A Review of Brief Risk Assessment Tools Validated for use in Correctional Settings

July 2015

THE UNIVERSITY OF UTAH

Utah Criminal Justice Center

COLLEGE OF SOCIAL WORK
COLLEGE OF SOCIAL & BEHAVIORAL SCIENCES
UTAH COMMISSION ON CRIMINAL AND JUVENILE JUSTICE
S.J. QUINNEY COLLEGE OF LAW

A Review of Brief Risk Assessment Tools Validated for use in Correctional Settings

Christian M. Sarver, M.S., M.S.W.
Kort C. Prince, Ph.D.
Jessica Seawright, M.S.W.
Robert P. Butters, Ph.D.

July 2015

Utah Criminal Justice Center, University of Utah

{THIS PAGE INTENTIONALLY LEFT BLANK}

Table of Contents

{THIS PAGE INTENTIONALLY LEFT BLANK}

Overview of Risk Assessment in Correctional Settings

Criminal justice researchers, government entities, and professional organizations endorse the use of structured risk assessments as best practice in correctional settings (Austin, 2004; Latessa & Lovins, 2010; Summers & Willis, 2010). When decisions regarding the placement, supervision, and treatment of offenders are informed by risk assessment, criminal justice systems are better able to address organizational goals related to public safety and efficiency, theoretically because offenders receive the specific type and amount of supervision and services necessary to reduce their risk of recidivism (Andrews, Bonta, & Hoge, 1990). Jurisdictions that adopt risk assessment systems have demonstrated reduced correctional costs in the form of fewer incarcerated individuals and shorter jail and prison stays (Lowenkamp, Pealer, Smith, & Latessa, 2006; Ostrom, Kleiman, Cheesman, Hansen, & Kauder, 2002). Despite such benefits, the economic reality of conducting an assessment with every individual who enters the criminal justice system can be prohibitive (Flores, Lowenkamp, Holsinger, & Latessa, 2006). Given the costs of assessment—including those related to staff time, training, and certification—many criminal justice entities lack sufficient resources to conduct in-depth assessments of the 12 million persons arrested in the U.S. each year (Federal Bureau of Investigation (FBI), 2014).

Risk screening instruments—brief questionnaires that can be administered by non-clinical staff—provide a less resource-intensive method for classifying offenders into groups according to risk of recidivism (Judicial Council of California, Administrative Office of the Courts, 2011; Latessa & Lovins, 2010). The primary purpose of risk screening is to inform decisions regarding who should receive a full risk/need assessment; as such, criminal justice professional organizations recommend that all offenders undergo basic screening at intake. Those classified as low-risk for recidivism should be considered for low level responses, without further assessment. In contrast, offenders classified as medium- to high-risk for reoffending should undergo further assessment in order to make individualized decisions regarding their program placement, supervision levels, and treatment needs (Christensen, Jannetta, & Buck-Willison, 2012). Brief screening procedures allow criminal justice entities to direct "scarce resources toward those who are most likely to benefit from a comprehensive assessment" (Wei & Parsons, 2012, p. 3).

As part of the risk assessment process, screening tools used in criminal justice settings should comply with the principles of effective correctional intervention, which are discussed briefly in the following section.

The Principles of Effective Correctional Intervention

The Risk-Need-Responsivity (RNR) model has been described as a map for offender rehabilitation that assists in the identification of *who* to provide services to (risk principle), *what* services to provide (need principle) and *how* to create a learning environment that will facilitate behavioral change (responsivity principle). In the past 20 years, research has demonstrated that criminal justice programs are more effective when operated in accordance with the RNR principles (Andrews & Bonta, 2010; Andrews & Dowden, 2007a;

Gendreau, Smith, & French, 2006). For example, Andrews and Bonta (2006) analyzed 44 vocational programs and found those adhering to all three principles demonstrated a 38 percentage point difference in recidivism rates between the treatment and comparison groups, while those adhering to none of the principles produced a five percentage point difference in recidivism. The three principles forming the basis of effective correctional interventions are discussed below (Andrews et al., 1990; Andrews & Bonta, 2010):

> **Risk principle:** The risk principle says that offenders' likelihood of committing future crimes can be predicted through risk assessments. As such, decisions about resource allocation (e.g., programming and treatment) should be based on assessed risk, with high-risk offenders receiving more intensive services. The implications of the risk principle are twofold: 1) correctional programs must provide services to high-risk offenders in sufficient dosage to reduce risk; and, 2) providing too many services to low-risk offenders can actually raise their risk of future crime (Lowenkamp & Latessa, 2004).

> **Need principle:** The need principle identifies eight factors—referred to as criminogenic needs—that are empirically linked with criminal behavior <u>and</u> are also amenable to intervention (e.g. risk factors that can be changed, like attitudes, as opposed to those that are not amenable to change, like age at first arrest). Correctional treatment should be based on assessed need and specifically target criminogenic needs: history of antisocial behavior; antisocial personality pattern; pro-criminal attitudes; social supports for crime; substance abuse; improved family relationships; education and employment deficits; and pro-social recreational opportunities.

> **Responsivity principle:** The responsivity principle identifies cognitive behavioral programming as the most effective way to teach new skills and behaviors. Additionally, the responsivity principle proposes that interventions should be tailored to offenders' individual characteristics, such as gender, cognitive abilities, or motivation to change.

The RNR model has been incorporated into best practice recommendations by the majority of U.S. criminal justice organizations. In the past ten years, The Center for Effective Public Policy (CEPP), The Council of State Governments Justice Center (CSG), The National Center for State Courts (NCSC), The National Institute of Corrections (NIC), The National Institute of Justice (NIJ), The National Parole Resource Center (NPRC), The Reentry Policy Council (RPC), and The Urban Institute (UI) have published best practice guidelines based on the RNR model.

Issues in Screening and Risk Assessment in Correctional Settings

Broadly speaking, risk assessment is a process for classifying offenders into groups based on their probability of future criminal behavior (Hamilton, Neuilly, Lee, & Barnoski, 2014), essentially applying "group statistics to individual decisions" (Baird, 2009, p. 3). Given the impossibility of identifying exactly which individuals will commit new crimes, assessment

instruments should be regarded as "advisory rather than peremptory" (Latessa & Lovins, 2010). Issues related to the evaluation of an instrument's ability to accurately classify offenders are discussed in the following section. In addition to accuracy, the utility of an instrument in a given setting depends upon a range of other factors, including intended purpose, ease of use, and consistency.

Assessment type. Evolving methods for collecting and interpreting information during assessment are often referred to as "generations" in criminal justice literature. The first generation of assessment, referred to as structured clinical judgment, was based on a practitioners' clinical expertise. This approach has less predictive validity when compared to statistical methods (Andrews & Bonta, 2010; Hilton, Harris, & Rice, 2006). Second generation assessments consist of instruments that make risk determinations based on statistical analyses. Second generation instruments classify offenders almost entirely on the basis of static (historical) risk factors, which means they have limited utility in the development of treatment plans. Third-generation instruments identify and quantify offenders' criminogenic needs and produce a list of treatment targets. Unlike second-generation tools, they are grounded in the assumption that behavior is malleable and that offenders can be reformed through treatment and environmental alterations. Third-generation instruments are empirically-based and theory-driven. Fourth-generation tools expanded upon third-generation instruments by systematically linking the results to treatment plans and case management strategies (Andrews & Bonta, 2006).

Typically, brief screening instruments are second generation, actuarial tools that can be completed, interpreted, and scored with relative ease and efficiency.

Validity and reliability. The accuracy of assessment instruments, including screening tools, is often described in terms of predictive validity, which means statistical testing confirms that the instrument accurately predicts intended outcomes for offender populations. Predictive validity may change across dimensions of race, class, gender, and geography and instruments should therefore be tested for the particular population with which they will be used (Christensen et al., 2012). In addition, screening and assessment tools should be validated within local jurisdictions, to confirm that offenders who are scored as high risk are actually recidivating at higher rates than individuals who score as low risk (Latessa & Lovins, 2010). A second dimension of an instruments' accuracy is referred to as reliability, which means the instrument produces similar results when administered by different individuals. Screening and assessment instruments are most relevant when they have been evaluated for reliability and validity in "real-world" settings, which means reliability and predictive validity have been tested in non-research settings, where the assessment is administered by correctional staff rather than researchers (Vincent, Guy, & Grisso, 2012).

Administering and scoring risk assessments. The ease with which a screening instrument can be completed depends on a number of factors, including: the number of questions; accessibility of the information needed to answer the questions; and the training and skill required to interpret responses. For screening purposes, instruments generally solicit information that is widely available (such as that found in official records) and does

not require an advanced degree to collect or interpret. Most commonly, that information is collected through a review of criminal history records and a structured interview, although the latter can be labor-intensive. Self-report instruments are cost-effective in terms of staff time: they do not require an interview and can be administered to multiple offenders at once (Campbell, French, & Gendreau, 2007). Self-report instruments often suffer from a lack of "face validity" with stakeholders, however, who may be skeptical about the reliability of information obtained directly from offenders (Austin, 2004).

Actuarial instruments characterize risk based on mathematical calculations: items are summed (and sometimes weighted according to their relative association with recidivism) to produce a total risk score. Actuarial instruments typically produce a risk classification as well: that is, an offender whose score falls within a certain range is classified as low, medium, or high risk for recidivism. Individual items may also contribute to larger domain scores, or sub-scales, that provide specific information such as risk for pretrial failure or violent recidivism.

In contrast to actuarial assessments, risk may also be characterized according to structured professional judgment (SJP). With this type of assessment, the relative importance of each item is determined by the assessor (even if the individual items are scored) who makes the ultimate decision regarding an individual's risk classification. Because the process for classifying offender risk is more nuanced than actuarial assessments (and may therefore require additional training or an advanced degree on the part of the assessor), relatively fewer SJP instruments are used as brief screening tools.

Items assessed. Screening and assessment instruments may be comprised of static items (not amenable to change, such as age at first offense) and/or dynamic items (amenable to change, such as pro-social associates). Included items may include both risk factors (items whose presence indicates an increased risk for criminal behavior, such as prior criminal history) and/or protective factors (items whose presence is associated with reduced risk of re-offending, such as being married). While each of the aforementioned types of data may contribute to an instrument's predictive validity, many of the static risk items associated with reoffending are available in official criminal records. In contrast, information related to dynamic needs and protective factors may require collateral sources of information and clinical diagnoses. Because of the relative ease of collecting static risk factors, when compared to dynamic and protective factors, many brief risk instruments are comprised solely of static factors. When compared to dynamic risk factors, static factors are often objective rather than subjective in nature. The comparative objectivity of the items on an assessment can be associated with increased reliability (Steinhart, 2006): it is more likely that three different assessors will agree on a question asking whether or not an offender has a prior felony arrest than whether or not he has an ongoing substance abuse problem. The use of instruments that rely on static and objective items is of particular relevance for brief risk screening; such tools require fewer resources to implement due to "the isolation of methodological performance differences, eliminating issues related to staff training, fidelity, and interrater reliability" (Hamilton et al., 2014, p. 9).

Assessment and decision-making. Risk assessments have been created for use at many stages of the criminal justice process, including decisions related to pretrial release, diversion, sentencing, prison-based programming, and community supervision. While some instruments are designed to be used at multiple decision-points, differences in the type of information available (for example, between sentenced and non-sentenced offenders) as well as the relevant outcomes (for example, new crimes, technical violations, or failure to appear at court proceedings) may limit an assessment's utility across settings (National Judicial College (NJC), 2012). In response to this problem, some jurisdictions have developed statewide assessment systems—comprised of multiple, validated tools— that can be used at various points in the criminal justice system (Ohio Department of Rehabilitation and Correction, 2015). Regardless of the decision to which an instrument is applied, its utility is increased when risk levels are associated with specific response options, such as supervision level (Steinhart, 2006)

The Current Study

At the behest of the Utah Governor's Office, local and regional criminal justice stakeholders are currently involved in an initiative to ensure the state is "achieving the best public safety return on corrections spending" (Utah Commission on Crime and Juvenile Justice (CCJJ), 2014, p. 3). As part of this criminal justice reinvestment initiative, CCJJ has asked the Utah Criminal Justice Center (UCJC) to review and evaluate the utility of brief instruments for risk screening in criminal justice settings. In particular, this project seeks to identify instruments that can be used to inform decisions regarding the allocation of correctional resources, with increased attention to both cost and public safety. The following report will evaluate and compare brief instruments for predicting general and violent recidivism (excluding sexual violence), failure to appear, and technical violations among adult offenders and juvenile delinquents who have been arrested, detained, charged, or convicted of an offense. A final section will examine instruments intended to predict domestic violence recidivism.

Methods

Instruments were broadly identified as appropriate for use in risk screening if they could be easily and quickly administered and interpreted. While "ease" and "quickness" are relative terms, the study authors sought instruments that met the following criteria:

- The instrument was comprised of less than 15 questions. This was not a hard-and-fast rule, as instruments that exceeded this number were also included, typically because they were identified as a screening version of a longer assessment. Additionally, longer instruments were sometimes included if the items were collected through offender self-report, which places less administrative burden on criminal justice staff.

- The instrument was comprised of questions that relied on one or two information sources that were commonly accessible to criminal justice personnel.

- The instrument could be scored and interpreted by staff from a wide range of educational backgrounds.

- The instrument authors identified the tool as a brief instrument and/or identified that the instrument took less than 15 minutes to complete.

- Instruments were excluded if they relied on information gathered by staff with a clinical background (such as a diagnosis). Instruments were also excluded if they required a score from another, longer tool.

Search Criteria

The following study was not intended to be a systematic review; however, methods for identifying instruments are described here to provide transparency with respect to the search process. Instruments were identified through searches of the following portals: EBSCO was used to search *CINAHL, Criminal Justice Abstracts, ERIC, Family & Society Studies Worldwide, MEDLINE,* and *PsychINFO*; Social Services Abstracts was used to search *Applied Social Sciences Index and Abstracts, Social Service Abstracts,* and *Sociological Abstracts.* Researchers also searched the National Criminal Justice Reference Service (NCJRS) and Google. In its initial stage, the search strategy was comprised of the following terms:

(risk AND (screen* OR instrument OR tool))

AND

(recid* OR rearrest OR reincarcerat* OR reconvict* OR revoke OR revoc* OR fail*)

AND

(predict* OR valid* OR reliabil*)

A similar search strategy was developed to identify instruments used in the juvenile justice system, with the addition of terms more relevant to that system (e.g., youth, delinquent, delinquency, detention). Finally, the reference lists of systematic reviews, meta-analyses, technical reports, and primary studies generated in the aforementioned search were consulted to identify additional instruments.

Instruments were exclude from the current study if they: 1) did not meet the criteria for "brief screening instrument" described above; 2) were not developed for use in adult correctional or juvenile justice settings; 3) had not been assessed for predictive validity with respect to some recidivism outcome (e.g., re-arrest, re-incarceration, re-offending, failure to appear, or technical violation). Brief instruments that had only been evaluated for predictive validity with respect to other outcomes of interest in a criminal justice setting (such as brief mental health or substance abuse screens) were excluded from the current study, even if developed for use in correctional settings.

After developing a list of eligible instruments, an additional search—using the same databases—was conducted for each instrument, using the following strategy:

(Full instrument name OR instrument acronym)

AND

(predict* OR valid* OR reliabil*)

The final list of instruments includes tools designed and evaluated for assessing pretrial and convicted adult offenders and juvenile delinquents. Relevant outcomes include: general recidivism, non-sexual violent recidivism, intimate partner violence recidivism, technical violations, and failure to appear for court-ordered proceedings. Studies were coded according to the items described in Table 1.

Table 1: Instrument Coding

Item	Description
Name	Full name of instrument & acronym
Related	Related instruments, including revised and modified versions, and the name of the full version for screening instruments
Intended population	Profile of population the instrument was developed to assess (e.g., adult male offenders)
Validated population(s)	If relevant, other populations the instrument has been validated with
Intended outcomes	Outcomes the instrument was designed to predict (e.g., new arrest for any crime, new arrest for domestic violence)
Validated outcome(s)	Outcomes the instrument accurately predicts, including performance indicators of predictive validity for each outcome
Setting	Criminal justice point of contact where the instrument is intended to be administered (e.g., at arrest, after conviction, prior to community release)
Decision(s)	What decisions was the instrument designed to inform (e.g., placement type, supervision level, disposition)
Assessor	Who can administer the instrument, including required training or certification
Source	What information sources are required to complete the instrument (e.g., defendant interview, administrative data, self-report)
Inter-rater Reliability	For instruments scored by others (not self-report), what is the degree of agreement among raters
Time	How long does it take to complete the instrument
Items	How many questions is the instrument comprised of
Range	What is the possible range of scores
Scoring	How is the instrument scored (e.g., by hand, simple addition; by hand, items are weighted; computer algorithm)
Risk Categories	How many risk categories does the instrument use to classify offenders
Static Risk Factors	What static risk factors are included on the instrument
Dynamic Risk Factors	Does the instrument include dynamic risk factors (criminogenic needs) and do they contribute to risk score and classification
Protective Factors	Does the instrument include protective factors (and do they contribute to risk score and classification)
k	How many studies have been conducted that evaluate the predictive validity or reliability of the instrument
Location	In what country was the instrument developed and where has it been validated

Evaluation Criteria

Juvenile and adult instruments reviewed in this report were evaluated for predictive validity and inter-rater reliability. Because researchers reported results using a variety of statistics, it was necessary to set evaluative thresholds to make the outcomes comparable across different metrics. Tables 2 and 3 (discussed more below) provide the criteria for establishing evaluative comparability between the different predictive validity and inter-rater reliability metrics.

Predictive validity. Predictive validity of instruments reviewed in this report was assessed using the five most commonly reported effect size metrics[1]: Cohen's d, correlation coefficients, odds ratios, area under the curve (AUC), and Somer's d/mean cost rating (hereafter referred to only as MCR[2]). Of these, Cohen's d, AUC, and MCR are considered relatively base rate insensitive[3]. The base rate in the case of recidivism is the percentage of individuals who recidivate (e.g., any new arrest, any new conviction, new domestic violence charge, failure to appear (FTA)) during the follow up period. Base rates can differ as a function of the population of interest, the outcome of interest, and the time interval over which individuals were followed. For example, studies examining FTA as an outcome, given the short interval that characterizes pretrial, will have lower base rates than general recidivism studies that often follow offenders for three years or more.

The property of being base rate insensitive makes Cohen's d, AUC, and MCR relatively more ideal for comparing predictive validity effect sizes across studies and assessments. The correlation coefficient, in contrast, is relatively more problematic for comparison purposes (though valid and commonly reported). The correlation coefficient is extremely sensitive to changes in base rates, but adjustments can be made to interpretation of correlational effect sizes as base rates depart from 50% (the assumed base rate at which Cohen's original effect size criteria were published[4]; Cohen, 1988; Hunter & Schmidt, 2004). These adjustments alter the thresholds for what are considered small, medium and large effects; essentially, when base rates depart substantially from 50%, smaller correlations are required to meet these criteria[5].

---

[1] These are more properly termed "indices of detection performance" in this context, but the more general term "predictive validity effect size" is used throughout this document, as it is no doubt a term with which most readers are more familiar.

[2] MCR is a special case of Somer's d (Greene, Hoffman, & Beck, 1994)

[3] Due to making relatively fewer assumptions about the nature of the data, MCR and AUC are generally considered the modern standards for determining predictive accuracy (see Rice & Harris, 2005; Greene, Hoffman, & Beck, 1994); indeed, with the relatively recent increase in accessibility to software that can report these outcomes, MCR and AUC have become more prevalent in the recidivism prediction literature than the once dominant correlation coefficient.

[4] The reader may be familiar with the values of .1, .3, and .5 corresponding to small, medium, and large correlational effect sizes, respectively. These values refer to effect sizes for two continuous measures. When one variable is ordinal or continuous, and the other is dichotomous (as in the case of recidivism: yes or no), the correlation coefficient becomes the point-biserial correlation coefficient (r_{pb}), and the respective values for small, medium, and large effect sizes (assuming a 50% base rate) are .100, .243, and .371 (Cohen, 1988; Rice & Harris, 2005).

[5] There is debate in the literature regarding the use of base rate sensitive versus base rate insensitive measures, and the debate over which is best remains unresolved (largely because the answer depends on context). A full accounting

Odds ratios are also problematic for comparison purposes, but for a different reason[6]. While standard criteria for small, medium, and large odds ratio effect sizes exist (Chen, Cohen, & Chen, 2010), these criteria assume a two-by-two contingency table. In other words, they assume that two groups (e.g., high and low risk) are being compared on a binary outcome (e.g., recidivism: yes or no). When the number of discrete categories is increased (e.g., low, moderate, high, and intensive risk), odds ratios that are equally meaningful in effect become smaller simply as a function of an increase in categories. Across a varying number of categories, or across assessments with varying ranges for total scores, odds ratios are not comparable.

When odds ratios reflected a two-by-two contingency table, the numeric values in Table 2 were used to classify the effect sizes. To create comparability between the predictive validity effect sizes, when authors reported odds ratios for a greater than two-by-two contingency table (i.e., a 2 x k table), UCJC researchers used the reported frequencies in k to reconstruct each dataset[7], and computed AUC as the outcome. This methodology could not be used for one instrument/study that did not report the frequencies across k; for that study, the odds ratio significance ("yes"/"no") is listed instead of the general classification criteria described below.

In the comparisons in this report, predictive validity effect sizes are classified (in accord with Desmarais & Singh, 2013) as "poor," "fair," "good," and "excellent." While these criteria remain the same for Cohen's d, AUC, and MCR, different criteria for categorizing effect sizes as "poor," "fair," "good," and "excellent" were used to evaluate correlation coefficients as a function of the varying base rates of the outcomes. Criteria for the lowest and highest base rate observed in the study are provided in Table 2 as examples, and the full criteria are available for review in Table A-1 in Appendix A[8,9]. These criteria are provided for reference

---

of this debate, and the arguments for each side, is beyond the scope of this report, but the reader is encouraged to consult McGrath and Meyer (2006) for an excellent review. For the purposes of this report, the decision has been made to adjust effect size interpretations to make the base rate sensitive point-biserial correlation coefficient (r_{pb}) comparable to the base rate insensitive metrics rather than the other way around. This decision was made based on the prevailing methods in the field of criminal justice. While it is defensibly true that information contained in base rates have inherent meaning (Cohen, Cohen, West, & Aiken, 2003), the field of criminal justice has adopted base rate insensitive measures, with increasing reliance on AUC and MCR. This methodological approach is also defensible in circumstances where the base rate is subject to change across contexts, as is the case with jurisdictional differences in recidivism base rates, and with differing base rates created by varying researcher defined tracking periods.

[6] Similar to Cohen's d, AUC, and MCR, odds ratios are actually relatively base rate insensitive (McGrath & Meyer, 2006). Odds ratios become base rate sensitive only at extremely low or high base rates. After the base rate exceeds about 5%, little difference exists in the interpretation of odds ratios as they relate to Cohen's d (see also Borenstein, Hedges, Higgin, & Rothstein, 2009)

[7] Datasets were recreated using SPSS syntax. First, a data file with cases equal to the number of study-reported participants was created, followed by a variable that grouped them into their assigned risk level (i.e., low, moderate, high). Next, cases within groups were assigned as having recidivated or not based on the frequency of recidivism in each group reported in the original study. Finally, calculations for AUC were performed.

