

Criminal Justice and Behavior

<http://cjb.sagepub.com>

A Meta-Analysis of Juvenile Justice Risk Assessment Instruments: Predictive Validity by Gender

Craig S. Schwalbe

Criminal Justice and Behavior 2008; 35; 1367

DOI: 10.1177/0093854808324377

The online version of this article can be found at:
<http://cjb.sagepub.com/cgi/content/abstract/35/11/1367>

Published by:



<http://www.sagepublications.com>

On behalf of:

[International Association for Correctional and Forensic Psychology](#)

Additional services and information for *Criminal Justice and Behavior* can be found at:

Email Alerts: <http://cjb.sagepub.com/cgi/alerts>

Subscriptions: <http://cjb.sagepub.com/subscriptions>

Reprints: <http://www.sagepub.com/journalsReprints.nav>

Permissions: <http://www.sagepub.com/journalsPermissions.nav>

Citations <http://cjb.sagepub.com/cgi/content/refs/35/11/1367>

A META-ANALYSIS OF JUVENILE JUSTICE RISK ASSESSMENT INSTRUMENTS



Predictive Validity by Gender

CRAIG S. SCHWALBE

Columbia University School of Social Work

Juvenile justice systems have widely adopted risk assessment instruments to support judicial and administrative decisions about sanctioning severity and restrictiveness of care. A little explored property of these instruments is the extent to which their predictive validity generalizes across gender. The article reports on a meta-analysis of risk assessment predictive validity with male and female offenders. Nineteen studies encompassing 20 unique samples met inclusion criteria. Findings indicated that predictive validity estimates are equivalent for male and female offenders and are consistent with results of other meta-analyses in the field. The findings also indicate that when gender differences are observed in individual studies, they provide evidence for gender biases in juvenile justice decision-making and case processing rather than for the ineffectiveness of risk assessment with female offenders.

Keywords: risk assessment; juvenile justice system; juvenile courts; gender; prediction; recidivism

Juvenile justice systems employ actuarial risk assessment to support sanctioning and intervention decisions. Since the 1990s, risk assessment utilization has grown from just 33% of juvenile justice jurisdictions to more than 86% currently (Griffin & Bozynski, 2003; Towberman, 1992). During this era of risk assessment development and implementation, primary efforts have been aimed at developing risk assessment instruments with high levels of predictive validity and increasingly high levels of clinical utility. These efforts correspond with the increasingly structured approach to juvenile justice decision-making that includes such companion decision aids as structured needs assessment, mental health screening instruments, and semideterminant disposition matrices (Howell, 2003).

Gender equity is a little explored property of actuarial risk assessments. To the extent that risk assessment instruments inform critical judicial decisions such as the need for institutional care versus community-based sanctions, gender disparities may disproportionately disadvantage female offenders. This can happen when risk is overestimated for some groups and correspondingly underestimated for others. With increasing rates of female referral to the juvenile justice system during the past decade (Snyder & Sickmund, 2006), the problem of gender equity has taken on greater urgency. This article examines the predictive validity of actuarial risk assessment instruments across gender and reports the results of a meta-analysis of risk assessment predictive validity comparing male and female juvenile offenders.

AUTHOR'S NOTE: Please address all correspondence to Craig S. Schwalbe, PhD, Assistant Professor, Columbia University School of Social Work, 1255 Amsterdam Ave., New York, NY 10027; e-mail: css2109@columbia.edu.

CRIMINAL JUSTICE AND BEHAVIOR, Vol. 35 No. 11, November 2008 1367-1381

DOI: 10.1177/0093854808324377

© 2008 International Association for Correctional and Forensic Psychology

INTRODUCTION TO ACTUARIAL RISK ASSESSMENT

Risk assessment instruments classify delinquent youths into groups that vary in their likelihood of repeat offending. Most begin by assigning numerical scores to a set of risk factors that are individually associated with repeat offending. Common risk factors measured in risk assessment instruments include offending history, substance abuse, family problems, peer delinquency, and school-related problems (Hoge, 2002; Johnson, Wagner, & Matthews, 2002; Schwalbe, Fraser, & Day, 2007; Wiebush, Wagner, & Ehrlich, 1999). Then, taking advantage of the increased predictive validity possible through the cumulative risk property (Fraser, Kirby, & Smokowski, 2004), risk scores are summed to yield a raw risk score. In most instances, raw risk scores are used to group juveniles into ordinal risk classes ranging from low risk to high risk.

