most well-validated instruments for juveniles, the *Youth Level of Service/Case Management Inventory* (YLS/CMI; Hoge & Andrews, 2002) and the *Structured Assessment of Violence Risk in Youth* (SAVRY; Borum, Bartel, & Forth, 2006), are based on different models of risk assessment. Users of the YLS/CMI score a set of items using specific instructions, sum the item scores, and use the resulting total score to assign a youth to one of four categorical risk levels. Used in this way, the YLS/CMI would be an *actuarial* risk assessment instrument because there are fixed, *a priori* determined rules for how the risk information is combined (see Meehl, 1954). However, the manual instructs users to then consider the presence of additional needs and special considerations related to the family and the youth. At the final step, in a departure from the actuarial model, YLS/CMI users provide their professional estimate of risk for the case using a four-level categorical scheme.

In contrast, the SAVRY is based on the *Structured Professional Judgment* (SPJ) model (e.g., Douglas, Hart, Webster, & Belfrage, 2013; Douglas & Kropp, 2002). The SPJ approach relies on evidence-based guidelines to help professionals identify which people pose a risk of violence and what steps are necessary to prevent violence. Depending on the particular SPJ instrument and the setting in which it is intended to be used, relatively more or fewer administration steps are recommended. Across all SPJ

measures, administration involves three broad categories: identifying facts, making meaning of those facts, and taking action to manage risk (see Douglas, Groscup, Hart, & Litwack, 2014; Guy, Douglas, & Hart, in press). Evaluators communicate their opinion regarding the evaluatee's risk for violence in part by making an overall summary judgment, also referred to as the summary risk rating (SRR), using a tri-level ranking system (low, moderate, high). SPJ instruments contain items that have operational definitions and instructions for assigning ratings (also using a three-level categorical scheme). In addition to rating whether a risk factor is present, users of SPJ instruments also consider the *individual manifestation* of each risk item. That is, how is it displayed by the individual (e.g., for substance use: Is the youth using cocaine or marijuana? Has frequency and/or severity been stable, increasing, or decreasing?). Also critical to estimating overall risk and developing risk management recommendations, evaluators make an informed judgment regarding the degree to which each risk factor has *individual relevance* for that person's risk for violence (e.g., a person may be using marijuana regularly at home before bed, but her marijuana use may not be related to her risk for another physical assault against a teacher). Finally, in the SPJ model, there is not necessarily a linear relation between the number of risk factors present and an individual's overall level of risk. That is, under some circumstances, a person with only few highly relevant risk factors present could be judged as being at high risk (and vice versa). This concept has been referred to as *linearity*. Although much research has been conducted regarding evaluators' agreement about the SRR and the presence of SAVRY items, to date there has been no empirical investigation regarding the degree to which evaluators understand the concepts of manifestation, relevance, and linearity when using an SPJ instrument.

PRESENT STUDY

This study investigates JPOs' expectations of and experiences using the YLS/CMI and SAVRY for case planning with probationers. It extends earlier work examining the implementation of these instruments in the *Risk/Needs Assessment in Juvenile Probation: Implementation Study* (RNAJP; Vincent, Guy, Gershenson, et al., 2012; Vincent, Paiva-Salisbury, et al., 2012). RNAJP was a multisite, pre-post longitudinal study in which data were collected in three juvenile probation offices (in each of two states) in exchange for technical assistance to support their risk assessment implementation. One state chose to implement the YLS/CMI and the other selected the SAVRY. Detailed procedures of the study are described elsewhere (Vincent, Paiva-Salisbury, et al., 2012).

The present study had three aims. First, we examined JPOs' broad perceptions about the usefulness and

difficulties of the instruments. Next, we investigated JPOs' experiences using the instruments with respect to (1) rating specific items and (2) making overall ratings about risk level. Third, among SAVRY users only, we investigated the degree to which JPOs reported rating the items and making a SRR in a manner consistent with the SPJ model's concepts of manifestation, relevance, and linearity.

METHOD

Participants

The sample comprised 71 JPOs across the six probation offices who participated in the RNAJP study.² Three offices were located in a northeastern state and implemented the YLS/CMI (27 JPOs), and three were in a southern state and implemented the SAVRY (44 JPOs). JPOs participated in two-day training workshops conducted by a co-author of the instrument, typically completed three post-workshop practice cases and received feedback, and participated in booster training six months following the initial workshop (see Vincent, Paiva-Salisbury, et al., 2012).

JPOs on average were 35.49 ($SD = 9.7$) years old, men (52.1%, $n = 37$), and Caucasian (63.2%, $n = 43$; African-American: 33.8%, $n = 23$; Other: 2.9%, $n = 2$), data were missing for three JPOs). Most had a bachelor's degree (75.8%, $n = 50$) and a few had a master's degree (24.2%, $n = 16$; data were missing for five JPOs). YLS/CMI users had significantly more years of experience working with juvenile justice-involved youth ($M = 14.3$, $SD = 10.0$) than SAVRY users ($M = 9.72$, $SD = 9.15$); $t(130) = 2.74$, $p = .007$; $d = .48$).

Risk Assessment Instruments

YLS/CMI (Hoge & Andrews, 2002).³ The YLS/CMI comprises 42 dichotomous items across eight risk/need scales. Item scores (0 or 1) are summed to yield a total score to assign youths to one of four categorical risk rankings: low (0–8), moderate (9–22), high (23–34), and very high (35–42). After considering additional needs and special considerations, evaluators decide whether the risk level should be adjusted (i.e., apply a *professional override*). There is good evidence for the instrument's interrater reliability and predictive validity (Olver, Stockdale, & Wormith, 2009; see Hoge, Vincent, & Guy, 2012, for a review).

²These 71 subjects were the subsample of JPOs from the RNAJP study ($N = 111$; 88 JPOs and 23 administrators) who completed the last of three interviews. Administrators were excluded from the present analyses because they typically did not conduct the risk assessments, and so many of our research questions were not applicable to them.

³The revised YLS/CMI manual (Hoge & Andrews, 2010) was published at the end of this project.

SAVRY (Borum et al., 2006). This SPJ instrument comprises six protective factors and 24 risk factors, and evaluators identify additional factors if relevant for a given case. Users make formal ratings regarding the presence of each risk (low, moderate, or high) and protective (present or absent) factor. They also consider the manifestation and individual relevance of each item. They then make summary judgments about the youth's overall level of risk and need for intervention (low, moderate, or high).⁴ Research has demonstrated sound inter-rater reliability of the instrument with JPOs in the field (Vincent, Guy, Fusco, & Gershenson, 2012) and predictive validity comparable to the YLS/CMI (see Hoge et al., 2012).

Measures and Procedures

Eight trained researchers interviewed JPOs three times about their case management practices and experiences supervising probationers: prior to implementation of and training on the risk assessment instrument, three months after implementation, and 10 months after implementation. Only data from the two post-implementation interviews are reported here, and subsequently are referred to as the first and second interviews, respectively. Most interviews were conducted over the phone (some in person) and all were audiotaped and transcribed. Staff turnover or difficulty reaching JPOs led to some attrition over time (for details, see Vincent, Paiva-Salisbury, et al., 2012), but every JPO was interviewed at least once. No JPOs declined to participate in the study. Of the 71 JPOs interviewed at the second interview, 68 also completed the first post-implementation interview.⁵

At both post-implementation interviews, JPOs were asked to describe whether the SAVRY or YLS/CMI was useful for making recommendations about disposition/placement, service referrals, and level of supervision. They were asked to discuss aspects of the instruments that were most and least helpful to them, including whether certain items were relatively more difficult to rate. JPOs were queried regarding how challenging it was to make the SRR (for SAVRY users) or the final risk estimate that could lead to a professional override (for YLS/CMI users). They also were asked to describe any factors they believed could make that process easier.