[8] Because this is not a meta-analytic review, correlation coefficient effect sizes are not adjusted for each study as a function of differing base rates; instead, studies were grouped so that base rates were more similar within comparisons. To facilitate these comparisons, studies were grouped by assessment type (whose outcomes differ in base rates), by outcome (e.g., arrest or conviction for general recidivism, arrest or new reported incident for domestic violence recidivism), and, within outcome, by follow time (e.g., recidivism follow periods of less than one

only; for ease of interpretation, translations of the effect size values in Tables 2 and A-1 have been performed by UCJC researchers. In the predictive validity tables that follow Table 2, numeric predictive validity effect size values (across all instruments and studies) are reported only in the user-friendly rating classification form (i.e., "poor," "fair," "good," and "excellent").

Table 2: Criteria for Determining Predictive Validity Effect Size Ratings

Effect Size Measure	Predictive Validity Rating				
	Poor	Fair	Good	Excellent	Significant
Cohen's d[a]	< .20	.20 – .49	.50 –.79	≥ .80	NA
Mean Cost Rating (MCR; Somer's d)[b]	< .10	.10 – .19	.20 –.29	≥ .30	NA
Area Under the Curve (AUC)[c]	<.56	.56 – .63	.64 –.71	≥ .72	NA
Correlation (r_{pb})[d]					
Lowest base rate (5%/95%)	<.04	.04 – .11	.12 – .17	≥ .18	NA
Highest base rate (50%/50%)	<.10	.10 – .24	.25 – .37	≥ .38	NA
Odds Ratio (2X2)[e]	(2Xk)	<1.44	1.45 – 2.48	2.49 – 4.27	≥ 4.28	Yes/No

[a] Values from Cohen (1988)
[b] Values adopted from Desmarais & Singh (2013)
[c] AUC values were anchored to Cohen's d effect sizes using formulas available in Ruscio (2008) and were checked against tabled values in Rice and Harris (2005).
[d] All base rate adjustments to correlation coefficients were anchored to Cohen's d; calculations were performed by hand using formulas available in McGrath and Meyer (2006).
[e] Values for a 2x2 contingency table were based on formulas from Borenstein, Hedges, Higgins, & Rothstein (2009).

Inter-rater reliability. Inter-rater reliability of the instruments reviewed in this report was evaluated using three common agreement metrics: kappa (K), intra-class correlation (ICC), and observer agreement. As with the predictive validity effect size metrics, intra-class correlations were classified as "poor," "fair," "good," and "excellent" according to the criteria in Table 3. The comparability across the different metrics was adopted from Desmarais and Singh (2013) and Cicchetti (2001).

Table 3: Criteria for Determining Inter-Rater Reliability (IRR) Ratings

IRR Rating	Reliability Measure		
	Kappa (K)	Intra-Class Correlation (ICC)	Observer Agreement (%)
Poor	.00 – .40	.00 – .40	<70.0
Fair	.40 – .59	.40 – .59	70.0 – 79.9
Good	.60 – .74	.60 – .74	80.0 – 89.9
Excellent	.75 – 1.00	.75 – 1.00	90.0 – 100.0
Notes: All values adopted from Desmarais & Singh (2013) and Cicchetti (2001, p. 697)			

---

year, one to three years, and more than three years). The base rates in Table A-1 represent the unweighted average of the base rates of the studies falling within the aforementioned groups. To the extent that base rates in any one study diverged greatly from the average, classifications may be slightly overinflated when the average or population base rate was smaller than the rate in the study, and underinflated when the average or population base rate was larger.
[9] Base rate insensitive effect sizes, such as Cohen's d, can be computed from the base rate sensitive point-biserial correlation and its associated base rates. Cohen's d can then be equated to AUC and MCR, as well as other metrics (F, eta-squared, odds ratios). While appropriate for a comprehensive meta-analysis, converting each study's individual effect size to a common language effect size (AUC) was beyond the scope of this report.

Classification rules. Because assessments often had more than one study addressing predictive validity or inter-rater reliability, classification rules were developed to address the situation in which different studies of the same assessment reported discrepant outcomes. The following rules were adopted:

- The mode or most frequent classification (i.e., "poor," "fair," "good," and "excellent") across studies was reported when there was only one study reporting a metric, or when more than one study reported a metric and one classification was most common.

  Example: Three studies report values falling into the classifications fair, good, and good; because "good" is the most frequently reported, the outcome is classified as good in the relevant table of this report (for an example of classifications used in this report, see Table 21).

- When an identical number of studies classified the instrument into two categories, and the categories were discrepant by *only one* classification, a dash was used to report the outcome.

  Example: Four studies report values falling into the classifications fair, fair, good, good; because no mode (or most common outcome) is present, the outcome is classified as "fair-good" in the relevant table of this report.

- When an identical number of studies classified the instrument into two categories, and the categories were discrepant by *more than one* classification, the average category (or the category falling between the two modes) was used to report the outcome.

  Example: Four studies report values falling into the classifications poor, poor, good, good; because no mode (or most common outcome) is present, and the outcomes are separated by more than one category, the outcome is classified as "fair", or the average of the two modes in the relevant table of this report.

Results

Thirty-six (36) instruments met the criteria for brief assessments of general or violent recidivism (excluding sexual violence), violent recidivism against an intimate partner (IPV), failure to appear, or technical violations among adult offenders. Twenty-eight (28) instruments met the criteria for predicting general and non-sexual violent recidivism among juvenile delinquents; however, the search identified no brief instruments for assessing domestic violence recidivism with juveniles.

A profile of each instrument was developed according to the coding criteria; instruments were then compared to each other with respect to administrative requirements, item

characteristics, validity, and reliability. Instruments were classified as *pre-conviction* if intended for use with adults or juveniles who had been arrested, detained, booked, or charged but not adjudicated. Instruments were classified as *post-conviction* if intended for use with adults or juveniles who had been adjudicated and convicted of a crime, even if they were still awaiting sentencing. The search process identified six (6) IPV assessments, all of which are presented together.

For ease of interpretation, this report will present adult and juvenile assessments separately, providing an overview of assessments intended for use at pre- and post-adjudication as well as intimate partner violence. Issues unique to screening and assessment of defendant, offenders, and intimate partner violence will be addressed within each section. Because instruments are sometimes used in a different setting than intended, analyses of the predictive validity and reliability of each instrument will be presented together, at the end of the respective adult and juvenile sections.

Brief Risk Assessment with Adults

      Defendants and the pretrial setting. One of the primary post-arrest decisions made by adult and juvenile justice systems is whether or not to hold a defendant pending resolution of the charges against them. Additional consideration is given to the circumstances of the release, in terms of bail, supervision, and other strategies to dissuade the defendant from committing additional crimes and encourage him/her to appear at future court proceedings. These decisions, which are most commonly made by a judge or other court officer, must balance concerns regarding public safety, the interest of the court in terms of ensuring the defendant attends future proceedings, and the due process rights of the defendant (VanNostrand, 2007). With respect to the latter, research shows that low-risk defendants who are detained pending trial receive more serious and lengthier sentences when compared to low-risk defendants who are released (Lowenkamp, VanNostrand, & Holsinger, 2013). Defendants also experience negative outcomes related to employment, income, health, and family as the result of incarceration (Subramanian, Delaney, Roberts, Fishman, & McGarry, 2015). Given the potential negative impact of detention, many advocacy groups recommend local policies that favor release for adult and juvenile defendants (Austin, 2004; Steinhart, 2006). Public safety is also better served through such policies, as indicated by research showing that the practice of "over" intervening, or providing supervision that exceeds risk level, may accrue additional public safety costs in the form of individuals who exit jail at a higher risk level than when they entered (Lowenkamp & Latessa, 2004).

In 2014, two-thirds of adult inmates in U.S. jails were pretrial defendants awaiting court action, a figure that has remained relatively steady since 2005 (Minton & Zeng, 2015). In federal district courts, 64% of pretrial defendants are detained for the duration of their case (Cohen, 2013). Such numbers suggest that current practice favors detention rather than release, a fact that may be partially explained by the nature of decision-making tools used by courts. Jurisdictions that make release decisions based solely on subjective criteria are twice as likely as those that use objective assessments to experience jail overcrowding, wherein the number of inmates exceeds the facility's capacity (Mamalian, 2011). Despite

clear evidence that the use of risk assessment at the pretrial stage can improve the court's ability to identify those offenders who can be safely released, relatively few U.S. court systems use validated assessment tools to inform release, bail, and supervision decisions with pretrial defendants (Subramanian et al., 2015)

Barriers to pretrial assessment. In 2012, U.S. law enforcement officers made more than 12 million arrests, most commonly for drug, traffic, and property offences (FBI, 2013). Ideally, all individuals arrested and booked into jail would be screened for risk of recidivism prior to their first court appearance, so that the release decision was informed by the risk of pretrial failure (Subramanian et al., 2015; see juvenile section, p. 38, for more on juvenile detention decisions). Gathering such information takes time, however, especially if the assessment requires a defendant interview. Many jurisdictions struggle to conduct assessments with so many people in compressed timeframes (VanNostrand, 2007). Compounding this problem is the relative lack of staffing for pretrial processing; the majority of assessment resources in correctional systems are devoted to planning treatment and supervision for sentenced offenders (Mamalian, 2011).

Much of the research on risk assessment has focused on convicted offenders; however, pretrial and post-conviction decision-making are not concerned with identical outcomes. When considering released defendants, the court is less concerned with punishment and rehabilitation and more narrowly focused on ensuring the defendant will be present at future court proceedings (Bechtel, Lowenkamp, & Holsinger, 2011). More recently, pretrial decision-making also takes into account public safety concerns and re-arrest, although some jurisdictions prohibit consideration of public safety as a factor when making release decisions with defendants (VanNostrand, 2007). Given such variation in relevant outcomes, traditional risk assessments may not have the same predictive accuracy when applied to defendants as they do with convicted offenders.

The Bureau of Justice Assistance makes the following recommendations when selecting a pretrial assessment instrument (Summers & Willis, 2010, p. 4):

- Pretrial risk assessment instruments (PRAI) should be consistent with the jurisdictional standards of relevant criteria for bail considerations, particularly with regard to race, ethnicity, gender, and financial status.

- Risk factors included in the PRAI need to be demonstrably related to failure to appear (FTA) and re-arrest rates, not solely to recidivism or general criminogenic factors.

- Risk factors and assessment terms should be clearly and unequivocally defined to ensure consistent evaluations.

- The instrument should be simple enough to use under day-to-day circumstances. Instruments that require specialized knowledge or overly time-intensive data collection are likely to be too burdensome for most jurisdictions.

- PRAIs must be validated and/or revised for the implementing jurisdiction. Jurisdictional variations in risk factors are likely.

- PRAIs should be relatively easy for criminal justice personnel to understand and administer.

The following section compares instruments designed for use with adult defendants, and the degree to which they align with the aforementioned goals and constraints of the pretrial setting.

   Overview of instruments. Table 4 describes the pre-conviction instruments for which eligible studies were located (N=11). The majority of instruments on this list would be classified as pretrial instruments, although the Service Planning Instrument (SPI) was designed for use in jails, with both pretrial defendants and convicted offenders, as a means for prioritizing reentry and discharge services. The RANT was designed to identify defendants who are appropriate for alternative dispositions such as drug court. Three-fourths of the instruments were designed for use in a single jurisdiction (city or county), which makes their utility unknown when administered in other locations. In contrast, the Federal Pretrial Services Risk Instrument (PTRA), which is used in federal jurisdictions across the nation, represents one attempt to create an instrument whose predictive ability transcends jurisdictional characteristics. Similarly, a shortened version of the Kentucky Pretrial Risk Assessment Instrument (KPRA) is currently being studied for use as a national model of pretrial risk assessment (called the PSA-Court). Of note, the oldest included study is from 2001, which is indicative of the relative infancy of pretrial risk assessment when compared to post-conviction risk assessment.

Table 4: Brief Instruments, Pre-conviction

Name	Acronym[a]	Related to	Versions	Date range	Intended Jurisdiction[b]
Federal Pretrial Services Risk Assessment	PTRA			2009-2012	National
Hennepin County Pretrial Risk Assessment	(HC-PRA)		2	2006-2007	County
Inventory of Need Pretrial Screening Tool	INPST	WRNA[c]		2014	County
Kentucky Pretrial Risk Assessment Instrument	KPRA	PSA-Court[d]	2	2012-2013	State
NYC Criminal Justice Agency Pretrial Release Recommendation System	CJA		2	2001-2006	City
Ohio Risk Assessment System Pretrial Assessment Tool	ORAS-PAT			2008-2009	State
Polk County Pretrial Release Point Schedule	(Polk)		2	2008	County

Name	Acronym[a]	Related to	Versions	Date range	Intended Jurisdiction[b]
Risk and Needs Triage	RANT			2009	County
Salt Lake County Pretrial Release Instrument	SLPRI			2011	County
Service Planning Instrument	SPI			2012	City

[a] Acronyms in parentheses were added by authors of the current report for purposes of referring to instruments more parsimoniously in subsequent tables; they are not the acronyms of the researchers who developed the instruments.
[b] For what type of jurisdiction was the instrument developed
[c] Women's Risk Needs Assessment
[d] A 7-item version of the KPRA, called the KPRA-S, does not require a defendant interview and is currently being studied for use nationally, called the PSA-Court

Table 5 provides an overview of the general population for which each instrument was intended as well as additional information related to the types of decisions the instrument informs. The vast majority assess for pretrial failure, which is defined as failure to appear or new arrest. The Risk and Needs Triage (RANT), which is the only instrument designed to inform disposition and placement (such as drug court), does not assess for failure to appear. Also, the SPI, which is used to assess both defendants and convicted offenders, does not assess for failure to appear. Because the included instruments are designed to inform pretrial decision-making, most classify defendants into at least three categories; as such, the risk category is often tied to specific suggestions with regard to release conditions. For example, the three categories of the Hennepin County Instrument correspond to the following release recommendations: low risk defendants are recommended for release without bail or supervision; medium risk defendants are recommended for release with conditional supervision; and high risk defendants are recommended held for review and bail. Of note, these recommendations are made by the pretrial unit to the judge, who maintains ultimate discretion for the final decision.

Table 5: Brief Instruments, Pre-conviction: Overview

Instrument	Intended Population	Intended Outcomes	Intended Decisions	Risk Categories
CJA	Adult Defendants (16+)	Pretrial failure	Release	3
HC-PRA	Adult Defendants, Felony/Targeted Misdemeanor[a]	Pretrial failure	Release	3
INPST	Adult Female Defendants	Pretrial failure	Release	
KPRA	Adult Defendants	Pretrial failure	Release	3[b]
ORAS-PAT	Adult Defendants	Pretrial failure	Release	3
Polk	Adult Defendants	Pretrial failure	Release	2
PTRA	Adult Defendants	Pretrial failure, including technical violations	Release	5
RANT	Adult Drug-involved Defendants	New crime	Disposition Placement	4
SLPRI	Adult Defendants	Pretrial failure	Release	FTA, 3 New Arrest, 4
SPI	Adult Defendants (& Offenders)	Short-term readmission to custody	Discharge planning	4

[a] Targeted misdemeanors include DUI or domestic violence
[b] The KPRA-S classifies into 5 risk categories

Three of the instruments were designed for use with specific types of defendants (Hennepin County, INPST, and RANT). While none of the included studies provided analyses of predictive validity by offender type, Table 6 describes the types of offenses for which the study participants were initially arrested (excluding the ORAS-PAT and the SLPRI, for which none of the included studies provided frequencies with respect to offense type). All study samples included property offenders and most included violent, drug, and felony offenders. Only one instrument, the Inventory of Need Pretrial Screening Tool (INPST) was designed for use specifically with female offenders. While studies were coded for validity with sub-groups, none evaluated differences in predictive validity with respect to gender[10].

Table 6: Brief Instruments, Pre-conviction: Characteristics of Study Populations

Instrument	Felony	Misdemeanor	Violent	Drug	Property	DUI	Other
CJA	X	X	X	X	X	X	X
HC-PRA	X		X	X	X		X
INPST			X		X		
KPRA	X	X	X		X	X	X
Polk			X	X	X		X
PTRA	X	X	X	X	X		X
RANT	X			X	X		
SLPRI	X	X	X	X	X	X	X

With the exception of the SPI, all of the instruments were designed for use by pretrial services staff, who are commonly trained probation officers (Table 7). Recall that the SPI was designed for screening both pretrial and convicted individuals and so is intended to be used by correctional staff in jail. None of the instruments required any specific professional license or degree, but three require specific training or certification. The lack of formal training requirements partially stems from the fact that the instruments were developed for use in specific jurisdictions; as such, studies often referred to in-house training for assessors, but no resources were devoted to developing training for individuals outside the jurisdiction. Very few of the studies indicated how long it takes to conduct the brief screen. All except the INPST require an interview with the defendant, however, which takes longer than self-report or simple file review. The INPST is the most flexible instrument in terms of collecting information, and is designed so that defendants can be assessed via interview or self-report. One-third of instruments had been field-tested, which was defined as having at least one study wherein validity or reliability was evaluated when pretrial staff, rather than researchers, was completing the instrument.

---

[10] The Polk County Pretrial Release Point Schedule was evaluated for predictive validity by gender, but there were 0 FTAs for females and the overall results (new crime and FTA) were not reported by gender.

Table 7: Brief Instruments, Pre-conviction: Administrative Requirements

Instrument	Intended Assessor	Training/ Certification	Time	Interview Required	Self-report instrument	Field-tested
CJA	Pretrial services			X		
HC-PRA[a]	Pretrial services			X		
INPST	Pretrial services			X[b]	X[b]	
KPRA	Pretrial services			X[c]		X
ORAS-PAT	Pretrial services	X	20	X	X	X
Polk	Pretrial services			X		
PTRA	Pretrial services	X		X		X
RANT	Pretrial services	X	15	X		X
SLPRI	Pretrial services			X		
SPI	Correctional officers in jail					

[a] Revised version
[b] Can be completed via interview or self-report
[c] KPRA-S does not require an interview

Items were coded as static, dynamic, or other (which means there is no current research demonstrating an association between the item and recidivism). When looking at the types of items included in the instruments, nearly all were comprised of primarily static risk items, Table 8. All but two (SPI and SLPRI) included at least one "other" item (not included in table), most commonly related to community ties, such as housing. While Baird (2009) argued for parsimonious assessment instruments that do not include extraneous items not statistically related to recidivism, pretrial instruments likely include those items due to: 1) statutory requirements mandating detention of certain types of offenders; and, 2) the specific interest in risk of failure to appear in addition to new crimes.

Similarly, the majority of instruments were comprised primarily of items that were characterized by the research team as objective, which meant that they were either easy to verify (for example, in criminal records) or structured according to simple criteria (e.g., ever attended substance abuse treatment versus has a substance abuse problem). For purposes of the current study, all static, dynamic, and other items were classified as objective or subjective. The two instruments that included primarily dynamic and subjective items were the INPST for males and females; because these instruments can be completed via self-report (see Table 7), inter-rater reliability—which is one of the primary issues with subjective assessment instruments—is not an issue.

Table 8: Brief Instruments, Pre-conviction: Items

Instrument	Items	Static Factors	Dynamic Factors	Protective Factors	Objective Items	Subjective Items
CJA[a]	6	2	1	--	5	1
HC-PRA	11	9	1	--	10	1
INPST, Males	30	8	12	Yes	16	13
INPST, Females	38	9	11	Yes	22	16
KPRA	12[b]	9	2	--	10	2
Polk	17	6	7	--	14	3
PTRA	11	6	3	--	10	1
RANT[a, c]	16	2[d]	5	--	3	7

Instrument	Items	Static Factors	Dynamic Factors	Protective Factors	Objective Items	Subjective Items
SLPRI[e]	7	6	1	--	6	1
SPI	4	4	--	--	4	--
ORAS-PAT	7	3	3	--	4	3

[a] Revised version
[b] KPRA-S has 7 items, all of which are static
[c] RANT is comprised of two scales, Risk and Need
[d] Exact list of items not available, numbers on this line refer to characteristics of broader domains
[e] The SLPRI is comprised of two scales, Failure to Appear and New Crime, which share some items

When looking at the specific domains assessed, Table 9 indicates that anti-social behavior (in the form of prior criminal history) was included in all of the instruments. Juvenile offense history was only included in three of the instruments, likely due to the relative difficulty of obtaining juvenile records in the short timeframe of pretrial. The majority of instruments also included questions related to employment, education, substance abuse, and housing stability. Fewer assessed for peers and mental health concerns and none asked questions related to the criminogenic need items of recreation and leisure or anti-social personality (not in table).

Table 9: Brief Instruments, Pre-Conviction: Domains

Instrument	Current Offense Characteristics	Anti-social behavior	Juvenile Offenses[a]	Relationships/ Peers	Work/ School	Substance Abuse	Mental Health	Housing
CJA		X[b]		X	X			X
HC-PRA	X	X			X	X		X
INPST, men	X	X		X	X	X	X	X
INPST, women	X	X		X	X	X	X	X
KPRA	X[c]	X			X	X		X
ORAS-PAT		X	X		X	X		X
Polk		X			X	X		X
PTRA	X	X			X	X		X
RANT	X	X	X	X	X	X	X	
SLPRI	X	X	X			X		
SPI	X	X						

[a] Items related to juvenile offense history (including asking about age at first arrest)
[b] Revised version includes criminal history; original did not.
[c] Only for full KPRA but not KPRA-S

A comparison of the predictive validity and inter-rater reliability of the pre-conviction instruments is discussed later in Tables 21 and 24 (pgs. 33-37).

Brief risk assessment, post-conviction. As noted earlier, the term post-conviction refers to the period after an offender has been convicted of a crime, whether they are awaiting sentencing, incarcerated, or on community supervision. Risk assessment at the post-conviction stage focuses primarily on the criminal justice goals of public safety and offender rehabilitation. As such, post-conviction risk assessment informs decisions related to: placement type; supervision level; treatment type and dosage; timing and conditions of release; and sanctioning technical violations (NJC, 2012). The application of risk prediction

methods to sentencing decisions—which focus on the charge rather than the individual—is more complicated because of potential conflict between the goals of public safety, offender rehabilitation, and punishment and accountability (Hannah-Moffat, 2010; Latessa & Lovins, 2010).

Barriers to risk assessment. While criminal justice entities working with convicted offenders typically have more assessment resources than pretrial agencies, the system still struggles to conduct comprehensive risk assessment with the 7 million people under correctional supervision at any given time (Glaze & Kaeble, 2014). Capacity is further strained by the fact that assessment should be an ongoing process, occurring at each major transition point of an individual's criminal justice involvement. In a given year in the U.S., there are close to 1 million prison intake events, 2.1 million supervision entrances, and 2.1 million supervision exits (Glaze & Kaeble, 2014; Herberman & Bonczar, 2015), which does not account for transitions in and out of jail and technical violations. Ideally, each of these events would include a risk assessment, although one instrument may not be suitable for all decisions (e.g., pretrial, prison intake, entry to community supervision). In the context of post-conviction, brief screening systems can be a cost-effective way to assess all offenders at the appropriate point while allocating additional resources to those individuals who will benefit most (NJC, 2012).

The inclusion of dynamic factors on assessment instruments is likely at the post-conviction stage, because of explicit system goals related to risk reduction in prison and probation/parole settings. The importance of including criminogenic need items at the screening stage is less clear. While some research demonstrates that dynamic need items can increase the predictive validity of an instrument (Latessa & Lowenkamp, 2005b), criminal justice researchers have also expressed concern that the inclusion of both static and dynamic items conflates risk classification with risk reduction (Baird, 2009). In addition, dynamic factors are inherently more subjective and therefore "more problematic from a methodological viewpoint because they are more difficult to measure" (Gottfredson & Moriarity, 2006, p. 191). At a minimum, the inclusion of dynamic and subjective items requires more organizational resources in terms of staff time and training.