The development of risk assessment instruments can be described in a historical context (Bonta, 1996; Ferguson, 2002). First-generation risk assessment involved the impressionistic assessments of individual juvenile justice professionals without the aid of structured assessment devices. The weaknesses of this approach compared to formal statistical strategies for prediction and classification are, however, well established (Dawes, Faust, & Meehl, 1989; Grove & Meehl, 1996). Second- and third-generation risk assessments are grounded in the statistical association between a risk assessment instrument and repeat offending, although they vary in purpose.

Second-generation risk assessment instruments are limited to prediction and classification to inform sanctioning and supervision levels. That is, higher-risk youth merit more restrictive dispositions compared to lower-risk youths, irrespective of any treatment needs. Thus, second-generation instruments emphasize predictive validity over other potential uses, such as treatment planning. The Model Risk Assessment (Howell, 1995) is typical of second-generation risk assessment (Wiebush, 2002). This 11-item instrument includes three offense history variables and single item ratings of an array of social risk factors ranging from school discipline and attendance to peer delinquency to parental supervision. It has been actively promoted by the National Council on Crime and Delinquency (NCCD) along with the Office of Juvenile Justice and Delinquency Prevention (OJJDP) for use in conjunction with more comprehensive needs assessment instruments.

Third-generation risk assessment instruments inform treatment planning in addition to their classification role. The Youth Level of Service/Case Management Inventory (YLS/CMI; Hoge & Andrews, 2003) is typical of such instruments. The YLS/CMI is a 42-item instrument that measures risk factors in eight domains (offense history, family circumstances or parenting, education or employment, peer relations, substance abuse, leisure or recreation, personality or behavior, responsibility). Because of the more expansive scope of the YLS/CMI, it may support treatment decisions in addition to sanctioning severity and supervision levels. To support the treatment planning function, the manual includes an integrated case-planning protocol that links interventions with keystone risk factors that were raised in the assessment.

RISK ASSESSMENT WITH FEMALE OFFENDERS

Justice scholars debate the utility of actuarial risk assessment instruments designed specifically for female offenders (Odgers, Moretti, & Reppucci, 2005). As statistically derived prediction instruments, their effectiveness depends on consistent empirical relations between risk

and recidivism across gender. However, research with youthful offenders suggests at least three reasons that we should not expect gender consistency: gendered risk profiles, gendered sampling biases, and gendered juvenile court interventions. Each is addressed below.

Many scholars observe that research on risk and protective factors, the main body of literature informing actuarial risk assessment instruments, does not account for variation between male and female offenders. Rather, they assert that this literature is dominated by research with male offenders (Chesney-Lind & Sheldon, 1998; Lipsey & Derzon, 1998; Stouthamer-Loeber, Loeber, Homish, & Wei, 2001). Despite this, research with female offenders is growing. The literature shows that, on average, girls have higher levels of co-occurring problems and mental health disorders than boys and in particular are more likely to have suffered extensive trauma histories than their male counterparts (Gavazzi, 2006; Gavazzi, Yarcheck, & Chesney-Lind, 2006). Statistically, the presence of multiple problems should not in itself be problematic from a risk assessment predictive validity standpoint; indeed, statistical measures of predictive validity are maximized when variation in risk is high. However, predictive validity suffers if female-dominant risk factors are omitted from risk assessment instruments.

Moreover, research indicates that boys and girls may be responsive to different thresholds of risk. For example, the importance of family dysfunction as a risk factor for delinquency and recidivism is well documented (Cottle, Lee, & Heilbrun, 2001). However, it appears that lower levels of family dysfunction may accelerate risk for girls more rapidly than for boys (Hipwell & Loeber, 2006). Risks associated with peer delinquency are also well documented (Cottle et al., 2001), but evidence suggests that this effect may also vary by gender. Research by Piquero, Gover, MacDonald, and Piquero (2005) and Mears, Ploeger, and Warr (1998) suggests that girls may have more internalized protective mechanisms in place to buffer the impact of peer delinquency. Consequently, peer delinquency may be a more potent predictor of recidivism for boys than it is for girls. To the extent that risk assessment instruments hinge on severity or prevalence ratings of risk factors such as these, such instruments will maximize predictive validity for males while diminishing predictive validity for females.