In the final interview, SAVRY users were queried regarding the ways in which they used the instrument within the framework of the SPJ model. They were asked to describe the process they followed when making the

⁴Additional steps addressing formulation, scenario planning, and risk management are key components of newer SPJ instruments (Douglas et al., 2013; Douglas et al., 2014; Guy et al., in press; Hart et al., 2003).

⁵To maximize power, we report data from any JPO who completed the particular post-implementation interview. Thus, *ns* for interviews 1 and 2 vary.

SRR and to answer specific (and face valid) questions to assess—in a rudimentary manner—the degree to which they understood assessment practices consistent with the SPJ model: (1) *Could the same risk factor mean different things for different people, or look differently for different people? Why?* (concept of manifestation); (2) *Could the same risk factor be more or less relevant for one person compared to another? Why?* (concept of individual relevance); and (3) *If two youth have the same five risk factors, will they or won't they have the same risk level? Why?* (concepts of linearity and individual relevance).

Data Analysis

A mixed-methods data analytic approach was used to assess how often certain assessment related events occurred, and to better understand the nature of the JPOs' experiences using these instruments. Quantitative data were collected at both interviews and comprised frequencies of responses to close-ended questions and consensus ratings of two coders for items that were categorically rated based on responses during the interview. Qualitative data were coded using a content analysis (Berg, 2004) approach through the use of verbatim interview transcripts to uncover common themes (see also Allen, Watt, & Hess, 2008). The content analysis comprised several steps. First, two researchers read 10 randomly selected transcripts of interviews with SAVRY users and developed a set of initial themes that captured JPOs' perceived advantages and disadvantages associated with use of the instrument and their experiences associated with use of the instruments in the field. Because the areas of inquiry regarding use of the measures were defined *a priori*, certain categories were expected to emerge during data analysis of the open-ended questions. Nevertheless, we used a constant comparison process (Pope, Ziebland, & Mays, 2000) in which each data element coded was compared with all other data elements coded to that point to establish analytical categories; as many categories as needed were added to reflect as many of the nuances in the data as possible. Following this initial step, 35 themes were generated that reflected a detailed breakdown of JPOs' satisfaction with and use of the SAVRY in the field.

Next, a more parsimonious set of second order themes was developed (comprising 19 themes). Two researchers coded two thirds of the SAVRY transcripts independently using this set of themes to examine whether each theme was present (yes or no) and then coded verbatim the data from the transcript relevant to the particular theme. The coders discussed the reasons for their independent ratings and made final consensus ratings for each case by discussing coding discrepancies (both presence/absence of the themes as well the verbatim data from the transcript supporting a decision of presence). The remaining one third of the SAVRY transcripts was coded independently by one of the original coders and a new trained coder using the

parsimonious set of themes. Final consensus ratings were achieved using a similar procedure. All analyses were based on final consensus ratings.

Because the majority of questions asked during interviews with YLS/CMI users were the same as those asked of SAVRY users, the YLS/CMI transcripts were coded using the parsimonious set of themes described above. Researchers coded all transcripts independently and consensus ratings were reached. The few questions specific to YLS/CMI users were close-ended and therefore were analyzed quantitatively.

RESULTS

Below we present results related to each of the study's three aims: (1) JPOs' perceptions about the usefulness and difficulties of the instruments; (2) their experiences rating specific items and making summary risk judgments; and (3) the degree to which JPOs reported using the SAVRY in a manner consistent with assessment practices of the SPJ model. For each aim, we report results for qualitative and then quantitative analyses. Quantitative themes are presented separately for YLS/CMI and SAVRY users. After presenting results related to these three aims, we report findings about unanticipated qualitative themes discovered during the analytic process.

Aim 1. Perceptions About the Usefulness and Difficulties of the Instruments

Perceived Helpfulness of the Risk Instruments

Quantitative Results

Twenty of 24 (83%) YLS/CMI users indicated at the first interview that the instrument had some or many helpful characteristics; no significant change was observed at the second interview, with 22 of 27 JPOs (82%) reporting similarly positive perceptions. Among SAVRY users, 29 of 31 (94%) reported the instrument had some or many helpful characteristics after using it for three months; a slight drop in this rate was observed at 10 months (39 of 44; 89%). Next, we examined consensus ratings of two raters regarding the degree to which JPOs' responses indicated they perceived the instrument to be 'not at all,' 'somewhat,' or 'very helpful' for each type of case management decision (see Table 1). Overall, the majority of users of both instruments perceived them to be 'somewhat' or 'very' helpful for making recommendations about disposition, services, and level of supervision across both follow-up periods.

Qualitative Themes

YLS/CMI. Relatively few (13 of 27) JPOs elaborated when prompted to discuss specific aspects of the YLS/CMI

TABLE 1
Perceived Helpfulness of Instruments for Case Management Decisions

	Post-implementation Interview 1 <i>n</i> (%)			Post-implementation Interview 2 <i>n</i> (%)		
	Not at all	Somewhat	Very	Not at all	Somewhat	Very
SAVRY						
Disposition/Placement	4 (11.1)	20 (55.6)	12 (33.3)	0 (0.0)	14 (36.8)	24 (63.2)
Service referrals	3 (7.9)	14 (31.8)	21 (55.3)	2 (4.9)	10 (24.4)	29 (70.7)
Level of Supervision	1 (3.6)	15 (53.6)	12 (42.9)	2 (5.9)	8 (23.5)	24 (70.6)
YLS/CMI						
Disposition/Placement	3 (33.3)	1 (11.1)	5 (55.6)	1 (7.1)	7 (50.0)	6 (42.9)
Service referrals	1 (5.6)	13 (72.2)	4 (22.2)	2 (7.4)	5 (35.7)	7 (50.0)
Level of Supervision	3 (21.4)	3 (21.4)	8 (57.1)	5 (29.4)	5 (29.4)	7 (41.2)

Note. *Ns* vary depending on the availability of data to make consensus ratings.

they perceived to be helpful. The most frequently identified theme was use of the instrument to “back up” their opinions about risk level and recommendations regarding services and level of supervision, which they believed they would have reached using only their professional experience. For example:

...I've worked (here) for 11 years (and) have a Master's degree in counseling...I'm pretty confident in my ability...I've not administered the YLS and been overwhelmed with “Oh wow! I didn't consider that!”...it's not a real eye-opener for...pointing out specific needs. Now for the case supervision plan that coincides with that, that's where it's been helpful because at least we have documentation that it's not just me with a gut feeling or a judgment call. We have a legitimate tool... it backs me up...

In some cases, YLS/CMI results were valued only if they supported the JPO's opinion. When asked whether she or he uses the instrument to make recommendations, a JPO remarked:

Yes, to the extent that I say it's supportive of what I'm recommending. I don't do it the other way around—I don't look at the score and say ‘because she scored this I should recommend that.’ I do my recommendation and the YLS and the scoring can be reflective/supplementary to what I'm recommending.

Other less frequently identified themes related to the perceived helpfulness of the instrument included more comprehensive gathering of risk-related information, the “user friendly” aspects associated with having the YLS/CMI items and scoring guidelines incorporated into an electronic data management system, and the consistency across probation offices for assessing risk for reoffending using the same criteria.

SAVRY. JPOs using the SAVRY provided more comments about usefulness than YLS/CMI users. The primary themes that emerged were the SAVRY's emphasis on professional judgment or lack of reliance on a total score,

enhanced data collection that led to more useful service recommendations, and increased knowledge about risk factors.