Overview of instruments. Table 10 describes the post-conviction instruments for which eligible studies were located (N=18). Almost half are screening versions of longer instruments that provide comprehensive risk and needs assessment. When compared to the pre-conviction instruments, these tools tend to be designed for use in a wider range of geographic and correctional settings. The date range of the studies is also indicative of the fact that many of these instruments have been around longer than the pre-conviction tools and have more research evaluating their validity.

Table 10: Brief Instruments for use Post-conviction[a]

Name	Acronym[b]	Screening Version	Related to	Versions	Date range	Locale[c]
Correctional Offender Management Profiling for Alternative Sanctions, General Recidivism Risk Scale	COMPAS Recidivism Scale (COMPAS-R)	X			2008-14	1[d]
Correctional Offender Management Profiling for Alternative Sanctions, Violent Recidivism Risk Scale	COMPAS Violence Scale (COMPAS-V)	X			2008-14	1[d]
Justice System Assessment & Training Proxy Score Risk Assessment	Proxy				2008-14	1[d]
Level of Service Inventory-Revised, Screening Version	LSI-R:SV	X			2005-11	2[d,g]
Lifestyle Criminality Screening Form	LCSF			2	1995-2014	2[h]
Minnesota Screening Tool Assessing Recidivism Risk	MnSTARR				2014	1[d]
Offender Group Reconviction Scale	OGRS		OASys GRP[i]	3	2004-14	1[j]
Offender Risk and Management System, Primary Risk Assessment	ORAMS PRA	X	WRA[k]		2002	1[h]
Ohio Risk Assessment System, Community Supervision Screening Tool	ORAS CSST	X[l]			2009	1[d]
Orange County California Probation Department Risk Assessment Instrument	(OC-RAI)	X	WRA[j]	2	2011	1[d]
Post-Conviction Risk Assessment	PCRA		PICTS[e]		2011-15	1[d]
Self-Appraisal Questionnaire	SAQ				2000-12	4[d,g,h,m]
Service Planning Inventory, Pre-screen	SPIn	X	SPIn-W[n]		2015	1[l]
Static Risk Assessment	SRA	X[o]	LSI-R	2	2008-14	1[d]
Statistical Information on Recidivism-Revised	SIR-R1			3	1984-2011	1[h]
Wisconsin Risk Assessment	WRA	X		2	2003-11	1[d]
Virginia Nonviolent Offender Risk Assessment Instrument	(NORAI)			2	2002-12	1[d]

[a] Instruments intended to assess specifically for risk of recidivism and other behaviors with sentenced offenders

[b] Acronyms in parentheses were added by authors of the current report for purposes of referring to instruments more parsimoniously in subsequent tables; they are not the acronyms of the researchers who developed the instruments.

[c] The number of different countries in which the instrument has been assessed for reliability and/or predictive validity

[d] US

[e] Includes items from Pyschological Inventory of Criminal Thinking; items do not contribute to risk score.

[f] Assessed across US federal districts

[g] Australia

[h] Canada

[i] Offender Assessment System General Reoffending Predictor; OGRS comprises the static component

[j] United Kingdom

[k] Wisconsin Risk Assessment

[l] Ohio Risk Assessment System Community Screening Tool

[m] Singapore

[n] SPIn-W for female offenders

[o] Is the static risk assessment of the STRONG

Table 11 provides an overview of the general population for which each instrument was intended as well as additional information related to the types of decisions the instrument informs. All assess for new crime and the majority assess for at least two outcomes. The inclusion of both facility-based behaviors, such as institutional misconduct, as well as community-based behaviors, such as technical violations of supervision conditions, reflects the range of settings in which these instruments are intended to be used. Primarily, these tools are intended to sort offenders into categories that allow for the development of appropriate supervision strategies; however, several—such as the LSI:SV and the SAQ—are also intended to identify treatment needs.

The majority of instruments classify offenders into three or four risk categories (e.g., low, medium, high, and very high). In contrast, the Ohio Risk Assessment Community Supervision Screening Tool (ORAS CSST) simply classifies offenders as low or moderate/high risk. This very brief tool has one purpose, which is to quickly screen those offenders for whom additional assessment is not necessary and refer the rest to the full community supervision screening tool. The Virginia Nonviolent Offender Risk Assessment Instrument (NORAI) also classifies offenders into two categories, which are eligible or not eligible for alternative dispositions, and is the only included instrument that was designed to inform sentencing decisions.

Table 11: Brief Instruments, Post-conviction: Overview

Instrument	Intended Population	Intended Outcomes	Intended Decisions	Risk Categories
COMPAS-R	Adult Offenders	New Crime	Full Assessment	3
COMPAS-V	Adult Offenders	New Violent Crime	Full Assessment Placement Supervision	3
LCSF	Adult Offenders	New Crime Institutional Misconduct	Placement Supervision	N/A
LSI-R:SV	Adult Offenders	New Crime Technical Violation Institutional Misconduct	Full Assessment	3
MnSTARR, Male	Adult Male Offenders	New Crime, Sex New Crime, Nonsexual Violent New Crime, Felony	Service Prioritization Supervision	4
MnSTARR, Female	Adult Female Offenders	New Crime, nonsexual violent New Crime, Nonviolent New Crime, Felony	Service Prioritization Supervision	4
NORAI	Adult Felony Drug and Property Offenders[b]	New Felony Conviction	Disposition[c]	2
OC-RAI	Adult Offenders	New Crime New Violent Crime	Supervision	3
OGRS	Adult Offenders with Prior Conviction	New Crime	Service Prioritization Supervision	4
ORAMS PRA	Adult Offenders	New Crime Technical Violation	Placement Supervision	3
ORAS CSST	Adult Offenders	New Crime	Full Assessment Supervision	2

Instrument	Intended Population	Intended Outcomes	Intended Decisions	Risk Categories
PCRA	Adult Offenders	New Crime Technical Violation	Supervision	4
Proxy	Adult Offenders	New Crime	Supervision	3
SAQ	Adult Male offenders	New General Crime New Violent Crime	Supervision Treatment Needs	4
SIR-R1	Adult Male offenders	New Crime Technical Violations	Supervision	5
SPIn	Adult Offenders	New Crime	Supervision Treatment Needs	10[a]
SRA	Adult Felony offenders	New Crime	Supervision	5
WRA	Adult Offenders	New Crime Technical Violations	Supervision	3

[a] Combines 4 possible risk categories with 4 possible strengths categories
[b] Without prior violence conviction
[c] Decision whether offender is eligible for alternative to incarceration

Only one instrument, the MnSTARR, has separate versions for use with male and female offenders. In comparison to the pre-conviction assessments, however, most of the post-conviction instruments have undergone substantially more evaluation with respect to predictive validity among sub-populations. The majority has been examined for validity with both male and female offenders (see Table 22 for more information)

Table 12: Brief Instruments, Post-conviction:
Assessed for Validity with Sub-groups

Instrument	Gender	Age	Offender Type
COMPAS-R	X		
COMPAS-V	X		
LCSF	X		
LSI-R:SV	X		Mentally ill, Dual diagnosis
MnSTARR, male			Felony, Sex[a], Violent
MnStarr, female			Violent, Felony
NORAI			
OC-RAI			
OGRS	X	X	Mentally ill, Sex, Violent
ORAMS PRA	X	X	
ORAS CSST	X		
PCRA	X		
Proxy			
SAQ	X		Drug
SIR-R1	X		
SPIn	X		
SRA			
WRA	X		

[a]Slightly modified version of instrument was developed for use with male sex offenders

All of the instruments were designed to be administered by correctional officers, in facility-based and/or community settings (Table 13). None of the instruments have professional restrictions—such as advanced degrees or clinical licensure—in terms of administration and scoring, although the Self-Appraisal Questionnaire (SAQ) must be interpreted by

someone with a clinical background. Close to half of the tools require formal training or certification to administer, which reflects the fact that the instruments were developed for use across jurisdictions and settings and resources were therefore allocated to developing training curricula. While very few of the studies indicated how long it takes to conduct the brief screen, most for which information was provided took less than 20 minutes. One-third of the instruments require an interview with the offender, while the rest can be completed via file review or self-report, both of which are comparatively less resource-intensive in terms of staff time and training. The reliance on official records is possible, in part, because of in-depth information that accrues while an offender is in custody or under supervision, including pre-sentence reports, treatment records, supervision history, and other sources. The majority of the instruments had been field-tested, which was defined as having at least one study wherein validity or reliability was evaluated when correctional officers, rather than researchers, were completing the instrument.

Table 13: Brief Instruments, Post-conviction: Administrative Requirements

Instrument	Intended Assessor	Training/ Certification	Time	Interview Required	Self-report instrument	Field-tested
COMPAS-R	Corrections officer[a]	2-day training	10	Possible[b]	Possible[b]	Yes
COMPAS-V	Corrections officer[a]	2-day training	10	Possible[b]	Possible[b]	Yes
LCSF	Corrections officer[c]		10	No[d]	No	Yes
LSI-R:SV	Corrections staff[c]	Certification	15	Yes	No	Yes
MnSTARR	Corrections officer[c]		30	No	No	No
NORAI	Probation Officer or Prosecutor			No[i]	No	Yes
OC-RAI	Community corrections			Yes	No	Yes
OGRS	Corrections officer[c]			No	No	No
ORAMS PRA	Community corrections			Yes	No	Yes
ORAS CSST	Community corrections	Certification		No	Yes	No
PCRA	Community corrections	16-hours		No[e]	No[e]	Yes
Proxy	Community corrections			No	No	Yes
SAQ	Corrections officer[c]	Yes[f]	20[g]	No	Yes	N/A
SIR-R1	Corrections officer[h]			Yes	No	Yes
SPIn	Corrections officer[c]	4-day training		Yes	No	No

Instrument	Intended Assessor	Training/ Certification	Time	Interview Required	Self-report instrument	Field-tested
SRA	Community corrections			No	No	Yes
WRA	Community corrections			Yes	No	Yes

[a] Including Institutional, Community Corrections, and Pre-sentence
[b] Can be completed via interview or self-report
[c] Institutional and community corrections
[d] Instrument itself is completed by chart review; however, identifying items are based on information gathered during pre-sentence investigation, which often includes an interview.
[e] Data available in records because collected during pre-sentence report; self-report portion does not contribute to risk score.
[f] Can be administered and scored by clerical or correctional staff; interpretation requires clinical background
[g] 15 to take, 5 to score
[h] Institutional staff at release
[i] Completed using information regularly collected during pre-sentence investigation

Approximately half of the brief screening instruments are comprised of 10 or fewer items (Table 14). Of the remaining instruments, one is self-report (SAQ) and seven do not require an offender interview (LCSF, MnSTARR, NORAI, PRA, PCRA, and SRA), both of which require less in the way of staff resources to administer. All the instruments are comprised of some static risk factors and the vast majority also includes at least one dynamic risk factor. For the most part, the instruments are comprised of more objective than subjective items, which is important when considering their utility as a brief screening instrument. Those instruments that include more subjective items may require more in the way of staff training in order to ensure congruence among different staff that is using the instrument. The research team also coded instruments for the scoring method, which, ideally, is straightforward if the instrument is intended for use as a screening tool. The majority were scored by either simple addition (LSI-R:SV), wherein the item is identified as present or absent, or through use of a weighted system, wherein items are given certain points (often derived from a regression analysis) and summed.

Table 14: Brief Instruments, Post-conviction: Items

Instrument	Items[a]	Scales	Scoring	Static Factors	Dynamic Factors	Objective Items	Subjective Items
COMPAS-R	4		Weighted	2	2[b]	2	2
COMPAS-V	5		Weighted	4	1[b]	5	--
LCSF	14	4[c]	Weighted	4	5[b]	8	6
LSI-R:SV	8		Simple	2	5[b]	3	5
MnSTARR, Male	23		Unknown	13	2[b]	20	3
MnSTARR, Female	20		Unknown	9	4[b]	14	6
NORAI	11		Weighted	8	3	10	1
OC-RAI	10		Weighted	5	4	7	3
ORAMS PRA	15		Weighted	4	7	8	7
OGRS	6[d]		Computer	6	--	6	--
ORAS CSST	4		Weighted	1	3[b]	2	2
PCRA	15	2[e]	Weighted	12	3	14	1
Proxy	3		Weighted	3	--	3	--
SAQ	72	7[f]	Unknown	3[g]	4[g]	Unknown	Unknown

Instrument	Items[a]	Scales	Scoring	Static Factors	Dynamic Factors	Objective Items	Subjective Items
SIR-R1	15		Weighted	10	2[b]	15	--
SPIn	35	2[h]	Unknown	21	7	22	13
SRA	26[i]	4[j]	Weighted	26	--	26	--
WRA[k]	11		Weighted	6	4	8	3

[a] Some of the items included in instruments were characterized as "other," meaning they have not been demonstrated to be static or dynamic predictors of recidivism; as such total items does not always equal the sum of static and dynamic factors.

[b] Items are classified as dynamic even if they are conceptualized statically (e.g., highest educational attainment) because brief screening instruments have better IRR when items are objective.

[c] Irresponsibility; Self-indulgence; Interpersonal intrusiveness; Social rule-breaking, all of which have a total score of 5-6.

[d] For the OGRS-3, previous version has 9 items

[e] Responsivity items comprise second scale but do not contribute to risk score and are not included here.

[f] Criminal tendencies; Antisocial personality problems; Conduct problems; SUD; Antisocial Associates; and Anger, which does not contribute to risk score.

[g] For domains rather than individual items, which were not available

[h] Risk/Need and Strengths

[i] Revised version has 23

[j] Produces separate scores for drug, property, violence and felony

[k] Initial version included adjudication for assaultive offense within last five years, which did not predict recidivism; this was replaced in most recent version with age at start of probation. Coding here is for revised version (Eisenberg, Bryl, & Fabelo, 2009; Henderson & Miller, 2011).

When looking at the specific domains assessed, Table 15 indicates that anti-social behavior (in the form of prior criminal history) was included in all of the instruments. The majority of instruments also included questions related to employment or education, substance abuse, and family relationships and peers. Demographic items—most commonly age— were also included on the majority of instruments. Fewer assessed for mental health concerns or personality problems, likely because such an assessment requires additional time or training, and none asked questions related to the criminogenic need items of recreation and leisure (not in table). While the latter may feel like a distal need, particularly for incarcerated offenders, it is significantly associated with recidivism and may be particularly relevant for offenders who are exiting prison to community supervision.

Table 15: Brief Instruments, Post-conviction: Domains

Instrument	Attitudes	Anti-social behavior	Relationships/ Peers	Work/ School	Substance Abuse	Mental Health	Housing	Demographics
COMPAS-R		X	X	X	X			X
COMPAS-V		X		X				X
LCSF		X	X	X	X			
LSI-R:SV	X	X	X	X	X	X		
MnSTARR, Mfale		X	X	X	X			X
MnSTARR, Female		X	X	X	X			X
OC-RAI	X	X		X	X		X	X
OGRS-3		X						X
ORAMS PRA	X	X	X	X	X	X	X	X
ORAS CSST		X	X	X				

Instrument	Attitudes	Anti-social behavior	Relationships/ Peers	Work/ School	Substance Abuse	Mental Health	Housing	Demographics
PCRA		X		X	X			X
Proxy		X						X
SAQ	X	X	X		X			
SIR-R1		X	X	X				X
SPIn	X	X	X	X	X	X	X	
SRA		X						X
WRA	X	X	X	X	X	X	X	X
NORAI		X	X	X				X

A comparison of the predictive validity and inter-rater reliability of the post-conviction instruments is discussed later in Table 22 (p. 34).

Brief risk assessments for intimate partner violence. Lifetime prevalence rates for physical abuse by an intimate partner approach 25% for U.S. women and 14% for men (Breiding, Chen, & Black, 2014). While many domestic incidents never enter the criminal justice system, a study using data from the Federal Bureau of Investigation's Incident Based Reporting System from 19 states (including Utah) found that law enforcement were called to nearly 600,000 incidents in a single year and made arrests in 37% of those cases (Hirschel, Buzawa, Pattavina, Faggiani, & Reuland, 2007). Such numbers suggest that domestic violence offenders comprise a substantial portion of the criminal justice population. While risk assessment instruments designed to predict violent recidivism have been used to predict intimate partner violence (IPV), many are inappropriate for use as screening tools because they must be administered by forensic mental health professionals and/or rely on subjective assessments regarding mental health diagnoses (Nicholls, Pritchard, Reeves, & Hilterman, 2013). More recently, instruments designed specifically to predict IPV have been developed.

The prediction of intimate partner violence (IPV) must attend to the following concerns: 1) many brief instruments are designed for use at the scene of an incident, prior to an individual being charged or even arrested for a crime; 2) risk management strategies must account for harm to a specific person, the victim, as well as the broader public; 3) IPV involves a victim and perpetrator with ongoing emotional, financial, or other attachments; and, 4) wide variability exists both within and between statutory, research, and victim advocate-based definitions of IPV. IPV risk assessment tools are often designed for use by first responders and are intended to inform a broad scope of decisions, including perpetrator risk (e.g., arrest, release, and supervision) and rehabilitation (treatment) and also victim concerns (e.g., safety planning and victim services) (Messing & Thaller, 2013). IPV assessments are used in a range of capacities that exceed the scope of many risk prediction tools, such as educating victims with respect to risk level. Risk classification categories often include recommendations for the victim as well as the perpetrator. In response to the unique characteristics of intimate partner risk prediction, many instruments rely on victim, as well as perpetrator interviews (Nicholls et al., 2013).

Instruments were included in the current study if they met the following criteria: 1) were evaluated according to their validity in predicting IPV recidivism; and, 2) met the criteria for a brief risk screen. While research has demonstrated that psychological abuse of an intimate partner is both harmful and predictive of physical assault (Nicholls et al., 2013), such behavior is not a crime under most state criminal codes and therefore would not be classified as recidivism for the purposes of the current study.

Overview of instruments. Table 16 describes the intimate partner violence instruments for which eligible studies were located (N=6). While these instruments share some similarity with the pretrial tools—in that they are often intended for use with individuals who have not been convicted of an offense—they are unlike those instruments in that all were designed for use across geographic locations, as evidenced by the many locales in which they have been implemented and evaluated.

Table 16: Brief Instruments for Assessing Risk of Intimate Partner Violence (IPV)

Name	Acronym[a]	Screening Version	Related to	Versions	Date range	Locale[b]
Brief Spousal Assault Form for the Evaluation of Risk	B-SAFER	X	SARA[c]		2011-2014	3[d]
Domestic Violence Screening Instrument	DVSI	X[e]		2[f]	2004-2014	7[g]
Domestic Violence Supplementary Report	DVSR				2004-2009	1[h]
Lethality Screen	(LS)		DA[i]		2015	1[j]
Ontario Domestic Assault Risk Assessment	ODARA		DVRAG[k]		2004-2015	5[l]
Partner Abuse Prognostic Scale	PAPS		CTS2[m] AUDIT		2003	1[j]

[a] Acronyms in parentheses were added by authors of the current report for purposes of referring to instruments more parsimoniously in subsequent tables; they are not the acronyms of the researchers who developed the instruments.
[b] The number of different jurisdictions in which the instrument has been assessed for reliability and/or predictive validity
[c] Spousal Risk Assessment Guide
[d] Europe
[e] Is intended to identify offenders who require additional assessment and has been studied as a pre-screen for the SARA
[f] Revised version, the DVSI-R, is intended to be administered by master's level clinicians
[g] Canada and United States
[h] Canada
[i] Comprised of items from the Danger Assessment
[j] United States
[k] Domestic Violence Risk Appraisal Guide, which combines the ODARA and PCL
[l] Europe and Canada
[m] Includes items from Revised Conflict Tactics Scales in relationship violence scale and AUDIT in the substance abuse scale

Table 17 provides an overview of the type of IPV offender for which each instrument was intended. Of note, many of these instruments are intended to assess risk at the scene of an incident; as such, some of the individuals with whom they are used will never be charged or arrested; the term perpetrator is used, rather than defendant or offender, to avoid connotations of a particular criminal justice status. In addition, most of the instruments are designed to assess risk to a specific victim and therefore include both victim and perpetrator as the intended focus of the assessment (both in terms of gathering

information and suggesting an intervention per the risk classification). The Lethality Screen, which is a short version of the Danger Assessment (DA), was designed to predict risk based on the victim's assessment, without input from the perpetrator. Half of the instruments were designed to predict risk of future IPV by a male perpetrator against a female spouse or partner.

Unlike the majority of instruments, the B-SAFER, is not an actuarial tool and does not sort perpetrators into risk categories. Points are assigned to items based on the interview, but the ultimate risk classification is made by the assessor (in this case, law enforcement). While this type of structured professional judgment tool would not typically meet the criteria of brief screen, the B-SAFER was specifically designed for use by officers in the field and is therefore included in the current study.

Table 17: Brief IPV Instruments: Overview

Instrument	Intended Population	Intended Outcomes	Intended Decisions	Risk Categories
B-SAFER	Adult IPV perpetrators	IPV	Full assessment Risk management	3
DVSI	Adult male IPV perpetrators	IPV Non-compliance Risk to victim	Full assessment Release Risk management	3
DVSR	Adult IPV perpetrators	IPV, female partner	Risk management	N/A
LS	Adult IPV victims	Fatal or severe IPV	Victim services	2
ODARA	Adult male IPV perpetrators	IPV, female partner, Severity & Frequency	Release Victim Services	7
PAPS	Adult male IPV perpetrators	IPV	Treatment progress	2

The included instruments were designed for use by law enforcement, treatment providers, and pretrial staff (Table 18). None require an advanced degree to administer or score; only two provided information regarding training or certification requirements. Four (B-SAFER, DVSI, Lethality Screen, and ODARA) were designed to be completed by law enforcement with information commonly available at the scene of an incident. Of note, a greater number rely on victim, rather than perpetrator interviews; this partially reflects research showing the predictive validity of victim input when assessing risk of IPV (Nicholls et al., 2013). One-third of instruments had been field-tested, which was defined as having at least one study wherein validity or reliability was evaluated when law enforcement, treatment staff, or probation officers, rather than researchers, was completing the instrument.

Table 18: Brief IPV Instruments: Administrative Requirements

Instrument	Intended Assessor	Training/ Certification	Time	Offender Interview	Victim Interview	Field-tested
B-SAFER	Police, at scene			X[a]	X[a]	X
DVSI	Probation officers, pre-arraignment	X[b]		X	X[c]	X
DVSR	Police, at scene	X[d]			X	
LS	First responders, at scene				X	
ODARA	Police, at scene	X[e]				
PAPS	Treatment providers		30	X	X	

[a] Preferred method is that information solicited from both parties
[b] Revised version, DVSI-R requires master's level clinician
[c] Also requires check of protective order registry
[d] Included in IPV safety planning course
[e] Online, 4-6 hours

Items were coded as static, dynamic, or other (meaning there was no current research demonstrating a significant association between the item and recidivism). Items were classified as static and dynamic factors due to a specific association with IPV recidivism (such as recent separation between intimate partners and prior threats to harm victim and family), even if the item was not associated with general recidivism. When looking at the types of items included in the instruments, nearly all were comprised of both static and dynamic risk items (Table 19). In comparison to the pre- and post-conviction instruments, the IPV screening tools were comprised of more items that were classified as subjective.