Apparent support for this hypothesis was provided in two risk assessment studies with court involved juveniles (Funk, 1999; Schwalbe, Fraser, Day, & Cooley, 2006). Funk (1999) examined risk factors in a file review study of 1,030 court involved youths (28% female). Findings indicated that girls had higher levels of family-related risks (i.e., poor parent relations, running away, child abuse, parental criminality) than boys, whereas boys had higher peer-related risk factors. In terms of recidivism risk, Funk found that although there was some overlap in the offense history domain, social risk factors were segregated by gender: Delinquent peers and school behavior problems drove male recidivism risk, whereas child maltreatment and history of running away drove female recidivism risk. In a key study finding, results of a multivariate analysis with the combined sample mirrored results for males; risk factors associated with female recidivism risk were not represented.

Similarly, Schwalbe et al. (2006) found gender differences in their study of delinquent youths in North Carolina ($N = 9,534$). Their study found that compared to other groups, eight of nine risk factors measured by the North Carolina Assessment of Risk (NCAR) predicted recidivism for White and Black males ($n = 3,387$ and $n = 3,955$, respectively), but that only five predicted recidivism for Black females ($n = 1,237$) and one predicted recidivism for White females ($n = 955$). As in Funk's (1999) analysis, gender variation was masked in an analysis of the full sample in which eight risk factors were statistically significant predictors of recidivism.

Although Funk (1999) and Schwalbe et al. (2006) demonstrated that risk factors underlying recidivism risk may vary according to gender, alternative explanations are possible. One explanation with growing empirical support is the high potential for gendered sampling bias. Gendered decision-making practices that influence entry into the juvenile justice system threatens to introduce sampling biases into all risk assessment research (Chesney-Lind & Sheldon, 1998; Gaarder, Rodriguez, & Zatz, 2004; Leiber & Mack, 2003; MacDonald & Chesney-Lind, 2001; McGuire & Kuhn, 2003). Gendered decision-making practices affect risk assessment research in at least three ways. First, gendered decision-making practices that influence the likelihood of detection and referral to the juvenile justice system alter population parameters at entry into the juvenile justice system. For example, the juvenile justice system tendency to divert female offenders more readily compared to male offenders may create groups with nonoverlapping population parameters for recidivism risk factors, thereby diminishing the predictive validity of risk assessment instruments with female offenders (Leiber & Mack, 2003). Second, gendered decision-making practices introduce measurement error into the key outcome measure of risk assessment research—officially documented recidivism. As detection and referral patterns confound most definitions of recidivism, these practices bias statistical measures of the relationship between measured risk and recidivism. Third, it is not at all clear that the criterion variable in all risk assessment research—recidivism—has the same meaning for boys and girls. Some justice scholars argue that the juvenile justice system tends to criminalize behaviors in girls that are in fact coping responses to trauma and abuse (Goodkind, Ng, & Sarri, 2006; Simkins & Katz, 2002). To the extent that this is borne out by evidence, it suggests strongly that risk assessment instruments predict different phenomena in boys and girls, thereby casting doubt on the generalizability of risk assessment findings.

Finally, gendered decision-making practices about court interventions aimed at suppressing the possibility of recidivism—out-of-home placements and institutional commitments, for example—may reduce statistical measures of predictive validity more sharply for boys compared to girls. Studies by Leiber and Mack (2003) and Schwalbe, Hatcher, and Maschi (in press) found that adjudicated female offenders are more likely to receive harsh dispositions (i.e., change of placement, institutional placements, or waiver to adult court) compared to male offenders with similar offending profiles. These findings suggest that recidivism rates may be artificially suppressed for female offenders over periods commonly examined in risk assessment research. This hypothesis was confirmed by Schwalbe, Fraser, and Day (2007), who showed that a gender difference in the predictive validity of the Joint Risk Matrix (JRM) was due to variation in rates of out-of-home placement. Specifically, out-of-home placements for girls were targeted toward high risk youths, whereas out-of-home placements for boys were distributed randomly. When Schwalbe et al. (2007) introduced a statistical control for length of out-of-home placement, predictive validity estimates for the JRM among girls increased such that they were indistinguishable from those of boys.

The juvenile justice system has responded to these threats to predictive validity for female offenders in one of three ways. In some jurisdictions, gender is included in risk assessment instruments as a risk factor. Miller and Lin (2007) described such an instrument developed for the juvenile courts in New York City. Their study ($N = 730$) compared two versions of a locally developed actuarial risk assessment instrument, one with gender and one without, with two versions of a generic risk assessment instrument that were both gender neutral. The local instrument inclusive of gender had higher predictive validity than other models. As a risk factor, gender tends to award more risk to males, in effect formally

acknowledging higher rates of offending among males and suppressing total risk scores for females.