Emphasis on professional judgment (as a positive). Most JPOs (33 of 43, 77%) reported preference for an instrument that did not produce a total score corresponding to a risk level. Most responses relevant to this theme mentioned the importance of having one's professional identity, training, and experience taken seriously and respected by decision-makers, rather than having case related decisions abdicated to an instrument. Some JPOs identified specific shortfalls associated with fixed risk levels. One JPO expressed this sentiment by comparing the SAVRY to an actuarial instrument used previously in that office:

(The previous instrument) said between 0 and 30 is this; and 30 and 60 is this...sometimes the child may have scored in that middle range...there may have been some critical items that jumped out at you...and you felt like that child should be a high kid. You had to mark that kid as a moderate kid...because of what the risk assessment said.

Enhanced data collection. As part of the implementation process, JPOs were instructed to use semi-structured interview guides with youth and parents, which entailed gathering more detailed information than they previously had obtained. One SAVRY user commented:

...a lot of times I will find out information that I did not know before, like one kid let me know there had been a time where he thought about killing himself and later on found out that this child probably needs to be evaluated for depression even though on the outside it doesn't look like anything is going on so (the SAVRY)...questions (go) a little bit deeper than I ask.

Increased knowledge about risk factors. Several JPOs indicated their knowledge about empirically based risk factors increased after completing the SAVRY training workshop and becoming familiar with the manual. For example:

...it's just really helped me to understand behavior and look at it from a different angle...the age...they started offending...the early initiation of violence...the history of maltreatment...we always ask those things, but...I didn't realize that had much of an impact...

Other minor themes observed related to positive aspects of the SAVRY included the promotion of objectivity and transparency in the risk assessment process, having a research based procedure "back up" their professional opinion and recommendations, the utility of the SAVRY for tracking changes in risk over time, ease of communication between professionals trained in the same instrument (e.g., speaking the "same language"), and increased ability to "pinpoint" the most critical criminogenic needs to be targeted for treatment.

Perceived Difficulties of the Instruments

Quantitative Results

At the first interview, eight of 24 (33%) YLS/CMI users reported having some problems initially and nine of 24 (38%) reported having persistent problems. At the second interview, relatively fewer YLS/CMI users reported having some problems that improved over time (three of 27, 11%), but considerably more JPOs reported having problems with the instrument that had not improved (16 of 27, 59%). The reverse was true for SAVRY users. Initially, 11 of 38 JPOs (29%) reported having some problems that desisted over time, and 18 of 38 (47%) reported having at least one persisting problem. At the second interview, even fewer SAVRY users reported experiencing problems: 10 of 44 (23%) had some problems that improved, and 15 of 44 (34%) reported a problem that persisted.

Qualitative Themes

Amongst both YLS/CMI and SAVRY users, the most frequently reported disadvantage was the increased length of time required to complete the pre-dispositional report (into which the instruments' 'results' were incorporated).

YLS/CMI users. Many JPOs cited the redundancy between the information gathering and decision-making practices they were using prior to YLS/CMI implementation and the practices put into place with the YLS/CMI. They ultimately voiced complaint about the increased drain on their time associated with what was perceived to be an unnecessary task. A prototypical quote illustrating this theme was:

...it's additional paperwork, it's time consuming...and that it doesn't allow me to do anything different prior to the YLS.

SAVRY users. The majority of SAVRY users expressed a preference for an instrument that incorporated professional

judgment, but a few expressed an inclination toward using an instrument that assigned a risk level automatically. For example, the following exchange occurred when a SAVRY user was asked whether she or he would prefer an 'automated' measure:

...Oh that would make things real simple wouldn't it, you mean like if I entered all of this into the computer and it said this child's risk is?...That would be really good...no room for my own personal judgment.

For a minority of SAVRY users, this preference was rooted in their desire to have a more structured instrument that provided immunity against negative outcomes. For example, when asked about whether she or he preferred having discretion versus a tool that was more standardized, another JPO replied:

Yes, you would like something to be more concrete but you like to be able to use your own discretion and...make the decision based on what you observe...But also, I would like something a little bit more concrete for liability reasons...if I have a child (who)...may be considered high but I rated him moderate...and he goes out and does something crazy...then (I would be asked) 'why did you rate this child at this point?' (and I would respond) 'Well I did it because the tool said I could rate him how I wanted to and it was my opinion.'

Other SAVRY users voiced concern that individual differences in JPOs' attitudes, orientation towards retribution, or tolerance for risk taking could affect the way in which the SRR is assigned. For example:

...it's kind of one of those things where you can kinda be your worst enemy. You're gonna second-guess yourself and kinda say, well, you know, should it be low? Maybe I should bump it up to moderate, just to be on the safe side.

Some SAVRY users expressed a desire to reduce the perceived subjectivity associated with assigning the SRR. For example:

I think if there was an actual scoring sheet where you count it up... I count everything up and look how many moderates does he have...I'd really like to see that...where you don't have opinions placed into it as much so if the score says he's moderate then he's moderate...end of story.

An unexpected finding that emerged suggested concern about misusing the flexibility of the SPJ approach to avoid additional work (because supervision requirements were tied to risk level by policy). For example, within the context of describing the approach she or he takes when assigning the SRR, one JPO reported:

Downloaded by [208.80.101.212] at 14:31 03 September 2014

I don't find it very difficult (to make the SRR)...it's...more hard to (move from)...a moderate and a high than it is a low to a moderate...you don't want to put high if they don't need to be, but yet...you know...that requires more of your time...then again you don't want to put them as a moderate if you know they need to be supervised.

A few JPOs indicated the SAVRY would be more helpful to less experienced JPOs.

I've been doing this for a long time. I know what to ask and what to look for...(for) some younger probation officers I think...it's a very good thing because it makes them sit down and get to know the family and what's going on.

Aim 2. Experiences Rating Specific Items and Making Overall Risk Ratings

Experiences Rating Instruments' Items

YLS/CMI users. Few JPOs reported finding specific items difficult to rate at the second interview (nine of 25, 36%). Items in the Attitudes/Orientation domain were identified most frequently ($n = 3$), followed by the Substance Abuse and Leisure/Recreation domains. Two items in the Family Circumstances/Parenting domain, Poor Relations (father-youth) and Poor Relations (mother-youth), were identified as being challenging to rate ($n = 2$). Among JPOs who elaborated regarding their experienced difficulty rating items, three of eight (37.5%) reported the challenge related to having insufficient data available to rate the item, and five of eight (62.5%) identified difficulty related to lack of clarity in the manual's item description. Some responses suggested frustration with the dichotomous item ratings. For example:

The YLS...in some respects is vague... Like where it says some delinquent/positive friends... Antisocial/procriminal—those kinds of things. They don't actively reject help, but it's not so cut and dry.

SAVRY users. At the second interview, 25 of 44 JPOs (57%) reported finding one or more items difficult to rate. Among JPOs who reported some difficulty, 23% ($n = 10$) indicated they had insufficient information, 21% ($n = 9$) cited lack of clarity in the manual's item description, and

11% ($n = 6$) reported that the source of the difficulty varied depending on the particular item.

Most SAVRY items identified as being difficult to rate were on the Social/Contextual scale: Peer Rejection ($n = 2$), Stress and Poor Coping ($n = 3$), Lack of Personal/Social Support ($n = 3$), and Community Disorganization ($n = 2$). Three items were identified on the Historical scale: History of Violence ($n = 1$), Childhood History of Maltreatment ($n = 2$), and Poor School Achievement ($n = 1$). Three items were identified on the Individual scale: Risk Taking/Impulsivity ($n = 2$), Substance Use Difficulties ($n = 1$), and Low Empathy/Remorse ($n = 3$). In addition, one JPO mentioned experiencing a general sense of subjectivity about the items. Another JPO reported being confused about whether rating an item as 'high' indicated that the variable of interest was present or absent for the youth. Two JPOs reported protective factors generally were difficult to rate. A few JPOs spontaneously praised the manual for its thoroughness, detail, and helpfulness.