Table 19: Brief IPV Instruments: Items

Instrument	Items	Scoring	Scales	Static	Dynamic	Objective	Subjective
B-SAFER	10	SJP	3[a]	5	5	4	6
DVSI	12[b]	Weighted		8	1	10	2
DVSR	19	Simple		4	4	2	17
LS	11[c]	Decision tree	2	8	3	2	9
ODARA	13	Simple		10	3	6	7
PAPS	17	Weighted	3[d]	10	7	10	7

[a] Imminent Risk; Long-term Risk; Risk of Extremely Serious Assault or Death
[b] DVSI-R has 11 items (5 objective) plus 2 subjective assessment questions that function as overrides
[c] Presence of any one of the first 3 items indicates increased danger OR presence of any 4 of the last 8 items indicates danger
[d] Relationship Violence History; Substance Use; Aggression History

The domains used to code items on the IPV screening tools were derived from The Colorado Domestic Violence Risk and Needs Assessment (DVRNA), which conceptualizes known IPV risk and need factors within an RNR framework (Gover, Richards, & Tomisch, 2015). Items related to previous IPV incidents, as well as other criminal history, were included as separate domains, as well as any recent changes in the relationship between the victim and the perpetrator, threats or use of weapons, and perpetrator suicidality. When looking at the domains assessed, Table 20 indicates that previous intimate partner violence on behalf of the perpetrator was included in all of the instruments while other criminal history was not. The majority of instruments included questions related to the

perpetrator's history of employment, and substance abuse. Circumstances—such as a recent separation between partners or the perpetrator's access to weapons—were included in half the instruments.

<p style="text-align:center">Table 20: Brief IPV Instruments: Domains[a]</p>

Instrument	Attitudes/ IPV	IPV	Criminal History[b]	Current Offense[c]	Work	SU[d]	MH[e]	Suicidality	Change/ Separation	Weapon[f]
B-SAFER	X	X	X	X	X	X	X			
DVSI		X	X	X	X	X			X	X
DVSR		X				X	X	X	X	X
LS		X			X			X	X	X
ODARA		X	X	X		X				
PAPS		X	X		X	X				

[a] Items revised from RNR model to incorporate IPV-specific assessment domains, per Gover, Richards, & Tomsich, 2015.
[b] Perpetrator's non-IPV criminal history
[c] Instrument assesses for characteristics of current offense
[d] Assesses for perpetrator substance use/abuse
[e] Assesses for perpetrator mental health issues
[f] Includes use of weapon in most recent incident (DVSI) and ongoing access to weapon (DVSR and LS)

Predictive validity effect size classifications. The following section provides the classification ratings for each of the adult assessments evaluated as part of the scope of this report. In keeping with the previous discussion, assessments are classified by the nature of the study outcome.

Predictive validity effect size classifications for adult pre-conviction outcomes are presented in Table 21. These outcomes fall into two broad categories: assessments that predict failure to appear (FTA) and assessments designed to predict future crime. Prediction of future crime contains two outcomes: pre-conviction crime (PCC) prediction and pre-conviction risk determination (PCRD). The placement of assessments used to determine pre-conviction risk determination into the pre-conviction table was based on when the assessments are typically administered; while assessments in this section tend to be administered pre-conviction, they should not be ruled out as viable post-conviction assessments of potential recidivism solely on that basis (many were used to predict long-term recidivism in the same way as post-conviction assessments discussed later in this report).

Table 21 provides the number of studies found and reviewed for each assessment (k_r), the number with the relevant outcome (k) and the overall classification rating based on the rules established in the methods section of this report. Where available, ratings for males and females are provided separately. Instruments lacking a male or female row had no studies specific to either gender, and classifications represent the combined classification for males and females[11]. Instruments lacking classification ratings in the primary row for that instrument either included only studies involving one gender (but not the other), or reported outcomes by gender only (and not overall). When columns in the table are

---

[11] It is important to note that the primary row, which indicates males and females were combined, generally heavily represents males, who are disproportionately more common in the criminal justice system. It may not be accurate to interpret the ratings from the primary row as applying equally to both males and females.

combined across the FTA and PCC/PCRD outcomes, it indicates only a combined FTA and PCC predictive validity metric was provided in the relevant study or studies.

As seen in Table 21, most instruments in these classifications have been evaluated for predictive validity only once, raising questions about generalizing the findings to other populations/jurisdictions. Also, the general pattern for predictive validities observed in Table 21 is one of inconsistency. For example, the Inventory of Need Pretrial Screening Tool (INPST) showed excellent predictive validity for FTA, and good predictive validity for PCC/PCRD. However, while it showed excellent predictive validity for females on FTA, it was rated only fair for females on PCC/PCRD. Only one other assessment received an excellent rating for FTA, the New York City Criminal Justice Agency Pretrial Release Recommendation Form (CJA-PRRS). Unfortunately, no separation by gender was reported, and the instrument has not been used in the extant literature to assess PCC/PCRD (which is also of interest when deciding whether to release an individual prior to a trial). The Ohio Risk Assessment System Pretrial Assessment Tool (ORAS-PAT) and the Salt Lake Pretrial Release Instrument (SLPRI) were both rated in the good to excellent range across FTA and PCC/PCRD, but neither has been evaluated by gender. Notably, no tool scored excellent on both FTA and PCC/PCRD. The Risk and Needs Triage (RANT) and Proxy performed most poorly overall; however, short-term recidivism was not an intended use for the Proxy when it was developed (the single study using the Proxy for this purpose was exploratory).

Table 21: Predictive Validity (PV) for Adult Pre-Conviction Outcomes: Failure to Appear (FTA) and Pre-Conviction Crime (Short-Term)/Pre Conviction Risk Determination (Long-Term) (PCC/PCRD) by Instrument

Instrument	FTA			PCC/PCRD[a]	
	k_r	k	Predictive Validity	k	Predictive Validity
CJA-PRRS	4	2	Excellent	--	--
HC-PRA	2	1	Good	1	Good
INPST	1	1	Excellent	1	Good
Male		1	Good	1	Good
Female		1	Excellent	1	Fair
KPRA	2	2	Fair-Good	2	Fair-Good
ORAS-PAT	2	2	Good	2	Good-Excellent
Polk	1	1	Poor[b]		
Proxy[c]	1	--	--	1	Poor
PTRA	3	3	Good		
RANT	1	--	--	1	Fair
SLPRI	1	1	Good	1	Excellent
SPI	1	--	--	--	--
Male		--	--	1	Good

Notes: k_r=total number of studies on assessment found/reviewed (this number may be greater than the number with a PV measure); k=number of studies with a PV measure for the outcome; instruments lacking a male or female row had no studies specific to either gender, and classifications represent the combined classification for males and females; instruments lacking classification ratings in the primary row for that instrument either included only studies involving one gender (but not the other), or reported outcomes by gender only (and not overall); when columns are combined across the FTA and PCC/PCRD outcomes, it indicates only a combined FTA and PSC predictive validity metric was provided.
[a] Determination of predictive validity correlational effects sizes were made using correction-based standards for "poor," "fair," "good," and "excellent" based on the different base rates for the short- and long-term outcomes (see Appendix A, Table A-1).
[b] Though a PV value was given for FTA only, it was based on just three total cases, and conclusions should not be drawn from such a small sample; it is also the case that the combined outcome shown in the table largely reflects ability to predict PSC, as only three FTA cases existed in the sample.
[c] Neither FTA nor PSC are intended uses for this tool; UCJC was asked to evaluate the tool for this purpose, but its poor performance in a domain for which it was not developed is not evidence against the tool for more appropriate uses (see post-sentence recidivism outcomes).

Predictive validity effect size classifications for adult post-conviction outcomes for general recidivism (including arrest and conviction), violent recidivism, and technical violations are presented in Table 22. Notably, the Minnesota Screening Tool Assessing Recidivism Risk (MnSTARR) and the Post-Conviction Risk Assessment (PCRA) were the only tools to receive excellent ratings in predicting both general and violent recidivism. For the MnSTARR, the finding is tempered somewhat by the fact that only one validation study has examined the tool. The PCRA, however, received an excellent rating across three studies of its efficacy predicting general recidivism. Although only one study examined the tool by gender, the PCRA also received excellent classifications for both males and females in prediction of general recidivism (no studies examined its utility in prediction of violent recidivism by gender).

A general finding observed in Table 22 was that, the more a tool had been studied, the more variability was found in its predictive validity classification. Therefore, tools that

performed particularly well across numerous validation studies warrant closer examination. The Self-Appraisal Questionnaire (SAQ) performed well across all outcomes in the table, but it lacked validation of violence and technical violation prediction for females, and only one study existed that examined its accuracy with a female population for general recidivism. The Statistical Information on Recidivism-Revision One (SIR-R1) tool performed exceptionally well for prediction of general and violent recidivism, but only one study tested these outcomes for females; like the SAQ, the majority of its evaluations focused on only males. While the Static Risk Assessment (SRA) performed excellent overall, it is interesting that the only study that split males and females found only good prediction for both (which is nevertheless on the higher end of all tools evaluated).

Other tools showed potential in specific domains, or in the specific domains for which they were designed. While the Lifestyle Criminality Screening Form (LCSF) was not remarkable for general recidivism, it was rated as good for violent recidivism, and excellent for technical violations, in two studies examining its efficacy with male offenders. The Offender Group Reconviction Scale (OGRS) showed promise for general and violent recidivism among males only. The Paper-Pencil Screen for Major Violence (PMAV) and Minor Violence (PMIV) both showed excellent violence prediction for males (particularly with serious male offenders) and good violence prediction for females.

Of note is the relative dearth of evaluations of these brief assessments with female offenders. Even assessments with numerous evaluations (such as the SAQ or SIR-R1 with 16 and 10 studies, respectively) were rarely examined in the context of female offenders (once each for these two assessments). No assessments indicated they were strictly for male offenders, but, without exception, further validation of assessments with females needs to be considered. Only the Offender Risk and Management System (ORAMS) and the Correctional Offender Management Profile for Alternative Sanctions – Recidivism Scale (COMPAS-R) were evaluated more than once with female offenders. While neither received an excellent rating on any outcome, they are notable in terms of predicting general recidivism in the fact that they: 1) received a good rating overall, 2) received a good rating for both males and females, and 3) have both been evaluated at least twice with a specifically female sample.

The Virginia Nonviolent Risk Assessment Instrument (NORAI) is not included in Table 22. Two reviewed studies examined the tool, but only one reported a predictive validity effect size. As mentioned earlier, this one study reported an odds ratio that could not be translated into a common language effect size by reconstructing the dataset. Accordingly, the outcome cannot be compared to other instruments in terms of predictive accuracy, but the reader should note that the odds ratio was significant in the tool's prediction of general recidivism.

Table 22: Predictive Validity (PV) for Adult Post-Conviction Outcomes: General Recidivism (Arrest, Conviction), Violent Recidivism, and Technical Violations by Instrument

Instrument	k_r	General Recidivism		Violent Recidivism		Technical Violations	
		k	Predictive Validity	k	Predictive Validity	k	Predictive Validity
COMPAS-R	8	4	Good	1	Fair	--	--
Male		*4*	*Good*	*1*	*Fair*	--	--
Female		*2*	*Good*	*1*	*Fair*	--	--
COMPAS-V	3	1	Good	1	Fair	--	--
Male		*1*	*Good*	*2*	*Fair-Good*	--	--
Female		*1*	*Good*	*1*	*Fair*	--	--
LCSF	10	--	--	--	--	--	--
Male		*2*	*Fair-Good*	*2*	*Good*	*2*	*Excellent*
LSI-R:SV	5	2	Good	1	Fair	--	--
Male		*1*	*Fair*	--	--	--	--
Female		*1*	*Fair*	--	--		--
MnSTARR	1	--	--	--	--	--	--
Male		*1*	*Excellent*	*1*	*Excellent*	--	--
Female		*1*	*Excellent*	*1*	*Excellent*	--	--
OC-RAI	1	1	Good	--	--	--	--
OGRS	9	2	Excellent	3	Good	--	--
Male		*2*	*Good-Excellent*	*3*	*Excellent*	--	--
Female		*1*	*Good*	*2*	*Fair*	--	--
ORAMS	2	2	Good	--	--	--	--
Male		*2*	*Good*	--	--	--	--
Female		*2*	*Good*	--	--	--	--
ORAS	1	--	--	--	--	--	--
Male		*1*	*Good*	--	--	--	--
Female		*1*	*Excellent*	--	--	--	--
PCRA	3	3	Excellent	1	Excellent	--	--
Male		*1*	*Excellent*	--	--	--	--
Female		*1*	*Excellent*	--	--	--	--
PMAV	1	--	--	--	--	--	--
Male[a]		--	--	*1*	*Excellent*	--	--
Female[a]		--	--	--	--	--	--
Male[b]		--	--	*1*	*Good*	--	--
Female[b]		--	--	*1*	*Good*	--	--
PMIV	1	--	--	--	--	--	--
Male[a]		--	--	*1*	*Excellent*	--	--
Female[a]		--	--	*1*	*Good*	--	--
Male[b]		--	--	*1*	*Good*	--	--
Female[b]		--	--	*1*	*Good*	--	--
Proxy	3	1	Good	1	Good	--	--
SAQ	16	--	--	-	--	--	--
Male		*5*	*Excellent*	*5*	*Good*	*1*	*Excellent*
Female		*1*	*Good*	--	--	--	--
SIR-R1	10	--	--	--	--	--	--
Male		*9*	*Excellent*	*6*	*Good*	--	--
Female		*1*	*Excellent*	*1*	*Excellent*	--	--
SPIn	1	1	Excellent	--	--	--	--
Male		*1*	*Excellent*	--	--	--	--
Female		*1*	*Excellent*	--	--	--	--

Instrument	k_r	General Recidivism		Violent Recidivism		Technical Violations	
		k	Predictive Validity	k	Predictive Validity	k	Predictive Validity
SRA[c]	4	3	Excellent	2	Excellent	--	--
Male		*1*	*Good*	*1*	*Good*	--	--
Female		*1*	*Good*	*1*	*Good*	--	--
WRA	5	2	Fair-Good	--	--	1	Good
Male		*1*	*Fair*	--	--	--	--

Notes: k_r=total number of studies on assessment found/reviewed (this number may be greater than the number with a PV measure); k=number of studies with a PV measure for the outcome; instruments lacking a male or female row had no studies specific to either gender, and classifications represent the combined classification for males and females; instruments lacking classification ratings in the primary row for that instrument either included only studies involving one gender (but not the other), or reported outcomes by gender only (and not overall).
[a] Serious offender population
[b] General offender population
[c] All SRA outcomes are for felony recidivism only

Predictive validity effect size classifications for adult post-conviction outcomes for intimate partner violence (IPV) and prediction of general recidivism among IPV offenders (including arrest and conviction) are presented in Table 23. Notably, none of the IPV tools was exceptional at predicting general reoffending. For the outcome of general recidivism, the Domestic Violence Screening Instrument-Revised (DVSI-R) is notable for receiving a good classification for both males and females (although only one study with females was found).

The DVSI-R also is notable for receiving good scores for prediction of both male and female IPV offending. Comparing these outcomes to the original DVSI, the revised version is a notable improvement. The best performing tool for predicting new IPV offenses among males was the Brief Spousal Assault Form for the Evaluation or Risk (B-SAFER). The Ontario Domestic Assault Risk Assessment (ODARA) performed best at predicting new female IPV offenses and also did quite well predicting new male IPV offenses.

Table 23: Predictive Validity (PV) for Adult Outcomes: Intimate Partner Violence (IPV) and General Recidivism (Arrest, Conviction) of IPV Offenders by Instrument

Instrument	k_r	IPV[a]		IPV Offender General Recidivism	
		k	Predictive Validity	k	Predictive Validity
B-SAFER	2	--	--	--	--
Male		*2*	*Good-Excellent*	--	--
DVSI	5	3	Fair	4	Good
Male		*3*	*Fair*	*2*	*Fair-Good*
Female		*3*	*Poor*	*1*	*Poor*
DVSI-R[b]	5	3	Good	1	Good
Male		*3*	*Good*	*1*	*Good*
Female		*1*	*Good*	*1*	*Good*
DVSR	2	--	--	--	--
Male		*2*	*Fair-Good*	--	--
LS	1	--	--	--	--

Instrument	k_r	k	IPV[a] Predictive Validity	k	IPV Offender General Recidivism Predictive Validity
Male[c]		1	Fair	--	--
ODARA	11	1	Good	--	--
Male		8	*Good*	*3*	*Good*
Female		2	*Good-Excellent*	*1*	*Fair*
PAPS	1	--	--	--	--
Male		*1*	*Good[d]*	*1*	*Good*

Notes: k_r=total number of studies on assessment found/reviewed (this number may be greater than the number with a PV measure); k=number of studies with a PV measure for the outcome; instruments lacking a male or female row had no studies specific to either gender, and classifications represent the combined classification for males and females; instruments lacking classification ratings in the primary row for that instrument either included only studies involving one gender (but not the other), or reported outcomes by gender only (and not overall).
[a] Includes violation of a protective order, threats, and physical injury
[b] Revised version of the DVSI
[c] Though participants in the study were female, the assessment is completed with respect to their male abusers (if same-sex abusers were included in the study, it was not noted by the authors).
[d] No overall DV recidivism PV value was given, so the interpreted effect size is the average of physical, severe, and injury IPV PV values.

Inter-rater reliability classifications. Very few instruments provided outcomes for inter-rater reliability. Those that did, however, generally revealed excellent classified inter-rater reliability (see Table 24). The only exception, the B-SAFER instrument (which is the only included assessment that is characterized as a structure professional judgment (SJP) tool), revealed a good classification.

Table 24: Inter-Rater Reliability (IRR) for Adult Instruments (When Available)

Instrument	k_r	k	IRR
B-SAFER	3	1	Good
LCSF	10	3	Excellent
LSI-R:SV	5	2	Excellent
MnStarr	1	1	Excellent
ODARA	11	8	Excellent
PCRA	3	1	Excellent
SIR-R1	10	4	Excellent

Notes: k_r=total number of studies on assessment/reviewed (this number may be greater than the number with an IRR value); k=number of studies with an IRR value.

Brief Risk Assessment with Juveniles

The Office of Juvenile Justice and Delinquency Prevention (OJJDP) estimates there are 1.5 million juvenile arrests per year in the U.S. (Puzzanchara, 2013) and more than 50,000 youth are held in a residential placement (detention or secure facility) on a given day (OJJDP, 2012). As with adult offenders, best practice dictates that decisions regarding placement, supervision, and treatment of delinquent youth are informed by risk

assessment at all transition points (Steinhart, 2006; Vincent, Guy, & Grisso, 2012). When compared to the criminal justice system, the juvenile justice system is relatively more concerned with youth's developmental stage as a factor in delinquent behavior; given that most individuals will "age" out of such behavior, even with no intervention, the juvenile justice system strives to minimize the negative impact of detention and out-of-home placements on an individual's life-course trajectory. In addition to the static and dynamic factors discussed earlier, juvenile risk assessments are likely to include items related to: parental control; parental criminality; gang involvement; and conduct problems in school or other settings that have not been referred to law enforcement.

The primary decision points within the juvenile justice system are described below (adapted from Vincent et al., 2012, p. 66):

- Diversion: the decision whether a youth will be formally processed in the juvenile court system, which could happen at probation intake or the prosecutor's office.

- Pretrial Detention: the decision whether a youth will be released or held in a locked facility awaiting adjudication. The main outcomes of concern at this point are failure to appear (FTA) and whether the youth poses a risk of harm to him or herself or others.

- Adjudication: the decision, made by the court, whether the youth has violated the law and committed a delinquent act. Models for Change, developed by the MacArthur Foundation, notes that this is the only decision that should not incorporate risk assessment tools.

- Corrections and Post-Disposition: the decision of when, where, and how the youth should re-enter the community, typically made by the courts or facilities.

A comparison of youth risk screening instruments is presented in the following sections, organized by pre- and post-adjudication outcomes. Of note, the instruments are classified according to assessed outcomes rather than intended outcomes. For example, instruments that are evaluated for predictive validity in the two years after adjudication will be discussed in the post-adjudication section, even if originally intended as a tool to inform pre-trial decision-making. Three instruments (CJRA, Missouri, and Y-ARAT-FO) are included in both sections; an overview of each is presented in the pre-adjudication section but is not repeated in the post-adjudication section.

  Instruments with pre-adjudication outcomes. At the pre-adjudication stage of juvenile justice court processing (e.g., diversion and pre-trial detention decisions), the primary barriers to risk assessment are: 1) the high volume of cases, which necessitates the use of brief instruments that can be feasibly administered to all youth referred to the court; and 2) concerns regarding self-incrimination among youth who have not been convicted or even charged with an offense (Administrative Office of the Courts, Judicial Council of California, 2011). Given such concerns, most assessment instruments used to make pre-

adjudication decisions are jurisdiction-specific brief screens comprised of relatively few items that are static and objective in nature (Steinhart, 2006). Included pre-adjudication instruments (N=13) are presented in Table 25.

Table 25: Instruments with Pre-adjudication Outcomes

Name	Acronym[a]	Screening Version	Related to	Date range of studies
Abscond Risk Instrument	ARI		CAPFA	2012
Alameda County Placement Risk Assessment	(Alameda RAI)			2000
Arkansas Division of Youth Services Risk Assessment Instrument	(Arkansas RAI)			2001
Colorado Juvenile Risk Assessment Pre Screen	CJRA	X		2013-14
Indiana Department of Corrections Risk Assessment Instrument	Indiana RAI		WJPARI[b]	1998
Iowa Juvenile Court Intake Risk Assessment[c]	(Iowa Court RA)			1996-99
Kansas Juvenile Detention Risk Assessment	Kansas JDRA			2003
Minnesota Juvenile Courtroom Risk Assessment	Minnesota JCRAI		JDC RAI[d]	2013
Missouri Juvenile Risk Assessment	Missouri JRA			2002-11
Montana Pre-Adjudicatory Risk Assessment Instrument	Montana RAI		VDAI[e]	2012
Ohio Youth Assessment System-Detention	OYAS-Detention		OYAS	2009-13
Ohio Youth Assessment System-Diversion	OYAS-Diversion		OYAS[f]	2009-13
Youth Actuarial Risk Assessment Tool for First-Time Offending	Y-ARAT-FO		Y-ARAT[g]	2014

[a] Acronyms/names in parentheses were added by authors of the current report for purposes of referring to instruments more parsimoniously in subsequent tables; they are not the acronyms of the researchers who developed the instruments.
[b] Wisconsin Juvenile Probation Aftercare & Risk Instrument
[c] The same instrument is used in sentencing decisions and called the Dispositional Risk Assessment
[d] Minnesota Juvenile Detention Center Risk Assessment
[e] Virginia Detention Assessment Instrument
[f] Ohio Youth Assessment System
[g] Youth Actuarial Risk Assessment Tool

With the exception of the Y-ARAT-FO, the instruments presented in Table 26 are intended for use with youth at the beginning of contact with the juvenile justice system, and are used to predict behavior between entering the system and case resolution (whether by adjudication, dismissal, or diversion). The Y-ARAT-FO is unique in that it was designed to identify youth who have been involved with law enforcement in a non-offender capacity (e.g., as victim or witness) in order to provide services to prevent future involvement. While the instrument does predict future delinquent acts, it is not predicting recidivism, per se, because the youth have not committed any known offenses at the time of the assessment. For the majority of instruments presented in Table 26, risk category recommendations fall broadly into the following options: release; release with conditions; and detain.