A second approach omits gender from the prediction models but optimizes risk factor weighting separately for boys and girls. A risk assessment project in Iowa demonstrated such a strategy (Huff & Prell, n.d.). Their study of 1,173 offenders (22% female) showed that a preliminary version of a locally developed risk assessment instrument overassessed recidivism risk for girls. That is, at similar levels of measured risk, recidivism rates were lower for female offenders than for male offenders. To correct this discrepancy, Huff and Prell modified the scale of the total risk score such that the classification of risk into heuristic categories (*medium low risk, medium high risk, high risk, very high risk*) differed for males and female offenders.

The most frequent approach reported in the literature ignores potential gender differences by adopting gender neutral risk assessment instruments irrespective of potential gender differences. For instance, the YLS/CMI, which has been validated across a range of juvenile justice settings, makes no adjustment for gender (Hoge & Andrews, 2003). An analysis of gender effects on YLS/CMI predictive validity found that predictive validity estimates did not vary by gender (Jung & Rawana, 1999). However, the NCAR has been subjected to three validation studies in North Carolina, each mirroring the results of Huff and Prell (n.d.) showing that girls reoffended at lower rates than boys at all levels of risk (Schwalbe et al., 2006, 2007; Schwalbe, Fraser, Day, & Arnold, 2004). The extent to which the gender disparities shown by Schwalbe and others (e.g., Sharkey et al., 2003) represent either the general effects of gender on risk assessment predictive validity or the idiosyncratic findings of specific jurisdictions has not been fully examined.

PRESENT STUDY

The present study is a meta-analysis of potential gender differences in risk assessment predictive validity. An earlier meta-analysis of juvenile justice risk assessment predictive validity identified 28 studies of 28 risk assessment instruments yielding 42 effect sizes (Schwalbe, 2007). Average predictive validity across all studies was modest ($r = .25$). Furthermore, the analysis showed that brief second-generation instruments had lower levels of predictive validity than longer, third-generation instruments. However, the study did not include a thorough analysis of the potential moderating effects of gender. The present study focused on a meta-analysis of research reports in which predictive validity was reported separately by gender. Following were the objectives of the study:

1. Compare the average predictive validity of risk assessment for recidivism in juvenile justice settings for male and female offenders.
2. Identify moderators to explain observed gender differences in the predictive validity of risk assessment.

METHOD

SAMPLE

The meta-analysis included reports of risk assessment predictive validity in which predictive validity estimates were reported separately by gender. Inclusion was restricted to

studies that estimated the predictive validity of a structured risk assessment instrument in a juvenile justice setting using a prospective longitudinal design. The predictive validity criterion variable was restricted to a measure of delinquency recidivism including rearrest and/or readjudication. Juvenile justice settings included traditional juvenile justice entities such as juvenile court probation departments and out-of-home placements sponsored and/or ordered by the juvenile court. Finally, included studies described their risk assessment instruments in enough detail so that they could be classified as described below and also provided sufficient information about predictive validity so that an effect size could be coded or calculated. To ensure a broad representation of risk assessment instruments, studies published in both peer review journals and studies published in other outlets (i.e., non-governmental research institutes, dissertations, government reports) were included.

A search of five electronic databases was conducted (National Criminal Justice Reference Service, Web of Science, Psycinfo, Sociological Abstracts, and Social Work Abstracts) using the following keywords: risk assessment, risk classification, risk prediction, delinquency, juvenile court, and juvenile justice. Studies published from the beginning of 1990 through the end of 2007 were eligible. To extend the search, additional searches were conducted in the Google search engine, from an examination of the National Center for Juvenile Justice State Juvenile Justice Profiles Web site (<http://www.ncjj.org/stateprofiles/>) and through contacts with private organizations and individuals who conduct risk assessment validation research. In total, 19 reports encompassing 20 unique samples were identified. Eleven reports were published in peer-review journals; five were the final reports of research organizations developing risk assessment instruments for individual jurisdictions (e.g., NCCD); two were reported in dissertations; one was a final report to an OJJDP-funded research study. These studies represent risk assessment research in the United States (11 studies), Canada (4 studies), Australia (2 studies), and the United Kingdom (1 study).

CODING

The author coded effect sizes, risk assessment characteristics, and methodological characteristics as described below.