Experiences Making Overall Risk Ratings

YLS/CMI users. At the time of data collection, Part IV of the YLS/CMI (i.e., decision whether to apply a professional override) was not yet implemented as standard practice in the sites. However, JPOs had received instruction on completing this step during the training, and some users reported considering the need for an override when using the instrument in practice. Approximately one quarter of YLS/CMI users (8 of 28, 29%) reported having never applied a professional override, despite having wanted to do so. Of those who had, most found making the override to be relatively easy (see Table 2). For example:

...the YLS is a good tool, but in the end you have to use your professional judgment...It might have been a moderate became a higher or lower (*risk*) kid. Never has it been from one end of the spectrum to the other.

SAVRY users. Most (28 of 34, 82%) JPOs reported at the first interview that they did not find it difficult to designate someone as High risk. This percentage increased slightly at the second follow-up (37 of 44 JPOs; 84%). JPOs also were asked questions to assess the degree to which they found making the SRR to be easy or challenging, and their views

TABLE 2
Perceived Ease With Which Structured Judgments about Overall Risk Could Be Made

	Post-implementation Interview 1 <i>n</i> (%)			Post-implementation Interview 2 <i>n</i> (%)		
	Not at all	Somewhat	Very	Not at all	Somewhat	Very
SAVRY Summary Risk Rating	2 (8.7)	9 (39.1)	12 (52.2)	0 (0.0)	21 (47.7)	23 (52.3)
YLS/CMI Professional Override	1 (4.8)	3 (14.3)	17 (81.0)	1 (3.7)	5 (18.5)	18 (66.7)

Note. *Ns* vary depending on the availability of data to make consensus ratings.

about how the process could be facilitated. Table 2 presents the frequencies for consensus ratings regarding the level of difficulty indicated in this task by each JPO's response. Over 90% of JPOs using the SAVRY reported at the first interview that making the SRR was 'very easy' (12 of 23, 52%) or 'somewhat easy' (nine of 23, 39%). By the second interview, the majority (23 of 44, 52%) reported the SRR was 'very easy' to generate.⁶

Among the minority of JPOs who described the process of selecting the SRR as being difficult at the first interview, only one provided a narrative explanation:

...I don't really find it super clear...it's just very subjective...I don't think there's really enough of anything concrete to say, okay, this is definitely high, medium or low...I think if there was something in the summary that said, "If you have like 5 high, 4 medium, 2 low, therefore it has to be a medium plus" or something.

Process for generating the SRR. JPOs were asked to describe how they approached the task of generating the overall summary judgment about risk. Responses indicated they generally followed the linearity concept, in which higher risk levels are associated with the presence of relatively more risk factors. A representative quote regarding consideration of the number of risk factors present was:

...pretty much I go back and review the number of lows...moderates and...highs and I compile them and I just come up with a decision. It's pretty easy, just looking at everything.

Only two JPOs provided a response that suggested she or he adhered to one of the core principles of the SPJ model in which the evaluator is permitted to deviate from linearity—either upward or downward in his or her estimate of overall risk—based on the presence of a few or even one critical risk factor. For example:

...I do look at the numbers that are rated low, moderate, high and then I also look at which items that I marked critical. I may have a kid who has more moderate items checked but... (if) I...marked them critical in areas such as substance abuse, low interest at school, poor parental management, I'd probably rate that kid high.

Some JPOs identified risk factors to which they routinely paid particular attention when assigning the SRR. The most commonly mentioned factor was History of Violence ($n = 8$). Most of the other items identified also were on the

Historical scale: History of Non-Violent Offending ($n = 2$), Early Initiation of Violence ($n = 1$), History of Self-Harm or Suicide Attempts ($n = 1$), Exposure to Violence in the Home ($n = 2$), Childhood History of Maltreatment ($n = 4$), Parental/Caregiver Criminality ($n = 1$), Early Caregiver Disruption ($n = 1$), and Poor School Achievement ($n = 2$). Only two items on the Social/Contextual scale were identified: Poor Parental Management ($n = 2$) and Community Disorganization ($n = 1$). Three items on the Individual scale were identified: Substance Use Difficulties ($n = 2$), Negative Attitudes ($n = 1$), and Anger Management ($n = 1$). Few JPOs considered Protective factors. Finally, one JPO reported that she or he placed emphasis on history of mental health problems as a risk factor.

Some JPOs reported considering the number of risk factors marked as high, moderate, or low (one JPO identified using a specific percentage cut-off), but then applying that information in a way that deviated from SPJ methods. For example, one JPO reported taking "an average" of the factors. Another JPO remarked:

I look and see what's the majority that I have circled, have I circled more moderate...then okay, it's moderate...I could have two or three high but if I feel like in my mind that it averages out then that child will be low so that right there, the summary part is a little bit left up to chance...

Aim 3. Among SAVRY Users, To What Degree Did JPOs Report Using the Instrument in a Manner Consistent With Principles of the SPJ Model?

In an attempt to assess in a more structured manner JPOs' understanding and application of basic SPJ rating practices, at the second follow-up they were posed three forced choice (yes/no) questions (Table 3 presents the frequencies), and asked to elaborate following their initial response. The first question, *Could the same risk factor mean different things for different people, or look differently for different people?*, was designed to assess their understanding of the concept of individual manifestation. Responses indicated that just over 90% ($n = 40$) of JPOs' understood that a single risk factor could be present for two youth for different reasons.

JPOs were asked *Could the same risk factor be more or less relevant for one person compared to another?* to assess their understanding of the concept of individual relevance. Most JPOs (89%, $n = 39$) replied in a manner that suggested that they understood the concept of individual relevance well. However, when asked the follow-up question, *If two youth have the same five risk factors, will they or won't they have the same risk level?*, to evaluate their understanding of the principle of individual relevance and, more indirectly, the concept of linearity, only a minority of respondents indicated (correctly) that youth with the same risk factors present would not necessarily have the same

⁶Although a few JPOs noted that the SRR is easier to make when there is a wide distribution on item ratings, most SAVRY users did not provide a detailed answer when asked to elaborate regarding the ease with which they felt able to arrive at a SRR or about any information or procedure that could be helpful for making the process easier.

TABLE 3
JPOs' Knowledge of the SPJ Principles of Linearity, Manifestation, and Individual Relevance

Question	Yes, <i>n</i> (%)	No, <i>n</i> (%)
If two youth have the same 5 risk factors, will they (by default) have the same risk level?	20 (45.5)	24 (54.5)
Could the same risk factor be more or less relevant for one person compared to another?	39 (88.6)	4 (9.1)
Could the same risk factors mean different things for different people?	40 (90.9)	4 (9.1)

Note. "Correct" answers to questions 1–3 are no, yes, and yes, respectively.

level of overall risk. When asked to elaborate, these 'correct' responders tended to provide responses consistent with the theme that youth had to be rated on a case-by-case basis because they are all "different." Conversely, almost half (46%, $n = 20$) of the JPOs reported (incorrectly) that two youth who have the same five risk factors present necessarily will have the same risk level. When queried about their responses, these JPOs tended to provide vague elaborative narratives lacking explanatory substance.

Post-hoc Themes Identified

Several themes related to implementation issues not part of the *a priori* constructs that served as the focus of the present study emerged during coding.

YLS/CMI users. The most prevalent theme was frustration associated with lack of buy-in from judges and attorneys, and the consequent lack of impact on case planning and risk management activities. For example:

...our judges (are not) on board with it...we can give all the information...but it doesn't...affect the disposition...it makes me aware of where the most necessary interventions are needed...for our own information and getting to know the youth the YLS is useful. For making decisions it's not (useful).