<p style="text-align:center">**Table 26: Instruments with Pre-adjudication Outcomes: Overview**</p>

Instrument	Intended Population	Intended Outcomes	Intended Decisions	Risk Categories
Alameda RAI	Juvenile Delinquents on Community Supervision	Recidivism	Supervision	3
ARI	Juvenile Delinquents on Community Supervision	Absconding	Supervision	2
Arkansas RAI	Juvenile Delinquents	Recidivism	Detention	3
CJRA	Juvenile Delinquents	Recidivism, General Recidivism, Felony	Placement Case Planning	3
Indiana RAI	Juvenile Delinquents	Probation Failure	Placement	3
Iowa Court RA	Juvenile Delinquents	Recidivism	Disposition Placement	4 (Boys) 5 (Girls)
Kansas JDRA	Juvenile Delinquents	Recidivism Technical Violations	Detention Conditions of Release	3
Minnesota JCRAI	High Risk Juvenile Delinquents	Pre-adjudication Failure	Detention Conditions of Release	3
Missouri JRA	Juvenile Delinquents	Recidivism	Release	3
Montana RAI	Juvenile Delinquents	Pre-adjudication Failure	Detention Conditions of Release	3
OYAS-Detention	Juvenile Delinquents	Pre-adjudication Failure	Detention Treatment Needs	6
OYAS-Diversion	Juvenile Delinquents	Pre-adjudication Failure	Eligibility for Diversion Treatment Needs	3
Y-ARAT-FO	Non-delinquent Juveniles[a]	Delinquent Behaviors	Referral to Services	10

[a] Who have been involved in an offense in a non-offender role such as victim or witness

As noted in Table 27, pre-adjudication screening is intended to be conducted by intake staff, whether that is in probation, the courts, or at a detention facility. Fewer than half of the instruments require formal training, although training is likely included in job-specific requirements for each agency. The relative lack of formal training curricula stems from the fact that many pre-adjudication instruments were developed for use in a local jurisdiction, and resources were not devoted for training beyond those confines. Approximately half of the included instruments require an interview with the delinquent youth, while the others rely on administrative data collected by the agency.

<p style="text-align:center">**Table 27: Instruments with Pre-adjudication Outcomes: Administrative Requirements**</p>

Instrument	Intended Assessor	Training/ Certification	Time	Interview Required	Self-report	Field-tested
Alameda ARI	Probation Officer					X
ARI	Probation Officer[a]	X[b]				
Arkansas ARI	Detention Staff	X		X		
CJRA		X[c]	60	X		
Indiana RAI	Detention Staff			X		X

Instrument	Intended Assessor	Training/ Certification	Time	Interview Required	Self-report	Field- tested
Iowa Court RA	Court Officers					X
Kansas JDRA	Intake & Assessment Center					X
Minnesota JCRAI	Detention staff					X
Missouri JRA	Probation officer			X		X
Montana RAI	Probation officer			X		X
OYAS-Detention	Detention Staff	X[d]	48	X	X	X
OYAS-Diversion	Detention Staff	X[d]	48	X	X	X
Y-ARAT-FO	Police officer					X

[a] Assessment is completed via file review from items collected during CAPFA assessment, which does require an interview
[b] CAPFA training
[c] 16 hours initially with annual re-certification that is 4 hours
[d] Two day training, with competence testing

The vast majority of instruments were comprised of less than ten items (Table 28) and only the Colorado Juvenile Risk Assessment Pre-screen (CJRA) produced a score for multiple scales. With the exception of the Montana and Minnesota instruments, all the screens included at least one dynamic factor. While researchers also coded for protective factors, they were only included on three instruments (CJRA, ARI, and Kansas JDRA).

Table 28: Instruments with Pre-adjudication Outcomes: Items

Instrument	Items	Scales	Static Factors	Dynamic Factors	Objective Items	Subjective Items
Alameda ARI	8		3	5	3	5
ARI	12		3	7	8	4
Arkansas ARI	10		7	3	7	3
CJRA	32	2[a]	17	11	18	14
Indiana RAI	8		4	4	4	4
Iowa Court RA	6		4	2	4	2
Kansas JDRA	4[b]					
Minnesota JCRAI	5		5		5	--
Missouri JRA	10		7	3	5	5
Montana RAI	7		7		7	--
OYAS-Detention	6		4	2	4	2
OYAS-Diversion	6		5	1	5	1
Y-ARAT-FO	5		3	1	5	--

[a] Criminal History and Social History
[b] Two of the items are actually scales (Aggravating Factors and Mitigating Factors), for which the exact number of items could not be identified

All instruments assessed for history of anti-social behavior, which included conduct problems in school and other contexts, even if the incidents had not been handled within the juvenile justice system. Problems with school and substance use were included on approximately half of the instruments. Problems with peers, usually operationalized as gang involvement or commission of delinquent acts with other juveniles, were assessed for in the majority of instruments. Family problems were typically operationalized as history of family criminality, child abuse and neglect complaints, and parenting skill deficits.

Personality problems were identified in only three instruments, likely because these are complex constructs that require in-depth information and subjective decision-making.

Table 29: Instruments with Pre-adjudication Outcomes: Domains

Instrument	Attitudes	Anti-social behavior	Personality	Peers	Family	School	Substance Use
Alameda ARI		X		X	X	X	X
ARI		X	X	X			X
Arkansas ARI		X		X		X	
CJRA	X	X	X	X	X	X	X
Indiana RAI		X		X		X	X
Iowa Court RA		X		X			X
Kansas JDRA[a]		X					
Minnesota JCRAI		X					
Missouri JRA		X		X	X	X	X
Montana RAI		X					
OYAS-Detention	X	X	X				
OYAS-Diversion		X			X		
Y-ARAT-FO		X		X			
[a] Presence of items in other domains unknown							

Instruments with post-adjudication outcomes. Instruments evaluated for post-adjudication outcomes are typically used to inform decisions related to placement (community or residential setting), supervision, and treatment. Ideally, all youth would have been assessed by this point in the juvenile justice process; if limited resources make that impossible, however, brief screening tools can provide a cost-effective means to classify youth into two broad categories: 1) youth who can receive the least intensive intervention without compromising public safety; and, 2) youth who require additional assessment to make decisions regarding placement and services. Table 30 describes instruments identified for the current study that have been evaluated for post-adjudication outcomes (e.g., delinquent behavior that occurs after adjudication; N=18). Of note, three instruments discussed in the previous section (CJRA, Missouri JRA, and Y-ARAT-FO) were also evaluated for post-adjudication outcomes but are not listed in Table 30. Information on the predictive validity of those instruments in both contexts will be presented later in the report; however, this section will not repeat information related to instrument characteristics.

When compared to the pre-adjudication instruments, the post-adjudication assessments tend to be older, as evidenced by the date range of included studies, and are designed for use in a range of jurisdictional and geographical settings. Three of the instruments in Table 30 are closely related to each other (the PACT, YASI, and WSJCA).

Table 30: Instruments with Post-Adjudication Outcomes

Name	Acronym[a]	Screening Version	Related to	Versions	Date range	Locale
Arizona Risk/Needs Assessment Instrument	ARNA		NCCD Scale Score[b]	3	1990-2013	1
El Paso Risk Assessment of	El Paso				2011-14	1

Name	Acronym[a]	Screening Version	Related to	Versions	Date range	Locale
Juveniles at Intake	RAJI					
Girls Link Risk Assessment	(Girls Link)		JAIS[c]		2013	1
Joint Risk Matrix			NCAR		2007	1
Juvenile Sanctions Center Risk Assessment	JSC		JAIS[c]		2013	1
North Carolina Assessment of Risk	NCAR				2004-07	1
Ohio Youth Assessment System-Disposition	OYAS-Disposition		OYAS		2009-13	1
Ohio Youth Assessment System-Residential	OYAS-Residential		OYAS		2009-13	1
Ohio Youth Assessment System-Reentry	OYAS-Reentry		OYAS		2009-13	1
Orange County Risk Assessment Instrument	(OC-RAI)				1990-2003	2
Oregon Youth Authority Recidivism Risk Assessment	OYA-ORRA				2011	1
Oregon Youth Authority Violent Crime	OYA-ORRA-V				2011	1
Positive Achievement Change Tool, Pre-Screen	PACT-PS		YASI, WSJCA, Back on Track!		2009-14	3
Washington State Juvenile Court Assessment, Pre Screen	WSJCA-PS		YASI		1998-2014	2
Wisconsin Juvenile Probation & Aftercare Risk Instrument	WJPARI				1988	1
Youth Actuarial Risk Assessment Tool	Y-ARAT				2013	1
Youth Assessment and Screening Instrument, Pre-Screen	YASI-PS		WSJCA, PACT	3[d]	2007-11	3
Youth Level of Service/Case Management Inventory: Screening Version	YLS/CMI: SV	X			2004-14	3

[a] Acronyms/names in parentheses were added by authors of the current report for purposes of referring to instruments more parsimoniously in subsequent tables; they are not the acronyms of the researchers who developed the instruments.
[b] National Council on Crime & Delinquency; Post Adjudication Score Instruments
[c] Juvenile Assessment & Intervention System
[d] At least three versions were identified in the studies reviewed for this report

While the majority of the instruments were designed for use with all juvenile delinquents, two are specifically used with first-time offenders (El Paso RAJI and Orange County). The Arizona Risk Need Assessment (ARNA) has three versions, depending on the number of prior referrals a youth has at the time of assessment. All but one of the instruments is designed to predict general recidivism, while the Oregon Youth Authority Violent Crime (ORRA-V) was designed specifically to predict violent recidivism. All but two of the instruments (ORRA and ORRA-V) classify youth by risk category; typical decisions informed by this type of instrument include supervision level, placement, and the need for

a full assessment. Both the PACT-PS and the Y-ARAT can be used to make decisions with pre-adjudicated youth (but are included in this section because they have only been evaluated for post-adjudication outcomes).

Table 31: Instruments with Post-Adjudication Outcomes: Overview

Instrument	Intended Population	Intended Outcomes	Intended Decisions	Risk Categories
ARNA	Juvenile Delinquents	Recidivism	Supervision	3[a]
El Paso RAJI	First-time Juvenile Delinquents in Diversion Program	Recidivism	Supervision	5[b]
Girls Link	Female Juvenile Delinquents	Recidivism	Supervision	3
Joint Risk Matrix	Juvenile Delinquents	Recidivism[c]	Placement Supervision	3 Static 3 Dynamic
JSC	Male Juvenile Delinquents	Recidivism	Supervision	3[d]
NCAR	Juvenile Delinquents	Recidivism	Placement Supervision	3
OC-RAI	First-time Juvenile Delinquents on Probation	Recidivism	Treatment	3
OYAS-Disposition	Juvenile Delinquents		Treatment[e]	3
OYAS-Residential	Juvenile Delinquents	Post-adjudication services	Placement[f] Services	3
OYAS-Reentry	Juvenile Delinquents	Post-adjudication services	Treatment[g]	3
OYA-ORRA	Juvenile Delinquents on Probation	Felony Recidivism	Supervision	None[h]
OYA-ORRA-V	Juvenile Delinquents on Probation	Violent Recidivism	Supervision	None[h]
PACT-PS	Juvenile Delinquents	Recidivism	Supervision Detention Full Assessment Referrals	4
WJPARI	Juvenile Parolees		Supervision	
WSJCA-PS	Juvenile Delinquents		Full Assessment	3[i]
Y-ARAT	Juvenile Delinquents	Recidivism	Diversion	5
YASI-PS	Juvenile Delinquents	Recidivism	Full Assessment	4[k]
YLS/CMI: SV	Juvenile Delinquents	Recidivism	Full Assessment	3

[a] Cut points for risk classification differ by number of previous referrals
[b] First version has 6 risk categories.
[c] Non-status offense
[d] Revised version
[e] Treatment prior to adjudication
[f] Length of stay recommendation
[g] Aftercare services
[h] Does not classify into risk category, but provides 0 to 100 probability-risk prediction
[i] Risk category combines criminal and social history scores
[j] Different scores by gender

The majority of instruments require an interview (Table 32), which is more than the pre-adjudication tools, reflecting less concern with self-incriminating statements at this stage of case processing. Many of the instruments require an interview with the parents/family of a youth as well. None of the instruments were self-report. All of the instruments were designed for use at intake, which could be into the juvenile justice system (court), probation/parole, or residential care. The PACT-PS can also be used to make detention decisions with pre-adjudicated youth.

Table 32: Instruments with Post-Adjudication Outcomes: Administrative Requirements

Instrument	Intended Assessor	Training/ Certification	Time	Interview Required	Field-tested
ARNA	Probation Officer		29	X	X
El Paso RAJI	Probation Officer[a]			X	X
Girls Link	Probation Officer	X	54	X	X
Joint Risk Matrix	Court Intake Staff	X		X	X
JSC	Probation Officer	X	54	X	X
NCAR	Court Intake Staff			X	X
OC-RAI	Probation Officer			X	X
OYAS-Disposition	Court Intake Staff	X[c]	48[d]	X[e]	X
OYAS-Residential	Secure Care Staff	X[c]	48[d]	X[e]	X
OYAS-Reentry	Probation Officer	X[c]	48[d]	X[e]	X
OYA-ORRA	Probation Officer				X
OYA-ORRA-V	Probation Officer				X
PACT-PS	Assessment Center Staff Detention Intake Staff Police officer	X[b]	25	X	X
WJPARI	Parole officer				X
WSJCA-PS	Court Intake Staff Probation Officer	X		X	X
Y-ARAT	Police officer				X
YASI-PS	Probation Officer Secure Care Staff	X[b]	97[f]	X	X
YLS/CMI: SV	Varies[g]	X[h]		X	X

[a] Case worker in a diversion program
[b] Two-day training
[d] For full assessment
[e] Also a self-report survey
[f] For full assessment
[g] Developed for use in school settings and with all practitioners involved in youth assessment, including law enforcement. In validation studies, assessor was law enforcement (k=1) and researchers (k=1)
[h] Three-day training

When compared to instruments with pre-adjudication outcomes, these tools are longer and comprised of relatively more dynamic items (Table 33). The inclusion of youth and family interviews allows for more insight into dynamic factors that should be addressed during placement or supervision. As noted earlier, such factors are also more subjective and therefore require more resources in terms of staff training, time, and monitoring.

Table 33: Post-adjudication Instruments: Items

Instrument	Items	Scales	Static Factors	Dynamic Factors	Objective Items	Subjective Items
ARNA	10[a]		3	5	5	5
El Paso RAJI	21		7	10	11	8
Girls Link	9		5	4	6	3
Joint Risk Matrix	14	2[b]	2	8	3	11
JSC	9		3	5	5	4
NCAR	9		4	5	5	4
OC-RAI	10		4	6	5	5
OYAS-Disposition	32	7[c]	9	21	11	22
OYAS-Residential	33	7[c]	9	23	10	23
OYAS-Reentry	42	7[c]	13	27	16	26
OYA-ORRA	12[d]		9	2	12	--
OYA-ORRA-V	9[e]		9	--	9	--
PACT-PS	46[f]	4[g]	14	10	14	13
WJPARI	8		3	5	3	5
WSJCA-PS	23[h]	3[i]	13	6	14	8
Y-ARAT	10		9[j]	--	1	--
YASI-PS	42[k]	2[l]	13	21		
YLS/CMI: SV	8		1	7	1	7

[a] 5 items in first-time offenders, 5 in second-time (with 3 shared), and 6 items in third-time (with 3 shared)
[b] Static, dynamic
[c] History; Family & Accommodations; Peers; Education & Employment; Pro-social Skills; Substance Abuse, Mental Health & Personality; Values, Beliefs, & Attitudes
[d] 12 items plus 3 interactions (prior criminal mischief & total prior misdemeanor referrals; prior criminal mischief and total prior thefts; prior criminal mischief & total prior runaway)
[e] 9 items plus 2 interactions (prior weapon offense & prior felony theft referrals; prior misdemeanor referrals & total prior felony referrals)
[f] Not all used in scoring; unclear which items contribute to risk score
[g] Criminal History, Social History, Mental Health, Attitudes and Behaviors
[h] Pre-screen assessment has 27 items, but does not score attitudes/behaviors (n=5) for the pre-screen score
[i] Criminal History, Social History, Attitudes and Behaviors (only first two contribute to risk classification for pre-screen)
[j] One additional item asks about youth's involvement in an incident in a non-offending role, which was classified as "other" rather than static
[k] Earlier versions had 28 and 32 items
[l] Legal History and Social History

Even when intended as a brief screen, the post-adjudication instruments are comprised of a wide range of domain items. All the instruments assessed for anti-social behavior, which includes delinquency history and conduct problems in other settings. The vast majority also assessed for academic performance, substance use, and family problems. The comparative rarity of attitudes, mental health and personality problems as assessment targets reflects the difficulty of identifying any of those items, which are complex constructs, using a brief instrument and/or staff with non-clinical backgrounds. Of note, the OYAS considers mental health as a factor at disposition but not for the residential and reentry instruments.

Table 34: Post-adjudication Assessment Instruments: Content Domains

Instrument	Attitudes	Anti-social behavior	Personality	Peers	Family	School	Substance Use	Mental Health
ARNA		X		X	X	X	X	X
El Paso RAJI		X		X	X	X	X	
Girls Link		X		X	x	X	X	
Joint Risk Matrix		X	X	X	X	X	X	X
JSC		X		X	X	X	X	
NCAR		X		X	X	X	X	
OC-RAI		X		X	X	X	X	
OYAS-Disposition	X	X	X	X	X	X	X	X
OYAS-Residential	X	X	X	X	X	X	X	
OYAS-Reentry	X	X	X	X	X	X	X	
OYA-ORRA		X					X	
OYA-ORRA-V		X						
PACT-PS		X		X	X	X	X	X
WJPARI		X		X	X	X	X	
WSJCA-PS	X[a]	X		X	X	X	X	X
Y-ARAT		X						
YASI-PS	X	X		X	X	X	X	X
YLS/CMI: SV	X	X	X	X	X	X	X	

[a] Included but do not contribute to score

Predictive validity. Predictive validity effect size classifications for juvenile pre-adjudication outcomes are presented in Table 35. These outcomes include failure to appear (FTA), pre-adjudication delinquency (short-term)/pre-adjudication risk determination (long-term) (PAD/PARD), and risk to abscond (RTA).

Only one instrument, the Abscond Risk Instrument (ARI) evaluated risk of juveniles absconding while under supervision. The ARI received an excellent classification on this outcome (which was the only outcome for which it is intended). Similarly, only one instrument was designed specifically to assess juvenile FTA. The Minnesota Juvenile Court Risk Assessment Instrument (Minnesota) received an excellent rating on this outcome. However, both the ARI and Minnesota tool had only one predictive validity study.

Several instruments were examined with respect to PAD/PARD. The Iowa Court Risk Assessment received an excellent overall classification in two studies of its predictive validity, and a good-excellent rating for males and females. The seemingly contradictory outcome occurred because one study found excellent predictive validity for males, but only good for females, while the other found excellent predictive validity for females, but only good for males. The overall rating in both studies was excellent, however.

The Youth Actuarial Risk Assessment Tool for First-Time Offending (Y-ARAT-FO) was the only other tool to receive an excellent rating for PAD/PARD predictive validity (albeit in only one study). This tool's intent is somewhat controversial, however. The tool was developed in the Netherlands as an assessment to be used, by police, when juveniles (who have not themselves committed a delinquent act) are in contact with police in a role other

46

than as a suspect (i.e., as a victim or witness to crime). The tool captures risk due to an ostensibly higher propensity to offend by those who have contact with the criminal justice system, even in non-delinquent roles. While the tool showed excellent prediction, it is not clear from the single validation whether it would also have utility in the U.S. juvenile justice system.

All other juvenile tools received poor, fair, or good predictive validity classifications for PAD/PARD. Among these other tools, the Missouri Juvenile Risk Assessment (JRA) is perhaps most notable in that three validations found the tool to have good predictive validity overall, and two studies found good predictive validity with both males and females. Other tools in the poor, fair, or good range lacked additional validation studies; it is, therefore, unclear whether they would perform better overall if additional studies were available, or whether the single-study ratings are an accurate representation of the tools predictive validities.

Table 35: Predictive Validity (PV) for Juvenile Pre-Adjudication Outcomes: Failure to Appear (FTA), Pre-Adjudication Delinquency (Short-Term)/Pre-Adjudication Risk Determination (Long-Term) (PAD/PARD), and Risk to Abscond (RTA)

Instrument	k_r	k	FTA[a] Predictive Validity	k	PAD/PARD[b] Predictive Validity	k	RTA Predictive Validity
Alameda RAI	1	--	--	1	Good	--	--
Male		--	--	*1*	*Good*	--	--
Female		--	--	*1*	*Good*	--	--
ARI	1	--	--	--	--	1	Excellent
Arkansas RAI	1	--	--	1	Poor	--	--
CJRA	1	--	--	1	Poor	--	--
Indiana RAI	1	--	--	1	Poor	--	--
Iowa Court RA	2	--	--	2	Excellent	--	--
Male		--	--	*2*	*Good-Excellent*	--	--
Female		--	--	*2*	*Good-Excellent*	--	--
Kansas JDRA	1	--	--	1	Fair	--	--
Minnesota JCRAI	1	1	Excellent	1	Good	--	--
Missouri JRA	3	--	--	3	Good	--	--
Male		--	--	*2*	*Good*	--	--
Female		--	--	*2*	*Good*	--	--
Montana RAI	2	--	--	1	Good	--	--
OYAS-Detention	1	--	--	1	Good	--	--
Male		--	--	*1*	*Good*	--	--
Female		--	--	*1*	*Fair*	--	--
OYAS-Diversion	1	--	--	1	Fair	--	--
Male		--	--	*1*	*Fair*	--	--
Female		--	--	*1*	*Fair*	--	--
Y-ARAT-FO[c]	1	--	--	1	Excellent	--	--

Notes: k_r=total number of studies on assessment found/reviewed (this number may be greater than the number with a PV measure); k=number of studies with a PV measure for the outcome; instruments lacking a male or female row had no studies specific to either gender, and classifications represent the combined classification for males and females; instruments lacking classification ratings in the primary row for that instrument either included only studies involving one gender (but not the other), or reported outcomes by gender only (and not overall).
[a] Four additional juvenile FTA assessments were located and examined for PV; however, none of the four offered a PV measure (or a means for analysts to compute one) in any of the reviewed studies. The additional tools, which could theoretically be used for FTA, were from Hennepin County, Minnesota; Multnomah County, Oregon; Virginia; and Montana.
[b] Determination of predictive validity correlational effects sizes were made using correction-based standards for "poor," "fair," "good," and "excellent" based on the different base rates for the short- and long-term outcomes (see Appendix A, Table A-1).
[c] Developed to predict risk for first-time offending (i.e., before an offense has occurred).

Predictive validity effect size classifications for juvenile post-adjudication outcomes are presented in Table 36. These outcomes include general delinquency (new arrest, adjudication) and new violent delinquency. Note that the CJRA, Missouri JRA, and Y-ARAT-FO from Table 35 are also present in Table 36. The Missouri JRA and Y-ARAT-FO are duplicated because they both also inform violent delinquency, while the CJRA is duplicated because it is used as both a pre- and post-adjudication tool. The respective outcomes of all three tools relating to general delinquency are duplicated in the PAD/PARD outcomes of Table 36.

One notable finding from Table 36 is that some of the most often studied tools (including the Arizona Risk/Needs Assessment (ARNA), Positive Achievement Change Tool –

Prescreen (PACT-PS), Washington State Juvenile Court Assessment – Prescreen (WSJCA-PS), and Youth Assessment and Screening Instrument – Prescreen (YASI-PS)) provided some of the poorest predictive validities. None of these instruments received higher than a fair-good classification in prediction of general delinquency overall, and two received only a fair classification.