Effect sizes. The literature on meta-analysis suggests two effect sizes for risk assessment studies: point-biserial correlation coefficients (r) and Area Under the Curve (AUC) from Receiver Operator Characteristic Curve analysis (Rice & Harris, 1995; Rosenthal, 1991; Rosenthal & DiMatteo, 2001). Although the AUC has favorable properties compared to traditional correlation coefficients (i.e., robust to variation in base rates, selection ratios, and truncated distributions), studies included in the analysis were neither uniform in their reporting of effect sizes, nor did they report all information necessary to calculate AUC statistics. However, all studies either reported the point-biserial correlation coefficient directly or provided information that would enable its hand calculation (Downie & Heath, 1983; Rice & Harris, 1995). Therefore, all analyses were conducted with point-biserial correlation coefficients.

Separate effect sizes were calculated for male and female offenders. The 20 samples from 19 studies yielded 49 effect sizes—25 for male offenders and 24 for female offenders. Where multiple risk assessment instruments were tested in the same samples (5 studies), effect sizes were coded for each risk assessment instrument.

Risk assessment characteristics. Risk assessment instruments were coded according to type. Type is a dichotomous indicator of second-generation versus third-generation risk assessment

type. As discussed earlier, second-generation instruments are either actuarial risk models or modeled after the OJJDP Model Risk Assessment (Howell, 1995, 2003). They measure common risk factors using fewer than 15 items. Risk assessment instruments that utilize alternative scoring protocols, that assume an underlying factor structure, or that measure constructs such as personality or functional impairment were coded in the third-generation category.

Methodological characteristics. Studies were coded for publication status (peer review vs. nonpeer review), sample size, sampling frame (general probation population vs. institutional population), definition of recidivism (new arrest/referral vs. new adjudication), base rate of recidivism, and length of follow-up. In addition, studies were coded for cross-validation. Cross-validation refers to the practice of reporting predictive validity effect sizes for samples independent of empirical procedures used to derive the instrument (Gottfredson & Tonry, 1987). In theory, cross-validation ensures that predictive validity estimates are robust to random sample variation.

ANALYSIS

A random effects model was used to estimate the average effect size and its statistical significance (Hunter & Schmidt, 2004). Fixed effects models assume that all study samples are representative of a single population and underestimate standard errors when this assumption is violated. In the present study, heterogeneity statistics supported the random effects model, $Q(49) = 636.8, p < .0001; I^2 = .92$. Hall and Brannick (2002) showed that the Hunter–Schmidt (S–H) random effects model (Hunter & Schmidt, 2000) produces credible confidence intervals compared to other methods and so was used for the present analysis. Confidence intervals were corrected for sampling error following Hunter and Schmidt (2004).

The file-drawer problem is endemic to meta-analysis where it is impossible to know with certainty that all relevant unpublished studies were obtained. A funnel plot, constructed to test for publication bias (Hunter & Schmidt, 2004), was inconclusive. Rosenthal (1991) recommended a procedure to describe the stability of findings based on the number of additional studies with null results needed to reduce the significance test of the average effect size to nonsignificance. This approach is based on a calculation of p values that are not commonly reported in prospective risk assessment studies. Rather, the present study adopts a similar approach by calculating a “failsafe N ” (Schwalbe, 2007). The failsafe N is the size of a simulated study with an average $r = .00$ that reduces the overall significance of the weighted mean effect size to $p > .05$. The S–H random effects model was used to simulate this additional study and calculate the required sample size.

Analysis of moderator effects explains variation in effect sizes (Lipsey, 2003). In the present study, potential moderators include gender, risk assessment characteristics, and methodological characteristics. Steel and Kammeyer-Mueller (2002) showed that bivariate statistics and ordinary least squares regression tend to overestimate the effects of moderator variables, whereas weighted least squares regression provides more accurate estimates. Thus, following Steel and Kammeyer-Mueller, weighted least squares regression was used to estimate the effects of moderator variables on effect sizes.

MISSING DATA

Six study reports failed to provide clear information about two methodological characteristics: definition of recidivism (four studies) and gender-specific recidivism base rates

(two studies). Attempts at completing this information by contacting the original study authors were not successful. To account for the effects of missing data, all multivariate analysis were conducted using two approaches: listwise deletion and multiple imputation (Allison, 2002). Listwise deletion provides unbiased estimates when data are missing completely at random, whereas multiple imputation provides unbiased estimates when data are missing at random, a less stringent assumption. Moreover, simulations with multiple imputation show that it is robust to modest departures from the missing at random assumption (Shafer & Graham, 2002). The imputation model included all study variables. Following convention, five data sets were imputed. Estimates from multivariate analyses were combined using SAS Proc MIANALYZE (SAS, 2007).