Another JPO, speaking about whether she or he uses the instrument to make decisions about supervision level, stated:

No... I have been trying to tell the judges about the YLS and they look at me like I'm crazy. I try to tell them that I don't need to see a kid as much as I have to because they are a low level...Usually, even the low level ones that I should be able to see one time every six weeks or so, I see them one time a month so I don't get in trouble.

SAVRY users. A theme emerged related to need for training in interviewing skills. Many JPOs indicated they used the semi-structured interview guide as an inflexible series of questions, all of which *had* to be asked. Several JPOs expressed frustration, which they attributed toward the SAVRY, that the information obtained from separate interviews with the youth and parent at times was discrepant. However, conducting separate interviews is sound clinical

practice, and so this policy was put into place when the SAVRY was implemented.

Finally, some JPOs expressed concern that more time was devoted to the assessment process at the expense of time supervising youth in the field. Other SAVRY users suggested that risk assessment procedures could be improved by the creation of an intake unit.

DISCUSSION

Our findings about JPOs' experiences using the YLS/CMI and SAVRY in their day-to-day practice add to the very small but developing literature base on the experiences of frontline users of validated risk assessment instruments. In addition, this is the first study in which an attempt was made to investigate the process through which evaluators (here, probation officers) using an SPJ instrument arrive at an estimate of overall level of risk, and whether that process is consistent with some of the assessment practices fundamental to the SPJ approach (individual manifestation and relevance, and linearity). As such, the study serves to bridge the well-acknowledged practitioner-researcher divide.

For several reasons, we cannot interpret our findings within the context of competing approaches (actuarial, SPJ) to risk assessment. First, the qualitative data from YLS/CMI users were substantially less rich than data provided by SAVRY users. More importantly, however, the instruments were implemented slightly differently in the two states, which could have affected JPOs' perceptions of them. Also, YLS/CMI users had significantly more experience as JPOs, likely related to differences in hiring procedures and salaries between the states. The more experience JPOs have, the greater the barriers administrators may have to overcome when attempting to implement risk assessment instruments (Vincent, Paiva-Salisbury, et al., 2012). Finally, social psychological research indicates that people tend to rate stimuli to which they have been exposed more positively than similar but unfamiliar stimuli (i.e., the familiarity principle, or mere-exposure effect; Bornstein, 1989; Zajonc, 1968). Merely because of their exposure to the YLS/CMI or SAVRY, JPOs may have rated the instrument positively. For this reason, coupled with the absence of a comparison group that did not experience implementation of a risk assessment instrument, it would be

difficult to attribute JPOs' perceptions about the experience of following new procedures that involved use of a risk assessment instrument to the particular instrument used.

Implementation Practices

High-quality implementation strategies are crucial in probation settings for risk assessment practices to effectuate change in decision-making. For example, Shook and Sarri (2007) found that JPOs were not likely to use a standardized instrument developed specifically for probation officers in Indiana, an outcome that was attributed to poor implementation efforts. There also is some evidence to suggest that anticipated system changes might take some time to materialize. For example, in a study examining the implementation of the Level of Service Inventory-Revised (LSI-R; Andrews & Bonta, 1995) in adult probation offices in a large state (Flores, Lowenkamp, Holsinger, & Latessa, 2006), risk level predicted recidivism, but the effect size was substantially larger when the instrument had been in place for at least three years (and used by JPOs with adequate training). The authors reasoned that the agency's experience with the instrument, as well as the extent to which the LSI-R became integrated into the routine "correctional landscape, thereby becoming a larger part of the agency's decision-making protocol" (Flores et al., 2006, p. 527) might have accounted for the finding. Moreover, implementation is difficult in organizations in which staff members are cynical about the changes (Farrell, Young, & Taxman, 2011). Thus, quality training with and buy-in from JPOs appear to be necessary pre-requisites to successful implementation of risk assessment systems. In the present study, a careful approach was coordinated to ensure that sound implementation practices were followed, in line with a model outlined by Vincent, Guy, and Grisso (2012). This included obtaining buy-in from relevant stakeholder groups (i.e., probation chiefs, JPOs, and attorneys) upfront (see Fixsen et al., 2005) by including them in the instrument selection process or discussions about how risk assessment information would be used and communicated in court.

The Role of Professional Judgment

The majority of SAVRY users cited as a positive aspect of the instrument its emphasis on the application of professional judgment. Indeed, 77% reported *not* wanting a measure that yielded a total score that corresponded with a risk level classification. On the other hand, nearly one quarter of SAVRY users (23%) reported they would prefer to use a risk assessment instrument that provided a total score. Some of these JPOs indicated they were in favor of such a high degree of structure because of concerns that any degree of subjectivity potentially could allow rater bias to

influence the assessment. However, research with these very JPOs indicated they demonstrated good inter-rater agreement when using the SAVRY in the field (Vincent, Guy, Fusco, & Gershenson, 2012). As such, there is little empirical justification for these apprehensions.

Despite concerns expressed by a minority of SAVRY users, professional discretion was broadly viewed as a positive aspect of the instrument. This is consistent with other surveys of probation officers' perceptions regarding the role of discretion in case planning and management activities. For example, Schneider and colleagues (1996) studied probation officers' perceptions of the implementation of a family of actuarial instruments (the Wisconsin risk/need instruments) in the Oklahoma Probation and Parole Department. Probation officers' attitudes toward these measures generally were negative or neutral, with fewer than half reporting that the instruments were helpful or appropriate for making decisions about level of supervision. More to the point, a substantial majority of the sample believed that probation officers should have more discretion in selecting the level of supervision (76%) and that the professional experience and knowledge of the probation officer was superior to what the instrument could offer (61%). Interestingly, in the present study, a similar theme—that JPOs' experience and knowledge were superior to any benefit offered by the instrument—was apparent only among users of the score-based YLS/CMI, and not among users of the SAVRY. However, this may be explained by the fact that YLS/CMI users had substantially more years of experience as JPOs compared with SAVRY users.

Administrators in future implementation trials may choose to alleviate similar fears by putting into place procedural safeguards. For example, activities to investigate the degree of inter-rater reliability amongst professionals in a given office could be undertaken to (hopefully) demonstrate that diverse users generate similar SRRs and rationales for those ratings. Such activities could be formal, such as a local field reliability investigation, or more informal, such as evaluations of performance on a booster training case delivered in a group setting. Agencies also could implement quality assurance guidelines to alleviate concerns over misuse or misapplication of the instrument (e.g., adjusting risk ratings to coincide with a more desirable supervision caseload).

Almost all SAVRY users reported at the first interview that making the SRR was 'very easy' or 'somewhat easy.' By the second interview, all JPOs held this view. These findings are consistent with those of Doyle, Lewis, and Brisbane (2008), who studied the implementation in a forensic mental health service in the U.K of another SPJ instrument, the Short-Term Assessment of Risk and Treatability (START; Webster et al., 2009). Doyle and colleagues reported that most staff (74.4%) indicated not having any difficulties in relation to the instrument's Risk Formulation section. Also studying implementation of the

START, Crocker and colleagues (2011) reported that among front line users of the instrument in a civil psychiatric hospital unit, 79% reported experiencing no difficulties in completing the Risk Formulation section, whereas 14% reported some difficulties (e.g., need for clarification around the concept of "risk"). Thus, it appears that clinicians and probation officers trained to use the SPJ model find the experience of applying SPJ instruments to be relatively straightforward. Whether strength of self-efficacy is associated with predictive accuracy, however, is unknown.