Turning to some of the less often evaluated instruments, the Oregon Youth Authority Recidivism Risk Assessment (OYA-ORRA) received an excellent rating for general delinquency prediction. The summary classification from two studies indicated the El Paso Risk Assessment of Juveniles at Intake (RAJI) had a good-excellent rating in prediction of general delinquency. Several Ohio Youth Assessment System (OYAS) tools performed well in prediction of delinquency. The disposition tool received a good classification with males, females, and overall. The residential tool received a good classification overall, while the reentry tool received an excellent classification.

Several tools are notable for performance in prediction of female delinquency as well as overall. The female specific Girls Link tool, Joint Risk Matrix, the previously mentioned Missouri JRA, and the North Carolina Assessment of Risk (NCAR) all received good ratings in prediction of female delinquency. The Girls Link male-specific counterpart, the Juvenile Sanctions Center (JSC) Risk Assessment, received a good rating with males, as did the Missouri JRA and the NCAR. The Joint Risk Matrix received an excellent classification for prediction with males and a good overall.

Turning to tools designed to predict violent delinquency, the Oregon Youth Authority Recidivism Risk Assessment – Violence (ORRA-V) received a good rating for both males and females in prediction of violent delinquency. The aforementioned Missouri JRA and YARAT-FO both received good classifications in predicting violence, as did the WSJCA-PS. Aside from the ORRA-V, none of these better-performing tools had validity studies of violence prediction by gender. Of the few other tools with studies evaluating violence predictive validity overall and by gender, the YASI-PS performed the poorest. The Youth Level of Service/Case Management Inventory (YLS/CMI) received a fair-good classification for males, and a good classification for females when predicting violent delinquency. No tools received an excellent classification for prediction of violent recidivism[12].

---

[12] One other tool was reviewed for violence prediction as part of the research project, but is not included in the table because it did not meet the definition of a brief assessment tool. It is worth mentioning, however, that the Child and Adolescent Risk for Violence (CARV) tool received the only excellent classification of any reviewed tools in its single study of predictive validity for violent delinquency. The tool is a modification of the WJPARI, but is most similar to the Indiana RAI, which, by itself, revealed poor predictive validity for Pre-Adjudication Delinquency (Short-Term)/Pre-Adjudication Risk Determination (Long-Term).

Table 36: Predictive Validity (PV) for Juvenile Post-Adjudication Outcomes: General Delinquency (New Arrest, Adjudication) and Violent Delinquency

Instrument	k_r	k	General Delinquency Predictive Validity	k	Violent Delinquency Predictive Validity
ARNA	5	2	Fair-Good	--	--
Male		*2*	*Fair-Good*	--	--
Female		*2*	*Fair-Good*	--	--
CJRA	1	1	Poor	--	--
El Paso RAJI	2	2	Good-Excellent	--	--
Girls Link RA	1	--	--	--	--
Female		*1*	*Good*	--	--
Joint Risk Matrix	1	1	Good	--	--
Male		*1*	*Excellent*	--	--
Female		*1*	*Good*	--	--
JSC	1	--	--	--	--
Male		*1*	*Good*	--	--
Missouri JRA	3	3	Good	1	Good
Male		*2*	*Good*	--	--
Female		*2*	*Good*	--	--
NCAR	3	3	Good	--	--
Male		*3*	*Good*	--	--
Female		*3*	*Good*	--	--
OC-RAI	2	1	Fair	--	--
Male		*1*	*Good*	--	--
Female		*1*	*Poor*	--	--
ORRA	1	1	Excellent[a]	--	--
ORRA-V	1	--	--	1	--[b]
Male		--	--	*1*	*Good[c]*
Female		--	--	*1*	*Good[c]*
OYAS-Disposition	3	2	Good	--	--
Male		*1*	*Good*	--	--
Female		*1*	*Good*	--	--
OYAS-Reentry	2	1	Excellent	--	--
OYAS-Residential	1	1	Good	--	--
PACT-PS	8	6	Fair	--	--[d]
Male		*4*	*Fair*	--	--
Female		*4*	*Fair*	--	--
WJPARI	1	1	Poor	--	--
WSJCA-PS	4	2	Fair-Good	1	Good
Male		*2*	*Fair-Good*	--	--
Female		*2*	*Good*	--	--
YARAT	1	--	--[e]	1	Good
Male		*1*	*Good*	--	--
Female		*1*	*Good*	--	--
YARAT-FO[f]	1	1	Excellent	1	Good
YASI-PS	6	3	Fair	1	Fair
Male		*3*	*Fair*	*1*	*Fair*
Female		*3*	*Fair*	*1*	*Fair*
YLS/CMI:SV	3	2	Good	1	Good
Male		*2*	*Good*	*2*	*Fair-Good*
Female		*2*	*Fair-Good*	*1*	*Good*

Notes: k_r=total number of studies on assessment found/reviewed (this number may be greater than the number with a PV measure); k=number of studies with a PV measure for the outcome;

instruments lacking a male or female row had no studies specific to either gender, and classifications represent the combined classification for males and females; instruments lacking classification ratings in the primary row for that instrument either included only studies involving one gender (but not the other), or reported outcomes by gender only (and not overall).

[a] Outcome is new felony adjudication rather than any new adjudication.

[b] An overall AUC was provided, but was calculated using a random sample of cases; accordingly, it is not reported because it did not represent the average of all male and female cases combined.

[c] Outcome is new violent felony adjudication rather than any new violent adjudication.

[d] One study did evaluate the PACT-PS for violence, but it was used to predict violence against staff or another youth in a locked facility; it was determined that this was not a use for which the instrument was intended, so the PV is not reported in the table (the PV value corresponded to "Fair" for this outcome).

[e] An overall PV was presented, but was out of range and so is not reported here.

[f] Developed to predict risk for first-time offending (i.e., before an offense has occurred).

Reliability. As with adults, few instruments provided outcomes for inter-rater reliability of juvenile brief assessments. Unlike adult instruments, however, considerable variability existed in the classification of juvenile instruments' inter-rater reliability. The Arizona Risk/Needs Assessment Instrument (ARNA), Positive Achievement Change Tool - Pre-Screen (PACT-PS), and Youth Level of Service/Case Management Inventory: Screening Version (YLS/CMI:SV) received the lowest inter-rater reliabilities (fair). All other youth brief assessments received either a good or excellent classification.

Table 37: Inter-Rater Reliability (IRR) for Juvenile Instruments (When Available)

Instrument	k_r	k	IRR
ARNA	2	2	Fair
Girls Link	1	1	Good
JSC	1	1	Excellent
OYAS-Disposition	3	1	Excellent
OYAS-Reentry	2	1	Excellent
PACT-PS	8	1	Fair
WSJCA-PS	4	1	Excellent
YASI-PS	6	1	Excellent
YLS/CMI:SV	2	2	Fair

Notes: k_r=total number of studies on assessment/reviewed (this number may be greater than the number with an IRR value); k=number of studies with an IRR value.

Discussion

Although the tables themselves serve as summaries of the findings from this study, several other observations warrant additional discussion. These topics are presented separately, although they are interrelated. The following discussion is intended to highlight these dependencies while also highlighting concerns that should be addressed when creating or adopting a brief assessment tool.

Brief Assessments Perform Best Where They Were Developed

Most of the instruments described in this report were limited in terms of the volume of research studies examining their psychometric properties. Those that were studied more frequently, including those that were developed in one jurisdiction and adopted by another, often did not perform well when transported outside the original jurisdiction. Exemplifying this finding, one study's examination of the New York City Juvenile Risk Assessment tool (Miller & Lin, 2007) found that the generic tool (an NCCD tool), even when locally normed, could not match the predictive accuracy of a locally developed tool, and actually performed more poorly than clinical judgment.

Such a finding was not at all unusual and local researchers often ended up making substantial modifications to stock tools (as observed in the tables above, many of the tools reviewed separately are related to one another and were developed through this type of process). This occurs for several reasons, the most obvious of which is a lack of statistical importance for some items when applied to a new population. Item-analysis often revealed that even the limited predictive value of a generic tool owed largely to one or two items rather than to the assessment as a whole. In rare (but important) instances, variables that were theoretically supported and predictively valid in one jurisdiction actually had the opposite relationship in another (i.e., a variable significantly predicted recidivism in one jurisdiction but predicted a lack of recidivism in another).

Another common problem with the application of a generic brief assessment to novel settings was that variables that were readily available in one setting were not in another, leading the new researchers to rely on proxies for the original variables. If a variable that is part of an assessment is not available when applied in a new jurisdiction (e.g., differing laws and policies allowing access to juvenile records), or is not reliably recorded, the utility of the tool is compromised.

The impact of poorly performing individual items is particularly important in the case of brief assessments. Longer assessments have the advantage that, if a variable is not particularly useful when applied to a new population, its poor predictive validity is often masked by the presence of several other variables that are predictive. Indeed, many of the studies only endeavored to locate poor performing items within generic tools because the tools as a whole were not predictive. Had they been predictive, one wonders if an attempt to examine individual items would have been as common; it may be that when generic tools fit 'well enough,' they are adopted without consideration of whether all items are important to prediction of the outcome (or are even detrimental).

Overall, the review of the literature examined in the course of this report has suggested that any jurisdiction would be wise to show caution when adopting a generic tool, or even using one as a starting point. When one seeks to predict an outcome, such as recidivism, with only a few variables, those variables must be carefully selected. If the process begins with a set of core variables that, for example, have been shown from a meta-analysis to be predictive across diverse settings, then the variables are, by definition, predictive in a general sense, yet the intent is to apply them in a very specific setting. Focusing on what is

predictive in general can lead to failure to consider the factors that make a jurisdiction unique. Those unique factors are often predictively important, and, while longer assessments can compensate for a lack of consideration of jurisdictional idiosyncrasies, brief assessments are generally unable to compensate for the loss of locally important predictors.

The synthesis of the research involving brief assessments suggests that researchers would do well to not only locally norm, but also locally develop brief assessments whenever possible in order to provide the possibility of addressing unique local characteristics (be they social, political, educational, or other) that drive recidivist behavior. This would involve consulting local experts in addition to the broader academic literature and would involve starting the process with a broader focus (including consideration of what is possible to assess/measure in a jurisdiction).

Assessments Do Not Usually Perform Well With Subgroups

Another common finding in the review of the relevant literature was that most brief assessments did not perform as well with 'subgroups' within the original population. In the report, attention was paid only to gender distinctions, and often assessments did not perform as well with females as with males (though this was not always the case, it was rare for a brief assessment to perform extremely well with both). To address this, many researchers appropriately developed gender-informed tools, or adopted modified versions by gender. Interestingly, this was rarely the case for other demographic divisions within the population.

This finding leads to the conclusion that local consideration must be paid to the original validation sample. If the original validation sample was largely comprised of one particular demographic, but the local population is predominately of another, there is little reason to expect that a brief assessment (which does not attempt to capture the broad theoretical, criminogenic profile of a full assessment) will be able to adequately address subgroup characteristics as moderators of criminal outcomes. While it is not practical to norm tools for all demographics, attention should be paid to whether the tool performs equally well for both genders, and, if a tool is adopted rather than locally developed, attention should be paid to notable demographic differences.

There is Not Much Research with Certain Subgroups

Related to the issue of how well instruments perform with subgroups is concern over the relative dearth of literature that has asked this question (in a substantive sense). There were several studies that examined the application of assessment tools to non-majority subgroups, but few had an explicit intent of developing tools for those subgroups. Instead, a tool was developed for the population as a whole (which was often rather homogenous), and then prediction was compared for various demographics. This is not the same as setting out to determine the best brief assessment predictors for each group. As one would expect, tools most often performed best when applied to the majority group, and diverged (sometimes dramatically) into poor prediction when applied to subgroups. As stated above,

it is not practical to either norm or develop a tool for all subgroups, but it is important that jurisdictions recognize that the current state of the literature on brief assessments and demographic-specific predictive validity is sparse. This translates into difficulties knowing whether theoretical constructs found in the literature are universally valid (at least within a brief assessment that cannot address them all).

Instruments Work Differently in Different Settings and for Different Outcomes

The review has also highlighted the concern that there is no 'one size fits all' tool; this applies to outcomes as well as demographics. Brief assessments rarely performed well across all outcomes (e.g., technical violations, general recidivism, and violent recidivism). Those that did (e.g., the SAQ) were often longer tools that met the criteria of this project because of their administrative ease rather than because they contained few items. This afforded these instruments the opportunity to assess items across domains that are relevant to several outcomes. Unless a longer tool with administrative ease (e.g., self-report) is adopted, jurisdictions may have to consider that a different tool may have to be used for different outcomes; this, of course, becomes an administrative burden in terms of cost and training (as well as other factors).

Assessments Can Perform Poorly for Other Reasons

Even with consideration of the myriad of concerns outlined above, it is possible that some of the evaluations of assessments reviewed revealed poor outcomes for reasons other than the tool's inherent predictive validity. For example, many of the tools that were evaluated in multiple studies were also evaluated by researchers other than those who developed the tool. Assessments tended to perform better when the creators evaluated the tool than when external researchers did. While this can occur for a number of reasons, one reason that undoubtedly created this discrepancy is failure to implement the tool as designed. This might be due to a lack of training or knowledge regarding implementation, or it might occur in much less obvious ways (for example, because variables that comprise the tool do not have the same meaning across jurisdictions, or are not recorded in the same way). Accordingly, when adopting a tool, consideration needs to be given to whether jurisdictions can implement the tool with fidelity.

All Tools Require Change

Related to the issue of implementation fidelity is the issue of systemic change. No matter how good an instrument is on paper, the system in which it will exist has to be prepared to change to conform to the requirements of the tool, and the tool also has to be integrated into organizational structure effectively. This may include changing how data is recorded and extracted, and will include allocating resources to on-going staff training, performance monitoring, and re-validation. On-going validation and refinement further increase organizational support and belief in the efficacy of the tool. Failure to continuously evaluate the adopted tool's efficacy, or to allocate sufficient resources to implementation, can also cause a tool to perform poorly for reasons unrelated to the tool itself.

There May Be an Upper Limit to Predictive Validity in Brief Assessments

Among the dozens of assessments that were examined, for this report, it is notable that very few brief assessments reached the 'excellent' predictive validity classification. Among those that were characterized as 'brief assessments' in the traditional sense (e.g., less than 10 items), even those that did reach the excellent classification were, without exception, on the lower end of the excellent range. There is likely a point of diminishing returns when trying to improve brief assessments; at a certain point, predictive validity cannot be further improved without the inclusion of additional and/or subjective and dynamic items, such that the tool no longer functions as a brief assessment.

While it is not uncommon for full-length assessments to reach considerably higher levels of predictive accuracy (e.g., AUCs of .8 or higher), such tools cannot meet the expediency, brevity, and cost-effectiveness goals of brief assessments. These outcomes suggest there is a tradeoff when adopting brief assessments. While a slight loss of predictive accuracy will likely occur, that loss can be attenuated by using these tools as guides rather than allowing them to fully dictate the handling of individual cases. As such, brief screening tools should be considered one step in a larger decision-making process, which includes written polices related to:

- The communication and interpretation of brief screening results—how are results communicated in recommendations and what specific placement and supervision options are associated with risk classification at each point of criminal justice processing.

- Use of overrides—overrides are a "method for deviating from the risk level assigned by a tool's total score" (Vincent et al., 2012, p. 62). There may be good reasons for deviating from a tool's recommendation; even with assessment tools, professional discretion is an important part of criminal justice decision-making (Latessa & Lovins, 2010). However, jurisdictions should develop policies and procedures guiding the use and monitoring of overrides, which should not exceed 10-15% of cases.

- Quality assurance measures for monitoring the administration and documentation of brief screening instruments.

- Data collection and reporting to monitor how results are used in decision-making.

References

[*Study was reviewed for inclusion in predictive validity or inter-rater reliability analysis]

Andrews, D. A., & Bonta, J. (2006). *The psychology of criminal conduct* (4th ed.). Newark, NJ: LexisNexis Mathew Bender.

Andrews, D. A., & Bonta, J. (2010). Rehabilitating criminal justice policy and practice. *Psychology, Public Policy, and Law, 16*(1), 39-55.

Andrews, D. A., Bonta, J., & Hoge, R. D. (1990). Classification for effective rehabilitation: Rediscovering psychology. *Criminal Justice and Behavior, 17*, 19-52.

Andrews, D. A., Bonta, J., & Wormith, S. J. (2006). The recent past and near future of risk/need assessment. *Crime and Delinquency, 52*, 7-27.

Andrews, D. A., & Dowden, C. (2007a). Managing correctional treatment for reduced recidivism: A meta-analytic review of programme integrity. *Legal and Criminological Psychology, 10*, 173-187.

*Ashford, J. B., & LeCroy, C. W. (1988). Predicting recidivism: An evaluation of the Wisconsin Juvenile Probation and Aftercare Risk Instrument. *Criminal Justice and Behavior, 15*(2), 141-151.

*Ashford, J. B., & LeCroy, C. W. (1990). Juvenile recidivism: A comparison of three predictive instruments. *Adolescence, 25*(98), 441-450.

*Assink, M., van der Put, C. E., & Stams, G. J. J. M. (2014). The development and validation of an actuarial risk assessment tool for prediction of first-time offending. *International Journal of Offender Therapy and Comparative Criminology,* 1-18.

Austin, J. (2004). The proper and improper use of risk assessment in corrections. *Federal Sentencing Reporter, 16*(3), 1-6.

*Austin, J., Bhati, A., Jones, M. & Ocker, R. (n.d.). *Florida pretrial risk assessment instrument*. Denver, CO: The JFA Institute.

*Austin, J., Ocker, R., & Bhati, A. (2012). *Kentucky Pretrial Risk Assessment Instrument validation*. Washington, DC: The JFA Institute.

* Baglivio, M. T. (2009). The assessment of risk to recidivate among a juvenile offending population. *Journal of Criminal Justice*, *37*, 596-607.

*Baglivio, M. T., & Jackowski, K. (2013). Examining the validity of a juvenile offending risk assessment instrument across gender and race/ethnicity. *Youth Violence and Juvenile Justice, 11*(1), 26-43.

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

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

Bechtel, K., Lowenkamp, C. T., Holsinger, A. (2011). Identifying the predictors of pretrial failure: A meta-analysis. *Federal Probation, 75*(2), 78-87.

*Behavioral Data Systems. (n.d.). *DVI: An inventory of scientific findings.* Phoenix, AZ: Author.

*Benda, B. B., Corwyn, R. F., & Toombs, N. J. (2001). From adolescent "serious offender" to adult felon. *Journal of Offender Rehabilitation*, *32*(3), 79-108.

*Blomberg, T., Bales, W., Mann, K., Meldrum, R., & Nedelec, J. (2010). *Validation of the COMPAS Risk Assessment Classification Instrument.* Tallahassee, FL: Center for Criminology and Public Policy Research, Florida State University.

*Bogue, B., Woodward, W., & Joplin, L. (2006). *Using a proxy score to pre-screen offenders for risk to reoffend.* Hawaii: Author.

*Bonta, J., Harman, W. G., Hann, R. G., & Cormier, R. B. (1996). *Canadian Journal of Criminology, 38*, 61-79.

Borenstein, M., Hedges, L. V., Higgins, J. P. T., & Rothstein, H. R. (2009). *Introduction to meta-analysis.* New York: John Wiley & Sons.

Breiding, M. J., Chen, J., & Black, M. C. (2014). Intimate partner violence in the United States—2010. Atlanta, GA: Center for Injury Prevention and Control, Centers for Disease Control and Prevention.

*Brennan, T. D., Dieterich, W., & Ehret, B. (2009). Evaluating the predictive validity of the COMPAS risk and needs assessment system. *Criminal Justice and Behavior, 36*(1), 21-40.

*Brown, S. L., St. Amand, M. D., & Zamble, E. (2009). The dynamic prediction of criminal recidivism: A three-wave prospective study. *Law and Human Behavior, 33*(1), 25-45.

*Butts, J. A., & Demuro, P. (1989). *Population profile and risk assessment study: Mississippi Department of Youth Services.* Ann Arbor, MI: University of Michigan Center for the Study of Youth Policy.

*Cadigan, T. P., Johnson, J. L., & Lowenkamp, C. T. (2012). The Re-validation of the Federal Pretrial Services Risk Assessment (PTRA). *Federal Probation, 76*(2), 3-9.

Campbell, M., French, S., & Gendreau, P. (2007). *Assessing the utility of risk assessment tools and personality measures in the prediction of violent recidivism for adult offenders.* Ottawa, Ontario, Canada: Department of Safety and Emergency Preparedness.

Campbell, M. A., French, S., & Gendreau, P. (2009). The prediction of violence in adult offenders: A meta-analytic comparison of instruments and methods of assessment. *Criminal Justice and Behavior*, *36*, 567-590.

*Caron, A. (2009). *Property drug calendar evaluation and RANT (Risk and Needs Triage) assessment.* Minneapolis, MN: Fourth Judicial District Research Division.

*Center for Criminal Justice Research (2010). Indiana Risk Assessment System – Community Supervision Tool (IRAS-CST) scoring guide. Cincinnati, OH: University of Cincinnati Center for Criminal Justice Research.

Chen, H., Cohen, P., & Chen, S. (2010). How big is a big odds ratio? Interpreting the magnitudes of odds ratios in epidemiological studies. *Communications in Statistics - Simulation and Computation, 39*, 860-864.

Christensen, G., Jannetta, J., & Buck-Willison, J. (2012). *The role of screening and assessment in jail reentry.* Washington, DC: The Urban Institute.

*Chu, C. M., Yu, H., Lee, Y., & Zeng, G. (2014). The utility of the YLS/CMI-SV for assessing youth offender in Singapore. *Criminal Justice and Behavior, 41*(12), 1437-1457.

Cicchetti, D. (2001). Methodological commentary the precision of reliability and validity estimates re-visited: Distinguishing between clinical and statistical significance of sample size requirements. *Journal of Clinical and Experimental Neuropsychology*, *23*(5), 695-700.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Erlbaum.

Cohen, T. H. (2013). *Pretrial detention and misconduct in federal district courts, 1995-2010.* (NCJ 239673). Washington, DC: Bureau of Justice Statistics.

Cohen, J., Cohen, P., West, S. G., & Aiken, L. S. (2003). *Applied multiple regression/correlation analysis for the behavioral sciences* (3rd ed.). Mahwah, NJ: Erlbaum.

*Coid, J., Yang, M., Ullrich, S., Zhang, T., Roberts, A., Roberts, C., Rogers, R., & Farrington, D. (2007). *Predicting and understanding risk of re-offending: The prisoner cohort study.* London, UK: Ministry of Justice

*Coid, J., Yang, M., Ullrich, S., Zhang, T., Sizmur, S., Roberts, C., & Rogers, R. D. (2009). Gender differences in structured risk assessment: Comparing the accuracy of five instruments. *Journal of Consulting and Clinical Psychology*, *77*(2), 337-348.

*Colorado Department of Human Services. (2013). *Recidivism evaluation of committed youth discharged in fiscal year 2010-11.* Denver, CO: Author.

*Colorado Department of Human Services. (2013). *Recidivism evaluation of DYC committed youth discharged in fiscal years 2010-2011 and 2011-2012.* Denver, CO: Author.

*Connolly, M. M. (2003). *A critical examination of actuarial offender-based prediction assessments: Guidance for the next generation of assessments* (Unpublished doctoral dissertation). Austin, TX: University of Texas at Austin.

*Copas, J., & Marshall, P. (2002). The Offender Group Reconviction Scale: A statistical reconviction score for use by probation officers. *Journal of The Royal Statistical Society Series C-Applied Statistics*, *47*(1), 159-171.