Implications for Training on Risk Assessment

Formal training is an important avenue by which one can develop skills for assessing risk for violence and other types of delinquent behavior. Research shows that standardized training is associated with skills acquisition. For example, using a pre-post design, Storey, Gibas, Reeves, and Hart (2011) investigated the degree to which participation in an 8-day training on the SPJ model and various instruments affected criminal justice professionals' knowledge about violence risk assessment concepts, violence risk assessment skills, and attitudes regarding competence at assessing violence risk. There were significant improvements on measures of knowledge about risk assessment (30% increase in correct responses, Cohen's $d = 1.27$); in skills in the analysis of risk of violence (identification of risk factors, risk level, and management strategies) based on a practice vignette; and in perceived confidence in conducting violence risk assessments.

To date, little research has been conducted on the effects of training on practice in the 'real world,' or on best practices for professionals who deliver such training programs. McNiel et al. (2008) investigated the impact of a five-hour workshop delivered to psychiatry residents and psychology interns on evidence based violence risk assessment, focusing on Version 2 of the Historical, Clinical, Risk Management-20 (HCR-20; Webster, Douglas, Eaves, & Hart, 1997). Workshop participation improved trainees' ability to identify risk and protective factors for violence, increased knowledge about and ability to manage violence, and improved the quality of clinical documentation (e.g., regarding the presence of risk factors as well as analysis of the significance of the risk factors for risk-reducing interventions). Evaluation of an HCR-20 training workshop for clinicians in a forensic psychiatric hospital in the U.K. also found workshop participation was associated with significantly improved risk management plans (Reynolds & Miles, 2009).

Results of the present study offer guidance for areas of curriculum that should be emphasized during training on risk assessment with youth populations. Users of both the YLS/CMI and the SAVRY reported that some items were relatively more difficult to rate than others. Although at times this difficulty was rooted in a lack of available

information needed to rate the item, on many occasions it was based on unclear rating instructions in the manual. With respect to the YLS/CMI, for example, clarification regarding how to rate dichotomous items when users feel the case information is more nuanced appears warranted during training workshops. Regarding the SAVRY, the present results suggest that trainers should devote ample time for instruction on the Social/Contextual risk factors.

For SAVRY users experiencing difficulty making the SRR, it may be beneficial during initial and subsequent booster training workshops to focus on development of case formulation skills (see Hart, Sturmey, Logan, & McMurrin, 2011). With respect to training about some of the standard SPJ assessment practices, our results indicate the concept of manifestation is relatively easy for JPOs to understand. Our findings suggest that SAVRY workshops could be improved by including more explicit instruction regarding the need to consider the idiographic relevance of each risk factor for the particular youth being evaluated, as well as guidance about how to apply information about items' relevance when making the SRR. In brief, when using any SPJ instrument, evaluators rate the presence of numerous (typically 20) nomothetically derived risk factors and then subsequently consider the individual relevance of each item for the given youth being evaluated. Although this step is a fundamental component of the SPJ model, not all SPJ instrument manuals offer explicit instruction on this step, including the SAVRY. Research using the HCR-20 demonstrated that relevance ratings were slightly more strongly associated with physical violence ($AUC = .81$) than were presence ratings ($AUC = .72$) in a postdictive study of 43 individuals in civil psychiatric and correctional settings (Blanchard & Douglas, 2011).

Limitations and Future Directions

Our results should be interpreted in light of a few limitations. First, the sample size of 71 was relatively small (although it was representative in that every JPO in each of the six offices was interviewed at least once as a part of the risk assessment implementation process). Second, the study did not measure pre-existing differences in the cultures of each site that may have impacted openness to use of a risk assessment instrument. Because each office in this study was part of a national juvenile justice reform initiative, it was assumed that all sites were at a fairly high stage of readiness for change prior to implementation of the instruments. There was buy-in from the majority of JPOs for implementation of a risk assessment instrument, which is a prerequisite for successful implementation and sustainability. However, there were some between-state differences in JPOs' education level and experience, as well as differences in philosophies with respect to use of a risk assessment instrument in disposition or placement decisions (see Vincent, Paiva-Salisbury, et al., 2012). Even though the

same general implementation and training procedures were followed in both states, state or site-level differences may have affected JPOs' perceptions and application of the risk assessment measure. Given these methodological limitations, findings regarding YLS/CMI and SAVRY users' perceived advantages or disadvantages of the instrument should not be attributed as being related to the particular instrument per se.

Results from any implementation study will depend, at least in part, on how the agency or system was operating prior to the introduction of the change being investigated (here, the use of the SAVRY or YLS/CMI and accompanying policies for practice). For example, in the probation offices studied, comprehensive information gathering about the youth and his or her family was not routine procedure before the risk assessment instruments were introduced. Had it been, then the criticism regarding the instruments adding significantly to the amount of time needed to complete case planning activities might have occurred less often. Also, how the agency was operating prior to being put into practice will affect the generalizability of results from implementation studies. Given that neither state had been using a structured process for risk assessment prior to using the SAVRY or YLS/CMI, our results may not generalize to probation offices where some form of routine procedure already was being followed. Furthermore, whether JPOs using instruments based on a particular model would be relatively more or less accepting of a new instrument that used a different model for decision-making (e.g., switching from using an actuarial to an SPJ instrument) cannot be addressed by the present study.

Another limitation is that data comprised JPOs' self-report regarding how they thought they conduct their assessments. Research suggests decision-makers and lay people generally have little awareness about the factors that influence their judgment (Arkes, 1981; Aspel, Willis, & Faust, 1998; Gauron & Dickinson, 1966). Thus, when JPOs reported that they mainly used the YLS/CMI to back-up conclusions they would have come to anyway, or reported that they considered specific SAVRY items for all cases when making their SRRs, it does not necessarily mean that these scenarios are what actually happen in practice. A research design helpful for exploring this issue would be one in which JPOs rated the risk assessment instrument and provided their case recommendations for a standardized case vignette.

Another important aspect of evaluating the effectiveness of a risk assessment instrument is how user-friendly it is for use in case management activities or when making recommendations about disposition. Surprisingly, despite the additional guidance of the YLS/CMI for identifying need areas (in the form of scales with explicit labels), there appeared to be little difference in JPOs' perceived helpfulness of these instruments for making recommendations about disposition, service referrals, or supervision level,

albeit this was not the main focus of the study. Developing a case plan with a high likelihood of reducing a youth's risk requires JPOs to identify and then triage the youth's most important dynamic risk factors (also known as criminogenic needs) to be targeted for intervention. SAVRY users mentioned a benefit of the instrument was its ability to identify dynamic risk and protective factors; therefore, the instrument ostensibly should lead to quality case planning. How well disposition recommendations and case plans align with youths' criminogenic needs and whether the alignment differs between the YLS/CMI and the SAVRY are areas in need of further study.

In most research studies using the YLS/CMI (including the present one for all intents and purposes), users tend to rate only Parts I (item scores) and II (scale scores and total score); data on Parts III (assessment of other needs and special considerations) and IV (use of the professional override) tend to be absent in the literature. It is strongly recommended that future research work towards filling this gap. In particular, the impact on predictive validity of considering the case specific risk and responsivity factors listed in Part III and on using a professional override in Part IV should be explored (e.g., see Vaswani & Merone, 2013).

In sum, our findings indicate that JPOs perceived both the YLS/CMI and SAVRY as being 'somewhat' or 'very' helpful for guiding case planning decisions. Most SAVRY users demonstrated good understanding of key components of the SPJ model. Both measures generally were experienced as being easy to use, including the aspect of the SAVRY requiring the most discretion - developing the overall summary judgment about the youth's risk level. The majority of SAVRY users expressed preference for risk assessment procedures that called for professional discretion rather than a score-based approach. Finally, our results suggest that many anticipated barriers to putting a risk assessment instrument into routine practice can be overcome when sound implementation procedures are followed, allowed frontline users and agencies more generally to benefit from research based developments in the science of risk assessment.

REFERENCES

- Allen, N. E., Watt, K. A., & Hess, J. Z. (2008). A qualitative study of the activities and outcomes of domestic violence coordinating councils. *American Journal of Community Psychology, 41*, 63-73. doi:10.1007/s10464-007-9149-5
- Andrews, D. A. & Bonta, J. (1995). *LSI-R: The Level of Service Inventory-Revised*. Toronto, ON, Canada: Multi-Health Systems, Inc.
- Arkes, H. R. (1981). Impediments to accurate clinical judgment and possible ways to minimize their impact. *Journal of Consulting and Clinical Psychology, 49*, 323-330. doi:10.1037/0022-006X.49.3.323
- Aspel, A. D., Willis, W. G., & Faust, D. (1998). School psychologists' diagnostic decision-making processes: Objective-subjective discrepancies. *Journal of School Psychology, 36*, 137-149. doi:10.1016/S0022-4405(98)00002-8

- Baird, C. (February 2009). *A question of evidence: A critique of risk assessment models used in the justice system*. Special Report. National Council on Crime and Delinquency.
- Berg, B. L. (2004). *Qualitative research methods for the social sciences* (5th ed.). Boston, MA: Pearson.
- Bilchik, S. (1999). *Juvenile Justice: A Century of Change* (Report). Washington, DC: Office of Juvenile and Delinquency Prevention. Available online at: <http://www.ncjrs.gov/pdffiles1/ojdp/178995.pdf>.
- Blanchard, A., & Douglas, K. S. (2011, March). *APLS Undergraduate Paper Award (First Place): The Historical-Clinical-Risk Management-Version 3: The inclusion of idiographic relevance ratings in violence risk assessment*. Invited poster presented at the annual convention of the American-Psychology Law Society, Miami, FL.
- Bornstein, R. F. (1989). Exposure and affect: overview and meta-analysis of research, 1968–1987. *Psychological Bulletin*, *106*, 265–289. doi: 10.1037/0033-2909.106.2.265
- Borum, R., Bartel, P., & Forth, A. (2006). *Structured Assessment of Violence Risk in Youth* (SAVRY). Odessa, FL: Psychological Assessment Resources, Inc.
- Crocker, A. G., Braithwaite, E., Laferrière, D., Gagnon, D., Venegas, C., & Jenkins, T. (2011). START changing practice: Implementing a risk assessment and management instrument in a civil psychiatric setting. *The International Journal of Forensic Mental Health*, *10*, 13–28. doi: 10.1080/14999013.2011.553146
- Desmarais, S. L., Sellers, B. G., Viljoen, J. L., Cruise, K. R., Nicholls, T. L., & Dvoskin, J. A. (2012). Pilot implementation and preliminary evaluation of START:AV assessments in secure juvenile correctional facilities. *The International Journal of Forensic Mental Health*, *11*(3), 150–164. doi:10.1080/14999013.2012.737405
- Desmarais, S. L., Van Dorn, R. A., Telford, R. P., Petrla, J., & Coffey, T. (2012). Characteristics of START assessments completed in mental health jail diversion programs. *Behavioral Sciences & the Law*, *30*, 448–469. doi: 10.1002/bsl.2022
- Douglas, K. S., & Kropp, P. R. (2002). A prevention-based paradigm for violence risk assessment: Clinical and research applications. *Criminal Justice and Behavior*, *29*(5), 617–658. doi: 10.1177/009385402236735
- Douglas, K. S., Hart, S. D., Groscup, J. L., & Litwack, T. R. (2014). Assessing violence risk. In I. B. Weiner & R. K. Otto (Eds.), *The Handbook of Forensic Psychology, 4th Edition* (pp. 385–441). New York, NY: Wiley.
- Douglas, K. S., Hart, S. D., Webster, C. D., & Belfrage, H. (2013). *HCR-20 V3: Assessing risk of violence – User guide*. Burnaby, Canada: Mental Health, Law, and Policy Institute, Simon Fraser University.
- Doyle, M., Lewis, G., & Brisbane, M. (2008). Implementing the Short-term Assessment of Risk and Treatability in a forensic mental health service. *Psychiatric Bulletin*, *32*, 406–408. doi:10.1192/pb.bp.108.019.794
- Farrell, J., Young, D. W., & Taxman, F. S. (2011). Effects of organizational factors on use of juvenile supervision practices. *Criminal Justice and Behavior*, *38*, 565v583. doi:10.1177/0093854811401786
- Fixsen, D. L., Naoom, S. F., Blase, K. A., Friedman, R. M., & Wallace, F. (2005). *Implementation research: A synthesis of the literature*. Tampa, FL: University of South Florida, Louis de la Parte Florida Mental Health Institute, The National Implementation Research Network (FMHI Publication #231).
- Flores, A. W., Lowenkamp, C. T., Holsinger, A. M., & Latessa, E. J. (2006). Predicting outcome with the Level of Service Inventory-Revised: The importance of implementation integrity. *Journal of Criminal Justice*, *34*, 523–529. doi:10.1016/j.jcrimjus.2006.09.007
- Gauro, E. F., & Dickinson, J. K. (1966). Diagnostic decision making in psychiatry. *Archives of General Psychiatry*, *14*, 225–232. doi:10.1001/archpsyc.1996.01730090001001
- Guy, L. S. (2008). *Performance indicators of the structured professional judgement approach for assessing risk for violence to others: A meta-analytic survey*. (Unpublished doctoral dissertation). Burnaby, BC, Canada: Simon Fraser University.
- Guy, L. S., Hart, S. D., & Douglas, K. S. (in press). Risk assessment and communication. In B. Cutler & P. Zapf (Eds.), *APA handbook of forensic psychology*. Washington, DC: American Psychological Association.
- 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.
- Hart, S. D., Kropp, P. R., Laws, D. R., Klaver, J., Logan, C., & Watt, K. A. (2003). *The risk for sexual violence protocol (RSVP): Structured professional guidelines for assessing risk of sexual violence*. Vancouver, BC, Canada: The Mental Health, Law, and Policy Institute, Simon Fraser University, Pacific Psychological Assessment Corporation, British Columbia Institute Against Family Violence.
- Hart, S. D., Sturmey, P., Logan, C., & McMurrin, M. (2011). Forensic case formulation. *The International Journal of Forensic Mental Health*, *10*, 118–126. doi:10.1080/14999013.2011.577137
- Hoge, R. D., & Andrews, D. A. (2002). *Youth Level of Service/Case Management Inventory: User's manual*, North Tonawanda, NY: Multi-Health Systems.
- Hoge, R. D., & Andrews, D. A. (2010). *Youth Level of Service/Case Management Inventory: User's manual, Version 2.0*. North Tonawanda, NY: Multi-Health Systems.
- Hoge, R. D., Vincent, G., & Guy, L. S. (2012). Prediction and risk/needs assessment. In R. Loeber & D. Farrington (Eds.), *From juvenile delinquency to adult crime: Criminal careers, justice policy, and prevention* (pp. 150–183). New York, NY: Oxford University Press.
- Juvenile Accountability Incentive Block Grant, P.L. 105–109.
- Juvenile Justice Delinquency and Prevention Act (JJDP). (2002). Public Law 107–273, 42.
- Kroppan, E., Nasset, M. B., Nonstad, K., Pedersen, T. W., Almvik, R., Palmstierna, T. (2011). Implementation of the Short Term Assessment of Risk and Treatability (START) in a forensic high secure unit. *International Journal of Forensic Mental Health*, *10*, 7–12. doi:10.1080/14999013.2011.552368
- Lipsey, M. W. (2009). The primary factors that characterize effective interventions with juvenile offenders: A meta-analytic overview. *Victim and Offenders*, *4*, 124–147. doi:10.1080/15564880802612573
- Lipsey, M. W., & Wilson, D. B. (1998). Effective intervention for serious juvenile offenders: A synthesis of research. In R. Loeber & D. P. Farrington (Eds.), *Serious & violent juvenile offenders: Risk factors and successful interventions* (p. 507). Thousand Oaks, CA: Sage Publications, Inc.
- Livsey, S. (2012). Juvenile delinquency probation caseload, 2009. OJJDP Factsheet (NCJ 239082). Washington, DC: Office of Juvenile Justice and Delinquency Prevention. Retrieved from http://www.ncjj.org/pdf/Livsey_JuvDelinquencyProbation09.pdf
- McNiel, D. E., Chamberlain, J. R., Weaver, C. M., Hall, S. E., Fordwood, S. R., & Binder, R. L. (2008). Impact of clinical training on violence risk assessment. *American Journal of Psychiatry*, *165*, 195–200. doi:10.1176/appi.ajp.2007.06081396
- Meehl, P. E. (1954). *Clinical versus statistical prediction: A theoretical analysis and a review of the evidence*. Minneapolis, MN: University of Minnesota Press.
- Mulvey, E. P., & Iselin, A. R. (2008). Improving professional judgments of risk and amenability in juvenile justice. *The Future of Children*, *18*, 35–57. doi: 10.1353/foc.0.0012
- Nicholls, T., Petersen, K. L., Brink, J., & Webster, C. (2011). A clinical and risk profile of forensic psychiatric patients: Treatment team STARTs in a Canadian service. *The International Journal of Forensic Mental Health*, *10*(3), 187–199. doi:10.1080/14999013.2011.600234
- Nonstad, K., & Webster, C. D. (2011). How to fail in the implementation of a risk assessment scheme or any other new procedure in your organization. *American Journal of Orthopsychiatry*, *81*, 94–99. doi:10.1111/j.1939-0025.2010.01076.x
- Olver, M., Stockdale, K., & Wormith, J. (2009). Risk assessment with young offenders: A meta-analysis of three assessment measures. *Criminal Justice and Behavior*, *36*, 329–353. doi:10.1177/0093854809331457