*Daffern, M., Ogloff, J. R., Ferguson, M., & Thomson, L. (2005). Assessing risk for aggression in a forensic psychiatric hospital using the Level of Service Inventory-Revised: Screening Version. *International Journal of Forensic Mental Health*,*4*(2), 201-206.

*Dedel, K., & Davies, G. (2007). *Validating Multnomah County's Juvenile Detention Risk Assessment Instrument.* Portland, OR: One in 37 Research, Inc.

*Dembo, R., Turner, G., Sue, C. C., Schmeidler, J., Borden, P., & Manning, D. (1994). An assessment of the Florida Department of Health and Rehabilitative Services Detention Risk Assessment Instrument on youths screened and processed at the Hillsborough County Juvenile Assessment Center. *Journal of Child and Adolescent Substance Abuse*, *4*(1), 45-71.

Desmarais, S. L., & Singh, J. P. (2013). *Risk assessment instruments validated and implemented in correctional settings in the United States*. New York: Council of State Governments Justice Center.

*Dhaliwal, G. K., Loza, W., & Reddon, J. R. (2007). Invalidity of "Disconfirmation of the Predictive Validity of the Self-Appraisal Questionnaire in a Sample of High-Risk Drug Offenders" (2006): A reply. *Journal of Interpersonal Violence*, *22*(8), 1077-1089.

*Duwe, G. (2014). The development, validity, and reliability of the Minnesota Screening Tool Assessing Recidivism Risk (MnSTARR). *Criminal Justice Policy Review, 25*(5), 579-613.

*Early, K. P., Hand, G. A., & Blankenship, J. L. (2012). *Validity and reliability of the Florida PACT risk and needs assessment instrument: A three-phase evaluation*. Tallahassee, FL: Justice Research Center.

*Eckberg, D., & Kubits, G. (2006). *Fourth Judicial District pretrial evaluation: Scale validation study*. Minneapolis, MN: Fourth Judicial District Research Division.

*Eisenberg, M., Bryl, J., & Fabelo, T. (2009). *Validation of the Wisconsin Department of Corrections risk assessment instrument.* Lexington, KY: Council of State Governments Justice Center.

*Eisenberg, M., Fabelo, T., & Tyler, J. (2011). *Validation of the Orange County California Probation Department Risk Assessment Instrument: Final report.* New York, NY: Council of State Governments Justice Center.

*Farabee, D., Zhang, S., Roberts, R. E. L., & Yang, J. (2010). *COMPAS validation study: Final Report*. Los Angeles, CA: California Department of Corrections.

*Fass, T. L., Heilbrun, K., DeMatteo, D., & Fretz, R. (2008). The LSI-R and the COMPAS: Validation data on two risk-needs tools. *Criminal Justice and Behavior, 1-14.*

*Feagans, D., & Mathews, H. (2012). *Analysis of General Risk Score Levels, Publication No. 13-290-DOC-01*. Massachusetts Department of Corrections.

Federal Bureau of Investigation. (2013). *Uniform crime reports: Crime in the United States 2012*. Washington, DC: Author.

Federal Bureau of Investigation. (2014). *Uniform crime reports: Crime in the United States 2013*. Washington, DC: Author.

*Ferguson, A. M., Ogloff, J. R., & Thomson, L. (2009). Predicting recidivism by mentally disordered offenders using the LSI-R:SV. *Criminal Justice and Behavior*, *36*(1), 5-20.

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.

*Fourth Judicial District of Minnesota Research Division (2006). *Fourth Judicial District pretrial evaluation: Scale validation study*. Hennepin County, MN: Minnesota Judicial Branch, Fourth Judicial District.

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

*Gehring, K. S. (2011). *Are needs related to pretrial outcomes? An examination of the Hamilton County Inventory of Need Pretrial Screening Tool (Doctoral dissertation).* University of Cincinnati: Cincinnati, OH.

*Gehring, K. S., & Van Voorhis, P. (2014). Needs and pretrial failure additional risk factors for female and male pretrial defendants. *Criminal Justice and Behavior, 41(8),* 943-970.

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

Gendreau, P., Smith, P., & French, S. (2006). The theory of effective correctional intervention: Empirical status and future directions. In F. T. Cullen, J. P. Wright, and K. R. Blevins (Eds.), *Taking stock: The status of criminological theory—Advances in criminological theory: Vol. 15* (pp. 419-446). New Brunswick, NJ: Transaction.

Glaze, L. E. & Kaeble, D. (2014). *Correctional Populations in the United States, 2013*. U.S. Department of Justice, Office of Justice Programs, Bureau of Justice Statistics.

*Glover, A. J. J., Nicholson, D. E., Hemmati, T., Bernfeld, G. A., & Quinsey, V. L. (2002). A comparison of predictors of general and violent recidivism among high-risk female offenders. *Criminal Justice and Behavior, 29*(3), 235-249.

Gover, A. R., Richards, T. A., & Tomsich, E. A. (2015). Colorado's innovative response to domestic violence offender treatment: Current achievements and recommendations for the future. Denver, CO: Buechner Institute for Governance, School of Public Affairs, University of Colorado.

Gottfredson, S. D., & Moriarity, L. J. (2006). Statistical risk assessment: Old problems and new applications. *Crime & Delinquency, 52*(1), 178-200.

*Gray, N. S., Snowden, R. J., MacCulloch, S., Phillips, H., Taylor, J., & MacCulloch, M. J. (2004). Relative efficacy of criminological, clinical, and personality measures of future risk of offending in mentally disordered offenders: A comparative study of HCR-20, PCL: SV, and OGRS. *Journal of Consulting and Clinical Psychology*, *72*(3), 523-530.

Greene, M., Hoffman, P., & Beck, J. (1994). The mean cost rating (MCR) is Somers' D: A methodological note. *Journal of Criminal Justice, 22*(1), 63-69.

*Hamilton, Z., Neuilly, M. A., Lee, S., & Barnoski, S. (2014). Isolating modeling effects in offender risk assessment. *Journal of Experimental Criminology, 11*(2), 299-318.

Hannah-Moffat, K. (2010). *Actuarial sentencing: An "unsettled" proposition*. Presentation for the September, 2010, Symposium on Sentencing, Albany, NY.

*Harris, P. M. (1994). Client management classification and prediction of probation outcome. *Crime & Delinquency*, *40*(2), 154-174.

Hedlund, J., & Cox, S. M. (2005). *Development and validation of an assessment for pretrial conditional release.* Department of Criminology and Criminal Justice, Central Connecticut State University.

Hedlund, J., Cox, S. M., & Wichrowski, S. (2003). *Validation of Connecticut's risk assessment for pretrial decision making*. Hartford, CT: Court Support Services Division, Judicial Branch, State of Connecticut.

*Henderson, H., & Miller, H. (2011). The (twice) failure of the Wisconsin Risk Need Assessment in a sample of probationers. *Criminal Justice Policy Review,* 1-23.

Herberman, E. J., & Bonczar, T. P. (2015). *Probation and parole in the United States, 2013*. (NCJ 248029). Washington, DC: Bureau of Justice Statistics.

*Herrschaft, B. A. (2014). *Evaluating the reliability and validity of the Correctional Offender Management Profiling for Alternative Sanctions (COMPAS) Tool: Implications for community corrections policy (Doctoral dissertation).* Rutgers University, Newark, NJ.

*Hess, J., & Turner, S. (2013). *Risk assessment accuracy in corrections population management: Testing the promise of tree based ensemble predictions*. Irvine, CA: University of California, Irvine, Center for Evidence-Based Corrections.

*Hickert, A. O., Worwood, E. B., & Prince, K. (2013). *Pretrial release risk study, validation & scoring: Final Report*. Salt Lake City, UT: Utah Criminal Justice Center.

*Hilton, N. Z., & Harris, G. T (2009). How nonrecidivism affects predictive accuracy: Evidence from a cross-validation of the Ontario Domestic Assault Risk Assessment (ODARA). *Journal of Interpersonal Violence, 24*(2), 326-337.

*Hilton, N. Z., Harris, G. T., Rice, M. E., Lang, C., Cormier, C. A., & Lines, K. J. (2004). A brief actuarial assessment for the prediction of wife assault recidivism: The Ontario Domestic Assault Risk Assessment. *Psychological Assessment, 16*(3), 267-275.

Hilton, N. Z., Harris, G. T., & Rice, M. E. (2006). Sixty-six years of research on the clinical versus actuarial prediction of violence. *Counseling Psychologist, 34*(3), 400-409.

*Hilton, Z., Harris, G. T., Rice, M. E., Houghton, R. E., & Eke, A. W. (2008). An indepth actuarial assessment for wife assault recidivism: The Domestic Violence Risk Appraisal guide. *Law and Human Behavior, 32(*2), 150-163

*Hisashima, J. (2008). *Validation study of the Domestic Violence Screening Instrument (DVSI).* Honolulu, HI: Hawaii State Department of Health.

Hirschel, D., Buzawa, E., Pattavina, A., Faggiani, D., & Reuland, M. (2007). *Explaining the prevalence, context, and consequences of dual arrest in intimate partner cases.* (USDOJ Document Number 218355). Washington, DC: Government Printing Office.

*Hollist, D., Coolidge, J., Delano, W., Greenwood, I., King, M., McLean, T., McKay, P., Burfeind, J., Harris, C., & Doyle, D. (2012). *The Montana Pre-Adjudicatory Detention Risk Assessment Instrument: A validation and assessment study.* Missoula, MT: University of Montana Social Science Research Laboratory.

*Howard, P. D., & Dixon, L. (2012). The construction and validation of the OASYS violence predictor: Advancing violence risk assessment in the English and Welsh correctional services. *Criminal Justice and Behavior*, *39*(3), 287-307.

*Howard, P. D., Francis, B., Soothill, K., & Humphreys, L. (2009). *OGRS 3: The revised Offender Group Reconviction Scale.* London, UK: Ministry of Justice.

*Huff, D. J. (1999). *Special report: Managing risks in juvenile justice.* Des Moines, Iowa: Iowa Department of Human Rights.

*Huff, D. J., & Prell, L. (n.d.). *The proposed Iowa Juvenile Court Intake Risk Assessment.* Des Moines, Iowa: Division of Criminal and Juvenile Justice Planning and Statistical Analysis Center.

Hunter, J. E., & Schmidt, F. L. (2004). *Methods of meta-analysis: Correcting error and bias in research findings* (2nd ed.). Thousand Oaks, CA: Sage.

*Interagency Advisory Committee on Adult and Juvenile Correctional Treatment. (2007). *Colorado reference guide: Juvenile screening and assessment instruments.* Denver, CO: Author.

*Johnson, D. L. (2013). *The risk of recidivism and the Positive Achievement Change Tool Pre-Screen (PACT-PS): An initial validation and basic program implementation evaluation* (Unpublished doctoral dissertation). University of Missouri, Kansas City, MO.

*Johnson, J. L., Lowenkamp, C. T., VanBenschoten, S. W., & Robinson, C. R. (2011). The construction and validation of the Federal Post Conviction Risk Assessment (PCRA). *Federal Probation, 1*6, 1-33.

*Johnson, K., Wagner, D., & Mathews, T. (2002). *Missouri Juvenile Risk Assessment re-validation report.* Madison, WI: National Council on Crime and Delinquency.

Jones, N. J. (2011). Merging theoretical frameworks to inform risk assessment for the young female offender* (Unpublished doctoral dissertation). Carlton University, Ottawa, Canada.

*Jones, N. J., Brown, S. L., Robinson, D., & Frey, D. (2015). Incorporating strengths into quantitative assessments of criminal risk for adult offenders: The Service Planning Instrument. *Criminal Justice and Behavior, 42*(3), 321-338.

*Jones, N. J., Brown, S. L., Wanamker, K. A., & Greiner, L. E. (2014). A quantitative exploration of gendered pathways to crime in a sample of male and female offenders. *Feminist Criminology, 9*(2), 113-136.

Judicial Council of California/Administrative Office of the Courts (2011). *Screening and assessments used in the juvenile justice system: Evaluating risks and needs of youth in the juvenile justice system.* San Francisco, CA: AOC Center for Families, Children, & Courts.

*Justice Research Center. (2006). *Pre-validation study of the Florida Department of Juvenile Justice Positive Achievement Change Tool (PACT): Executive summary.* Tallahassee, FL: Justice Research Center.

*Kallis, C., Bui, L., Yang, M., & Coid, J. W. (2014). Static screening instruments for risk of minor and major violence. *The Journal of Forensic Psychiatry & Psychology, 25*(4), 397-410.

*Kingston, W. L., MacTavish, A., & Loza-Fanous, A. (2007). A nine-year follow-up study on the predictive validity of the Self-Appraisal Questionnaire for predicting violent and nonviolent recidivism. *Journal of Interpersonal Violence*, *22*(9), 1144-1155.

*Kroner, D. G., & Loza, W. (2001). Evidence for the efficacy of the self-report in predicting nonviolent and violent criminal recidivism. *Journal of Interpersonal Violence, 16*(2), 168-177.

*Kroner, D. G., & Mills, J. F. (2001). The accuracy of five risk appraisal instruments in predicting institutional misconduct and new convictions. *Criminal Justice and Behavior, 28*(4), 471–489.

*Krysik, J., & LeCroy, C. W. (2002). The empirical validation of an instrument to predict risk of recidivism among juvenile offenders. *Research on Social Work Practice, 12*(1), 71-81.

*Kubiak, S. P., Kim, W. J., Bybee, D., & Eshelman, L. (2014). Assessing the validity of the Self-Appraisal Questionnaire in differentiating high-risk and violent female offenders. *The Prison Journal, 94*(3), 305-327.

*Lansing, S. (2012). *New York State COMPAS-probation risk and need assessment study: Examining the Recidivism Scale's effectiveness and predictive accuracy.* New York: New York State Division of Criminal Justice Services.

*Latessa, E.J., Lemke, R., Makarios, M., Smith, P., & Lowenkamp, C.T. (2010). The creation and validation of the Ohio risk assessment system (ORAS). *Federal Probation 74*(1)16-22.

Latessa, E. J., & Lovins, B. (2010). The role of offender risk assessment: A policy maker guide. *Victims & Offenders, 5*(3), 203-219.

*Latessa, E., Lovins, B., & Lux, J. (2014). *The Ohio risk Assessment System, Misdemeanor Assessment Tool (ORAS-MAT) and Misdemeanor Screening Tool (ORAS-MST.* Cincinnati, OH: University of Cincinnati Center for Criminal Justice Research.

*Latessa, E., Lovins, B., & Makarios, M. (2013). *Validation of the Indiana Risk Assessment System: Final report.* Cincinnati, OH: University of Cincinnati School of Criminal Justice, Center for Criminal Justice.

*Latessa, E., Lovins, B., & Ostrowski, K. (2009). *The Ohio Youth Assessment System: Final Report.* Cincinnati, OH: University of Cincinnati Center for Criminal Justice Research

Latessa, E. J., & Lowenkamp, C. T. (2005). What are criminogenic needs and why are they important? *For the Record, 4th Quarter,* 18-20.

*Latessa, E. J., Smith, P., Lemke, R., Makarios, M., & Lownkamp, C. (2009). *Creation and validation of the Ohio Risk Assessment System: Final report.* Cincinnati, OH: University of Cincinnati, Center for Criminal Justice Research.

*Latessa, E., Smith, P., Lemke, R., Makarios, M., & Lowenkamp, C. (2010). Creation and validation of the Ohio Risk Assessment System (ORAS). *Federal Probation, 74*(1), 16-22.

*Laura and John Arnold Foundation (2013). *Developing a national model for pretrial risk assessment (Research Summary).* Houston, TX: Laura and John Arnold Foundation.

*Laura and John Arnold Foundation (2014). *Results from the first six months of the Public Safety Assessment – Court in Kentucky.* Houston, TX: Laura and John Arnold Foundation.

*LeCroy, C. W., Krysik, J., & Palumbo, D. (1998). *Empirical validation of the Arizona Risk/Needs Instrument and assessment process.* Tucson, AZ: LeCroy & Milligan Associates, Inc.

*Lemmon, R. A., & Calhoon, S. K. (1998). Predicting juvenile recidivism using the Indiana Department of Correction's Risk Assessment Instrument. *Juvenile and Family Court Journal, 49*, 55-61.

*Levin, D. J. (2011). *Development of a validated pretrial risk assessment tool for Lee County, Florida.* Washington, DC: Pretrial Justice Institute.

*Loinaz, I. (2013). Typologies, risk and recidivism in partner-violent men with the B-SAFER: A pilot study. *Psychology, Crime and Law, 20*(2), 183-198.

*Lovins, B., & Latessa, E. (2013). Creation and validation of the Ohio youth assessment system (OYAS) and strategies for successful implementation. *Justice Research and Policy, 15*(1), 67-93.

*Lowenkamp, C. T., Holsinger, A. M., & Cohen, T. H. (2015). PCRA revisited: Testing the validity of the Federal Post Conviction Risk Assessment (PCRA). *Psychological Services, 12*(2), 149-157.

*Lowenkamp, C. T., Johnson, J. L., Holsinger, A. M., VanBenschoten, S. W., & Robinson, C. R. (2013). The Federal Post Conviction Risk Assessment (PCRA): A construction and validation study. *Psychological Services, 10*(1), 87-96.

Lowenkamp, C. T., & Latessa, E. J. (2004). Understanding the risk principle: How and why correctional interventions can harm low-risk offenders. *Topics in Community Corrections,* 3-8.

*Lowenkamp, C. T., & Whetzel, J. (2009). The development of an actuarial risk assessment instrument for US Pretrial Services. *Federal Probation, 73(2),* 33-36.

*Lowenkamp, C. T., Lemke, R., & Latessa, E. (2008). The development and validation of a pretrial screening tool. *Federal Probation, 72*(3)2-9.

*Lowenkamp, C. T., Lovins, B., & Latessa, E. J. (2009). Validating the Level of Service Inventory-Revised and the Level of Service Inventory: Screening Version with a sample of probationers. *Prison Journal*, *89*(2), 192-204.

Lowenkamp, C. T., Pealer, J., Smith, P., & Latessa, E. J. (2006). Adhering to the Risk and Need Principles: Does it Matter for Supervision-Based Programs? *Federal Probation*, *70*(3), 3-8.

Lowenkamp, C. T., VanNostrand, M., & Holsinger, A. (2013). *The hidden costs of pretrial detention*. Houston, TX: The Laura and John Arnold Foundation.

*Lowenkamp, C. T., & Whetzel, J. (2009). The development of an actuarial risk assessment instrument for U.S. pretrial services. *Federal Probation Journal, 73*(2).

*Loza, W., Conley, M., & Warren, B. (2004). Concurrent cross-validation of the Self-Appraisal Questionnaire: A tool for assessing violent and nonviolent recidivism and institutional adjustment on a sample of North Carolina offenders. *International Journal of Offender Therapy and Comparative Criminology*, *48*(1), 85-95.

*Loza, W., Cumbleton, A., Shahinfar, A., Neo, L. H., Evans, M., Conley, M., & Summers, R. (2004). Cross-validation of the Self-Appraisal Questionnaire (SAQ): An offender risk and need assessment measure on Australian, British, Canadian, Singaporean, and American offenders. *Journal of Interpersonal Violence*, *19*(10), 1172-1190.

*Loza, W., Dhaliwal, G., Kroner, D. G., & Loza-Fanous, A. (2000). Reliability, construct, and concurrent validities of the Self-Appraisal Questionnaire: A tool for assessing violent and non-violent recidivism. *Criminal Justice and Behavior, 27*, 356-374.

*Loza, W., & Green, K. (2003). The Self-appraisal Questionnaire: A self-report measure for predicting recidivism versus clinician-administered measures: A 5-year follow-up study. *Journal of Interpersonal Violence*, *18*(7), 781-797.

*Loza, W., & Loza-Fanous, A. (2000a). Predictive validity of the Self-Appraisal Questionnaire (SAQ): A tool for assessing violent and nonviolent release failures. *Journal of Interpersonal Violence*, *15*(11), 1183-1191.

*Loza, W., & Loza-Fanous, A. (2000b). The effectiveness of the self-appraisal questionnaire in predicting offenders' postrelease outcome: A comparison study. *Criminal Justice and Behavior, 28*(1), 105-121.

*Loza, W., & Loza-Fanous, A. (2003). More evidence of the validity of the Self-Appraisal Questionnaire for predicting violent and nonviolent recidivism: A 5-year follow-up study. *Criminal Justice and Behavior, 30*, 709-721.

*Loza, W., Loza-Fanous, A., & Heseltine, K. (2007). The myth of offenders' deception on self-report measure predicting recidivism: Example from the Self-Appraisal Questionnaire (SAQ). *Journal of Interpersonal Violence*, *22*(6), 671-683.

*Loza, W., Neo, L. H., Shahinfar, A., & Loza-Fanous, A. (2005). Cross-validation of the Self-Appraisal Questionnaire: A tool for assessing violent and nonviolent recidivism with female offenders. *International Journal of Offender Therapy and Comparative Criminology*, *49*(5), 547-560.

Mamalian, C. A. (2001). *State of the science of pretrial risk assessment*. Gaithersburg, MD: Pretrial Justice Institute.

*Marlow, D. B., Festinger, D. S., Dugosh, K. L., Caron, A., Podkopacz, M. R., & Clements, N. T. (2011). Targeting dispositions for drug-involved offenders: A field trial of the Risk and Needs Triage (RANT). *Journal of Criminal Justice, 39*, 253-260.

*Martin, J. H. (2012). *Predicting risk to reoffend: Establishing the validity of the Positive Achievement Change Tool* (Unpublished doctoral dissertation). Orlando, FL: University of Central Florida.

*McCafferty, J. T. (2013). *The predictive validity of the Ohio Youth Assessment System – Dispositional Instrument: A revalidation study* (Unpublished doctoral dissertation). University of Cincinnati, Cincinnati, OH.

*McElfresh, R. A. (2011). *Missouri Juvenile Risk Assessment re-validation report III.* Jefferson City, MO: Missouri Supreme Court Office of State Courts Administrator.

Mcgrath, R., & Meyer, G. (2006). When effect sizes disagree: The case of r and d. *Psychological Methods, 11*(4), 386-401.

*McKay, P. D. (2012). *Rescoring the Montana Risk Assessment Instrument: A comparison of methods* (Unpublished master's thesis). University of Montana, Missoula, MT.

*Meredith, T. (2014). *Ramsey County Proxy tool norming and validation results.* Atlanta, GA: Applied Research Services.

*Messing, J. T., Campbell, J., Wilson, J. S., Brown, S., & Patchell, B. (2015). The Lethality Screen: The predictive validity of an intimated partner violence risk assessment for first responders. *Journal of Interpersonal Violence, 1-22.*

Messing, J. T., & Thaller, J. (2013). The average predictive validity of intimate partner violence risk assessment instruments. *Journal of Interpersonal Violence, 28*(7), 1537-1558.

*Miller, H. (2006). A dynamic assessment of offender risk, needs, and strengths in a sample of pre-release general offenders. *Behavioral Sciences and the Law, 24*, 767-782.

*Miller, J., & Lin, J. (2007). Applying a generic juvenile risk assessment instrument to a local context: Some practical and theoretical lessons. *Crime & Delinquency*, *53* (4), 552-580.

*Mills, J. F., & Gray, A. L. (2013). Two-tiered violence risk estimates: A validation study of an integrated-actuarial risk assessment instrument. *Psychological Services, 10*(4), 361-371.

*Mills, J. F., & Kroner, D. G. (2006). The effect of discordance among violence and general recidivism risk estimates on predictive accuracy. *Criminal Behaviour and Mental Health*, *16*, 155-166.