- Pope, C., Ziebland, S., & Mays, N. (2000). Qualitative research in health care: Analyzing qualitative data. *British Medical Journal*, *320*, 114–116.
- Powell, B. J., McMillen, J., Proctor, E. K., Carpenter, C. R., Griffey, R. T., Bunger, A. C., . . . , & York, J. L. (2012). A compilation of strategies for implementing clinical innovations in health and mental health. *Medical Care Research And Review*, *69*(2), 123–157. doi:10.1177/1077558711430690
- Proctor, E. K., Landsverk, J., Aarons, G., Chambers, D., Glisson, C., & Mittman, B. (2009). Implementation research in mental health services: An emerging science with conceptual, methodological, and training challenges. *Administration And Policy In Mental Health And Mental Health Services Research*, *36*(1), 24–34. doi:10.1007/s10488-008-0197-4
- Proctor, E. K., & Rosen, A. (2008). From knowledge production to implementation: Research challenges and imperatives. *Research on Social Work Practice*, *18*(4), 285–291. doi:10.1177/1049731507302263
- Proctor, E., Silmere, H., Raghavan, R., Hovmand, P., Aarons, G., Bunger, A., . . . , & Hensley, M. (2011). Outcomes for implementation research: Conceptual distinctions, measurement challenges, and research agenda. *Administration and Policy in Mental Health and Mental Health Services Research*, *38*(2), 65–76. doi:10.1007/s10488-010-0319-7
- Reynolds, K., & Miles, H. L. (2009). The effect of training on the quality of HCR-20 violence risk assessments in forensic secure services. *The Journal of Forensic Psychiatry & Psychology*, *20*, 473–480. doi:10.1080/14789940802638366
- Savaya, R., Monnickendam, M., & Waysman, M. (2000). An assessment of the utilization of a Computerized Decision Support System for youth probation officers. *Journal of Technology in Human Services*, *17*, 1–14. doi:10.1300/J017v17n04_01
- Schneider, A. L., Ervin, L., & Snyder-Joy, Z. (1996). Further exploration of the flight from discretion: The role of risk/needs instruments in probation supervision decisions. *Journal of Criminal Justice*, *24*, 109–121. doi:10.1016/0047-2352(95)00059-3
- 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
- Storey, J. E., Gibas, A. L., Reeves, K. A., & Hart, S. D. (2011). Evaluation of a violence risk (threat) assessment program for police and other criminal justice professionals. *Criminal Justice and Behavior*, *38*, 554–564. doi:10.1177/0093854811403123
- Towberman, D. B. (1992). A national survey of juvenile risk assessment. *Juvenile & Family Court Journal*, *43*, 61–67. doi:10.1111/j.1755-6988.1992.tb00720.x
- Vaswani, N., & Merone, L. (2013). Are there risks with risk assessment? A study of the predictive accuracy of the Youth Level of Service–Case Management Inventory with young offenders in Scotland. *British Journal of Social Work*, 1–19. doi:10.1093/bjsw/bct059
- Vieira, T. A., Skilling, T. A., & Peterson-Badali, M. (2009). Matching court-ordered services with treatment needs. *Criminal Justice and Behavior*, *36*, 385–401. doi:10.1177/0093854808331249
- Vincent, G. M., Guy, L. S., & Grisso, T. (2012). *Risk assessment in juvenile justice: A guidebook for implementation*. John D. & Catherine T. MacArthur Foundation. Available online at: <http://www.nysap.us/Risk%20Guidebook.pdf>
- Vincent, G. M., Guy, L. S., Fusco, S. L., & Gershenson, B. G. (2012). Field reliability of the SAVRY with probation officers: Implications for training. *Law and Human Behavior*, *36*, 225–236. doi:10.1007/s10979-011-9284-2
- Vincent, G. M., Guy, L. S., & Gershenson, B. G., & McCabe, P. (2012). Does risk assessment make a difference? Results of implementing the SAVRY in juvenile probation. *Behavioral Sciences and the Law*, *30*, 384–405. doi: 10.1002/bsl.2014
- Vincent, G. M., Paiva-Salisbury, M., Cook, N. E., Guy, L. S., & Perrault, R. T. (2012). Impact of risk/needs assessment on juvenile probation officers' decision-making: Importance of implementation. *Psychology, Public Policy, and Law*, *18*, 549–576. doi:10.1037/a0027186
- Vincent, G. M., Terry, A., & Maney, S. (2009). Risk/needs tools for antisocial behavior and violence among youthful populations. In J. Andrade (Ed.), *Handbook of violence risk assessment and treatment for forensic mental health practitioners* (pp. 337–424). New York, NY: Springer.
- Webster, C. D., Douglas, K., Eaves, D. & Hart, S. (1997). *HCR-20: Assessing risk for violence, Version 2*. Burnaby, BC, Canada: Simon Fraser University.
- Webster, C. D., Martin, M., Brink, J., Nicholls, T. L., & Desmarais S. L. (2009). *Manual for the Short-Term Assessment of Risk and Treatability (START) (Version 1.1)*. Coquitlam, Canada: Forensic Psychiatric Services Commission.
- Zajonc, R. B. (1968). Attitudinal effects of mere exposure. *Journal of Personality and Social Psychology*, *9*, 1–27. doi:10.1037/h0025848. ISSN 1939-1315.