*Mills, J. F., Loza, W., & Kroner, D. G. (2003). Predictive validity despite social desirability: Evidence for the robustness of self-report among offenders. *Criminal Behaviour and Mental Health*, *13*, 140-150.

Minton, T. D., & Zeng, Z. (2015). *Jail inmates at midyear 2014.* (NCJ 248629). Washington, DC: Bureau of Justice Statistics.

*Mitchell, O., Caudy, M. S., & MacKenzie, D. L. (2012). A reanalysis of the Self-Appraisal questionnaire: Psychometric properties and predictive validity. *International Journal of Offender Therapy and Comparative Criminology, 57*(4), 445-459.

*Mitchell, O., & Mackenzie, D. L. (2006). Disconfirmation of the predictive validity of the Self-Appraisal Questionnaire in a sample of high-risk drug offenders. *Criminal Justice and Behavior*, *33*(4), 449-466.

*Mulinix, J. A. (n.d.). *Validity and accuracy of risk assessment in criminal justice: An examination of the Ohio Risk Assessment System*. Unpublished manuscript: Defiance College.

*Nafekh, M., & Motiuk, L. L. (2002). *The Statistical Information on Recidivism – Revised 1 (SIR-R1) Scale: A psychometric examination.* Ottawa, Ontario, Canada: Correctional Service of Canada.

*National Council on Crime and Delinquency. (2000). *Alameda County Placement Risk Assessment validation.* Oakland, CA: Author.

National Judicial College. (2012). Principles of an effective criminal justice response to the challenges and needs of drug-involved individuals. Reno, NV: Author.

Nicholls, T. L., Pritchard, M. M., Reeves, K. A., & Hilterman, E. (2013). Risk assessment in intimate partner violence: A systematic review of contemporary approaches. *Partner Abuse, 4*(1), 76-168.

*O'Byrne, P. L. (2002). *Exploring the predictive ability of risk needs assessment in Saskatchewan corrections* (Unpublished master's thesis). University of Regina, Regina, Canada.

Ohio Department of Rehabilitation and Correction. (2010, June 15). *Ohio Risk Assessment System*. Retrieved from http://www.drc.ohio.gov/web/ORAS.htm.

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

*Orbis Partners, Inc. (2011). *Validation of the Youth Assessment and Screening Instrument (YASI) for use by the Vermont Department for Children and Families.* Ottawa, Ontario, Canada: Author.

*Oregon Youth Authority Research and Evaluation. (2011). *OYA Recidivism Risk Assessment: Modeling risk to recidivate.* Salem, OR: Author.

*Oregon Youth Authority Research and Evaluation. (2011). *OYA Recidivism Risk Assessment – Violent Crime (ORRA-V): Modeling risk to recidivate with a violent crime.* Salem, OR: Author.

*Ostrom, B., Kleiman, M., Cheesman, F., Hansen, R., & Kauder, N. (2002). *Offender Risk Assessment in Virginia: A three-stage evaluation*. Williamsburg, VA: National Center for State Courts.

*Pitocco, K. (n.d.). *ORAS (Ohio Risk Assessment System): Purpose and benefits* (PowerPoint slides). Cincinnati, OH: University of Cincinnati Corrections Institute.

*Podkopacz, M.R. (2006). *Fourth Judicial District of Minnesota pretrial evaluation: Scale validation study* (PDF document of PowerPoint slides). Minneapolis, MN: Fourth Judicial District Research Division.

*Podkopacz, M. R. (2010). *Building and validating the 2007 Hennepin County Adult Pretrial Scale*. Minneapolis, MN: Fourth Judicial District Research Division.

*Podkopacz, M. R. (2009). *Hennepin County Juvenile Detention Alternative Initiative: Validation of the risk assessment instrument.* Hennepin County, MN: Fourth Judicial District Research Division.

*Prell, L. (2008). *Validation of the 5th Judicial District Pretrial Release Point Schedule on Polk County defendants*. Des Moines, IA: Iowa Department of Corrections.

Puzzanchera, C. (2013). *Juvenile Arrests 2011*. In Juvenile offenders and victims: National report series. U.S. Washington, DC: U.S. Department of Justice, Office of Justice Programs, Office of Juvenile Justice and Delinquency Prevention.

*Pyrooz, D. (2012). Absconding among juvenile parolees in Arizona: Constructing and validating a gender-responsive risk instrument. *Criminal Justice and Behavior, 39*(7), 923-937.

*Ranadive, D. (2014). *Cross-validation of the El Paso Risk Assessment of Juveniles at Intake Scale (El Paso RAJI)* (Unpublished master's thesis). The University of Texas at El Paso, El Paso, TX.

*Reiner, S., Miller, J. B., & Gangal, T. (2007). Public safety outcomes of Virginia's Detention Assessment Instrument. *Juvenile and Family Court Journal*, *58*(3), 31-41.

*Rembert, D. A. (2013). *The utility of the Positive Achievement Change Tool in predicting assault among state committed youths* (Doctoral dissertation). Retrieved from ProQuest Dissertations and Theses database. (3571395).

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

*Roehl, J., O'Sullivan, C., Webster, D., & Campbell, J. (2005). *Intimate partner violence risk assessment validation study: Final report.* (USDOJ Document Number 209731). Washington, DC: Government Printing Office.

*Rud, J., Skilling, N., & Nonemaker, D. (2010). *DOCCR validation of two domestic violence risk screening instruments: Domestic Violence Screening Instrument (DVSI) & Spousal Abuse Risk Assessment (SARA)*. Minneapolis, MN: Hennepin County Office of Planning, Policy and Evaluation.

Ruscio, J. (2008). A probability-based measure of effect size: Robustness to base rates and other factors. *Psychological Methods, 13*, 19-30.

*Schwalbe, C. S. (2007). *Revalidation of the Arizona Risk/Needs Assessment Instrument.* New York: Columbia University School of Social Work.

*Schwalbe, C. S. (2008a). A meta-analysis of juvenile justice risk assessment instruments: Predictive validity by gender. *Criminal Justice and Behavior, 35*(11), 1367-1381.

*Schwalbe, C. S. (2008b). Risk assessment stability: A revalidation study of the Arizona Risk/Needs Assessment Instrument. *Research on Social Work Practice*, 1-9.

*Schwalbe, C. S., Fraser, M. W., & Day, S. H. (2007). Predictive validity of the Joint Risk Matrix with juvenile offenders: A focus on gender and race/ethnicity. *Criminal Justice and Behavior*, *34*(3), 348-361.

*Schwalbe, C. S., Fraser, M. W., Day, S. H., & Arnold, E. M. (2004). North Carolina Assessment of Risk (NCAR). *Journal of Offender Rehabilitation*, *40*(1/2), 1-22.

*Schwalbe, C. S., Fraser, M. W., Day, S. H., & Cooley, V. (2006). Classifying juvenile offenders according to risk of recidivism: Predictive validity, race/ethnicity, and gender. *Criminal Justice and Behavior*, *33*(3), 305-324.

*Scott, C. K., Grella, C. E., Dennis, M. L., & Funk, R. R. (2014). Predictors of recidivism over 3 years among substance-using women released from jail. *Criminal Justice and Behavior, 41*(11), 1257-1289.

*Seifert, K., Phillips, S., & Parker, S. (2001). Child and Adolescent Risk for Violence (CARV): A tool to assess juvenile risk. *The Journal of Psychiatry and Law, 29*, 329-346.

*Sharkey, J. D., Furlong, M. J., Jimerson, S. R., & O'Brien, K. M. (2003). Evaluating the utility of a risk assessment to predict recidivism among male and female adolescents. *Education and Treatment of Children, 26*(4), 467-494.

*Siddiqi, Q. (2001). *Prediction of pretrial failure to appear and alternative pretrial release risk-classification schemes in New York City: A validation study.* New York, NY: New York City Criminal Justice Agency, Inc.

*Siddiqi, Q. (2004). CJA's new release-recommendation system: Looking beyond community ties to assess risk of flight. *Research Brief*, no. 5. New York, NY: New York City Criminal Justice Agency, Inc.

*Siddiqi, Q. (2004). *An evaluation of the new pretrial release recommendation system in New York City: Phase I of the post-implementation research*. New York, NY: New York City Criminal Justice Agency, Inc.

*Siddiqi, Q. (2005). *An evaluation of the new pretrial release recommendation system in New York City: Phase II of the post-implementation research.* New York, NY: New York City Criminal Justice Agency, Inc.

*Siddiqi, Q. (2006). *Predicting the likelihood of pretrial re-arrest for violent felony offenses and examining the risk of pretrial failure among New York City defendants: An analysis of 2001 dataset.* New York, NY: New York City Criminal Justice Agency, Inc.

*Siddiqi, Q. (2007). An evaluation of CJA's new release-recommendation system. *Research Brief*, no. 13. New York, NY: New York City Criminal Justice Agency, Inc.

*Siddiqi, Q. (2009). *Predicting the likelihood of pretrial failure to appear and/or re-arrest for a violent offense among New York City defendants: An analysis of the 2001 dataset.* New York, NY: New York City Criminal Justice Agency, Inc.

*Siddiqi, Q. (2009). Pretrial failure among New York City defendants. *Research Brief*, no. 19. New York, NY: New York City Criminal Justice Agency, Inc.

*Snowden, R. J., Gray, N. S., Taylor, J., & MacCulloch, M. J. (2007). Actuarial prediction of violent recidivism in mentally disordered offenders. *Psychological Medicine*, *37*, 1539-1549.

*Sokolov, E. (2002). *Offender Risk Assessment/Risk Management System* (Unpublished master's thesis). University of Manitoba, Winnipeg, Canada.

*Stansfield, R. Williams, K. R. (2014). Predicting family violence recidivism using the DVSI-R: Integrating survival analysis and perpetrator characteristics. *Criminal Justice & Behavior, 41*(2), 163-180.

Steinhart, D. (2006). *Juvenile detention risk assessment: A practice guide to juvenile detention reform. Baltimore*, MD: Annie E. Casey Foundation.

*Stewart, C. (2011). *Risk assessment of federal female offenders* (Unpublished master's thesis). University of Saskatchewan, Saskatoon, Canada.

*Storey, J. E., Kropp, P. R., Hart, S. D., Belfrage, H., & Strand, A. (2014). Assessment and management of risk for intimate partner violence by police officers using the Brief Spousal Assault form for the Evaluation of Risk. *Criminal Justice and Behavior, 41*(2), 256-271.

*Subramanian, R., Delaney, R., Roberts, S., Fishman, N., & McGarry, P. (2015). Incarceration's front door: The misuse of jails in America. New York: VERA Institute of Justice.

*Summers, R., & Loza, W. (2004). Cross validation of the Self-Appraisal Questionnaire (SAQ): A tool for assessing violent and nonviolent recidivism in Australian offenders. *Psychiatry, Psychology and Law*, *11*(2), 254-262.

Summers, C., & Willis, T. (2010). *Pretrial risk assessment: Research summary*. Arlington, VA: Bureau of Justice Assistance, U.S. Department of Justice.

*Thijssen, J., & de Ruiter, C. (2011) Identifying subtypes of spousal assaulters using the B-SAFER. *Journal of Interpersonal Violence, 26*(7), 1307-1321.

*Turner, S., Hess, J., Bradstreet, C., Chapman, S., & Murphy, A. (2013). *Development of the California Static Risk Assessment (CSRA): Recidivism risk prediction in the California Department of Corrections and Rehabilitation.* Irvine, CA: University of California, Irvine, Center for Evidence-Based Corrections.

*Turner, S., Hess, J., & Jannetta, J. (2009). Development of the California Static Risk Assessment Instrument (CSRA). Irvine, CA: University of California, Irvine, Center for Evidence-Based Corrections.

*Utah Criminal Justice Center. (2008). *Norming the Proxy risk assessment for Salt Lake County.* Salt Lake City, UT: University of Utah, Utah Criminal Justice Center.

*Valenzuela, Y. (2011). *Predictors of recidivism in a predominantly Hispanic sample of first-time juvenile offenders* (Unpublished master's thesis). The University of Texas at El Paso, El Paso, TX.

*van der Put, C. E. (2013). Youth Actuarial Risk Assessment Tool (Y-ARAT): The development of an actuarial risk assessment instrument for predicting general offense recidivism on the basis of police records. *Assessment*, *21*(3), 340-351.

*van der Put, C. E., van Vugt, E. S., Stams, G. J. J. M., Dekovic, M., & van der Laan, P. H. (2012). Short-term general recidivism risk of juvenile sex offenders: Validation of the Washington State Juvenile Court Prescreen Assessment. *International Journal of Offender Therapy and Comparative Criminology, 55*(11), 1374-1392.

*van der Put, C. E., van Vugt, E. S., Stams, G. J. J. M., Dekovic, M., & van der Laan, P. H. (2014). Predictive validity of the Washington State Juvenile Court Pre-Screen Assessment in the Netherlands: The development of a new scoring system. *Assessment, 21*(1), 92-107.

van der Ven, J. T. C. (2004). Assessment of risk and need factors and service use in diverted youth* (Unpublished doctoral dissertation). Carlton University, Ottawa, Ontario, Canada.

*VanNostrand, M. (2003). Assessing risk among pretrial defendants in Virginia: The Virginia Pretrial Risk Assessment Instrument. Richmond, VA: Virginia Department of Criminal Justice Services.

VanNostrand, M. (2007). *Legal and Evidence-Based Practices: Applications of Legal Principles, Laws, and Research to the Field of Pretrial Services*. Report for the Crime and Justice Institute and the National Institute of Corrections. Washington, DC: U.S. Department of Justice, National Institute of Corrections.

*VanNostrand, M., & Keebler, G. (2009). Pretrial risk assessment in the Federal Court. *Federal Probation, 73(2*), 3-29.

*VanNostrand, M., & Lowenkamp, C. T. (2013). *Assessing pretrial risk without a defendant interview*. Houston, TX: Laura and John Arnold Foundation.

*VanNostrand, M., & Rose, K. (2009). *Pretrial Risk Assessment in Virginia.* Richmond, VA: Virginia Department of Criminal Justice Services.

*Villeneuve, D. B., Oliver, N., & Loza, W. (2003). Cross-validation of the Self-Appraisal Questionnaire with a maximum-security psychiatric population. *Journal of Interpersonal Violence*, *18*(11), 1325-1334.

Vincent, G., Guy, L., & Grisso, T. (2012). *Risk assessment in juvenile justice: A guidebook for implementation*. Models for Change: Systems reform in juvenile justice.

*Virginia Criminal Sentencing Commission (2011). *2011 Annual report*. Richmond, VA: Author.

*Virginia Criminal Sentencing Commission (2011). *Re-validation of the nonviolent offender risk assessment instrument: Preliminary findings* (PDF document of PowerPoint slides handout). Richmond. VA: Author.

*Virginia Criminal Sentencing Commission (2011). *Re-validation of the nonviolent offender risk assessment instrument: Study Update* (PDF document of PowerPoint slides handout). Richmond. VA: Author.

*Virginia Criminal Sentencing Commission (2012). *2012 Annual report*. Richmond. VA: Author.

*Vitacco, M., Neumann, C., Robertson, A., & Durrant, S. (2002). Contributions of impulsivity and callousness in the assessment of adjudicated male adolescents: A prospective study. *Journal of Personality Assessment*, *78*(1), 87-103.

*Wakeling, H. C., Howard, P., & Barnett, G. (2011). Comparing the validity of the RM2000 scales and OGRS3 for predicting recidivism by internet sexual offenders. *Sexual Abuse: A Journal of Research and Treatment, 23*(1), 146-168.

*Walters, G. D. (1991). Predicting the disciplinary adjustment of maximum and minimum security prison inmates using the Lifestyle Criminality Screening Form. *International Journal of Offender Therapy and Comparative Criminology*, *35*(1), 63-71.

*Walters, G. D. (1995a). Factor structure of the Lifestyle Criminality Screening Form. *International Journal of Offender Therapy and Comparative Criminology*, *39*(2), 99-108.

*Walters, G. D. (1995b). Predictive validity of the Drug Lifestyle Screening Interview: A two-year follow-up. *American Journal of Drug and Alcohol Abuse*, *21*(2), 187-194.

*Walters, G. D. (2003). Predicting criminal justice outcomes with the Psychopathy Checklist and Lifestyle Criminality Screening Form: A meta-analytic comparison. *Behavioral Sciences & The Law*, *21*, 89-102.

*Walters, G. D. (2007). The latent structure of the criminal lifestyle: A taxometric analysis of the Lifestyle Criminality Screening Form and Psychological Inventory of Criminal Thinking Styles. *Criminal Justice and Behavior*, *34*(12), 1623-1637.

*Walters, G. D. (2011). Predicting recidivism with the Psychological Inventory of Criminal Thinking Styles and Level of Service Inventory-Revised: Screening Version. *Law and Human Behavior*, *35*, 211-220.

*Walters, G. D., & Chlumsky, M. L. (1993). The lifestyle criminality screening form and antisocial personality disorder: Predicting release outcome in a state prison sample. *Behavioral Sciences & The Law*, *11*, 111-115.

*Walters, G. D., & Schlauch, C. (2008). The Psychological Inventory of Criminal Thinking Styles and Level of Service Inventory-Revised: Screening Version as predictors of official and self-reported disciplinary infractions. *Law and Human Behavior*, *32*, 454-462.

*Walters, G. D., Revella, L., & Baltrusaitis II, W. J. (1990). Predicting parole/probation outcome with the aid of the Lifestyle Criminality Screening Form. *Psychological Assessment*, *2*(3), 313-316.

*Walters, G. D., White, T. W., & Denney, D. (1991). The Lifestyle Criminality Screening Form. *Criminal Justice and Behavior, 18*(4), 406-418.

*Warman-Hall, M. E. (2011). *A juvenile delinquency risk assessment: A predictive validity study* (Doctoral dissertation). Retrieved from ProQuest Dissertations and Theses database. (3486999)

*Washington State Institute for Public Policy. (1998). *Validation of the Washington State Juvenile Court Assessment: Interim report.* Olympia, WA: Author.

*Washington State Institute for Public Policy. (2004). *Assessing risk for re-offense: Validating the Washington State Juvenile Court Assessment.* Olympia, WA: Author.

*Washington State Institute for Public Policy. (2007). *Washington's Offender Accountability Act: Department of Corrections' Static Risk Assessment.* Olympia, WA: Author.

*Washington State Institute for Public Policy. (2014). *Predicting criminal recidivism: A systematic review of offender risk assessment in Washington state.* Olympia, WA: Author.

*Wei, Q., & Parsons, J. (2012). *Using administrative data to prioritize jail reentry services: Findings from the comprehensive transition planning project.* New York, NY: VERA Institute of Justice.

*Williams, K. (2012). Family violence risk assessment: A predictive cross-validation study of the Domestic Violence Screening Instrument Revised (DVSI-R). *Law and Human Behavior, 36*(2), 120-129.

*Williams, K. R., & Houghton, A. B. (2004). Assessing the risk of domestic violence reoffending: A validation study. *Law and Human Behavior, 28*(4), 437-455.

Williams, K. R., & Grant, S. R. (2006). Empirically examining the risk of intimate partner violence: The Revised Domestic Violence Screening Instrument (DVSI-R). Public Health Reports, 121*, 400-408.

*Withrow, B. (2003). Juvenile Detention Risk Assessment: An evaluation of the Sedgwick County instrument. *Corrections Compendium*, *28*(9), 1-4.

*Wormith, J. S., & Goldstone, C. S. (1984). The clinical and statistical prediction of recidivism. *Criminal Justice and Behavior, 11*(1), 3-34.

*Yacus, G. M. (1998). *Validation of the risk and needs assessment used in the classification for parole and probation of Virginia's adult criminal offenders* (Unpublished doctoral dissertation). Old Dominion University, Norfolk, VA.

*Zhang, S. X., Roberts, R. E. L., & Farabee, D. (2014). An analysis of prisoner reentry and parole risk using COMPAS and traditional criminal history measures. *Crime & Delinquency, 60*(2), 167-192.

Appendix A: Full Predictive Validity Classification Criteria

Table: A-1: Full Criteria for Determining Predictive Validity Effect Size Ratings

Effect Size Measure	Poor	Fair	Good	Excellent	Significant
	Predictive Validity Rating				
Cohen's d[a], any outcome	< .20	.20 – .49	.50 –.79	≥ .80	NA
Mean Cost Rating (MCR; Somer's d)[b], any outcome	< .10	.10 – .19	.20 –.29	≥ .30	NA
Area Under the Curve (AUC)[c], any outcome	<.56	.56 – .63	.64 – .71	≥ .72	NA
Correlation (r_{pb})[d,e]					
Adult Recidivism					
Any new arrest, ≤ 1 year follow, 28% BR	<.09	.09 – .22	.23 – .34	≥ .35	NA
Any new arrest, 1 – 3 year follow, 60% BR	<.10	.10 – .24	.25 – .36	≥ .37	NA
Any new arrest, >3 year follow, 73% BR	<.09	.09 – .22	.23 – .33	≥ .34	NA
Any new conviction, > 1 year follow, 15% BR	<.07	.07 – .18	.19 – .27	≥ .28	NA
Any new conviction, 1 – 3 year follow, 41% BR	<.10	.10 – .24	.25 – .36	≥ .37	NA
Any new conviction, > 3 year follow, 56% BR	<.10	.10 – .24	.25 – .36	≥ .37	NA
Technical/parole violation, under supervision, 50% BR	<.10	.10 – .24	.25 – .36	≥ .37	NA
New IPV, any time period, 35% BR[f]	<.10	.10 – .23	.24 – .36	≥ .37	NA
Failure to appear, pretrial, 14% BR	<.07	.07 – .17	.18 – .27	≥ .27	NA
New arrest, pretrial, 10% BR	<.06	.06 – .15	.16 – .23	≥ .24	NA
New violence arrest, any time period, 33% BR	<.09	.09 – .23	.24 – .35	≥ .36	NA
New violence conviction, any time period, 30% BR	<.09	.09 – .22	.23 – .34	≥ .35	NA
Juvenile Delinquency					
New arrest/referral, any time period[g], 36% BR	<.10	.10 – .23	.24 – .36	≥ .37	NA
New delinquent finding, any time period, 26% BR	<.09	.09 – .21	.22 – .33	≥ .34	NA
Failure to appear, pretrial, 5% BR	<.04	.04 – .11	.12 – .17	≥ .18	NA
New arrest, pretrial, 11% BR	<.06	.06 – .15	.16 – .24	≥ .25	NA
New violence arrest/conviction, any time period, 20% BR	<.08	.08 – .20	.21 – .30	≥ .31	NA
Odds Ratio (2X2)[h] | (2Xk)	<1.44	1.45 – 2.48	2.49 – 4.27	≥ 4.28	Yes/No

[a] Values from Cohen (1988)

[b] Values adopted from Desmarais & Singh (2013).

[c] AUC values were anchored to Cohen's d effect sizes using formulas available in Ruscio (2008) and were checked against tabled values in Rice and Harris (2005).

[d] All base rate adjustments to correlation coefficients were anchored to Cohen's d; calculations were performed by hand using formulas available in McGrath and Meyer (2006).

[e] BR=Base rate

[f] Includes violation of a protective order, threats, and physical injury

[g] There was little variability in the follow time of studies of juveniles, owing partly to the fact that they age out of the juvenile system. Because most studies involved 12 months of follow time, variations in base rates by follow time are not provided for juveniles as they were for adults.

[h] Values for a 2x2 contingency table were based on formulas from Borenstein, Hedges, Higgins, & Rothstein (2009).