RESULTS

The 19 studies, encompassing 20 independent samples, yielded 25 effect sizes for males and 24 effect sizes for females ($k = 49$). Table 1 shows individual studies, sample characteristics, risk assessment instruments, and their associated effect sizes. Across all studies, risk of recidivism was assessed for 57,938 youths (31% female). The overall average effect size of $r = .27$ across all studies is comparable to the findings of Schwalbe (2007; $r = .25$). Effect sizes for male offenders range from $r = .13$ to $.44$; effect sizes for female offenders range from $r = .03$ to $.57$.

Tables 2 and 3 provide descriptive statistics for the methodological moderators. All methodological moderators which are exogenous to gender were similar across gender. Sixty-one percent ($k = 30$) of the effect sizes were obtained with brief, second-generation risk assessment instruments. About two thirds ($k = 33$) of the effect sizes were obtained from articles published in peer-review journals; 84% ($k = 41$) used samples derived from probation caseloads, whereas 16% ($k = 8$) used institutional samples; 61% ($k = 30$) defined recidivism as a new referral or complaint; three quarters ($k = 37$) employed cross-validation; and the average length of follow-up was 13 months. Average sample sizes for male offenders were over two times larger than female offenders ($n = 2002$ vs. $n = 942$) and average recidivism base rates among boys were 20% larger than among girls (40% vs. 32%).

Tables 2 and 3 also show the association between individual moderators and effect size estimates. Average effect sizes for the categorical moderators ranges from $r = .13$ to $r = .34$ and shows no appreciable variation by gender. Moreover, the only within-gender comparisons to achieve statistical significance ($p < .05$) was definition of recidivism; studies which defined recidivism as a new referral or complaint had higher effect sizes than studies that defined recidivism as a new adjudication. Like other categorical moderators, this effect was invariant across gender. However, pronounced differences emerged for the interval-level moderators shown in Table 3 (sample size, base rate, length of follow-up). The relationship between base-rate and effect sizes was clearly of a large magnitude for the male and female offenders ($r = .89$ and $r = .83$, respectively), whereas length of follow-up was inversely related to effect size ($r = -.39$ and $r = -.38$, respectively). The association between sample size and effect size differed by gender; whereas among boys there was a relatively stronger inverse relationship between sample size and effect size ($r = -.24$), among girls there was a slight positive relationship ($r = .09$). However, confidence intervals for these estimates overlapped substantially (95% $CI_{\text{male}} = -0.58$ to 0.17 ; 95% $CI_{\text{female}} = -0.32$ to 0.48) indicating that the differences were not statistically significant.

TABLE 1: Description of Studies Included in the Analysis

Reference	n	Percentage of Female	Base Rate:		Instrument	Effect Size:	
			Male	Female		Male	Female
Baker, Jones, Roberts, and Merrington (2003)	1,081	18.5	.53	.40	ASSET	.34	.30
Flores (2004)	1,697	21.3	NR	NR	YLS/CMI	.31	.32
Ilaqqa, Coulson, Lombardo, and Nutbrown (1999)	164	50.0	.73	.73	YO-LSI	.25	.20
Johnson, Wagner, and Mathews (2002)	2,911	30.7	.38	.25	NCCD-MO	.29	.28
Jung and Rawana (1999)	263	34.2	.31	.30	NCCD-MO-r	.31	.30
Krysiak and LeCroy (2002); LeCroy, Krysiak, and Palumbo (1998)	9,832	32.7	.56	.49	YLS/CMI	.37	.37
NCCD (2000)	954	28.0	.36	.21	ARNA-pp	.44	.48
Putins (2005)	458	<0.00	NR	NR	Alameda	.27	.30
Rowe (2002)	408	19.9	.67	.47	SECAPS	.32	.53
Schmidt, Hoge, and Gomez (2005)	107	37.1	.52	.38	YLS/CMI	.32	.57
Schwalbe, Fraser, Day, and Arnold (2004)	395	24.0	.43	.41	YLS/CMI	.25	.14
Schwalbe, Fraser, Day, and Cooley (2006)	9,534	23.0	.14	.12	NCAR	.16	.24
Schwalbe, Fraser, and Day (2007)	536	32.0	.24	.21	NCAR	.13	.08
Schwalbe (in press)	9,398	35.7	.36	.30	JRM	.29	.19
					ARNA-pp	.33	.27
					ARNA-cr	.25	.26
					ARNA-pp	.23	.27
					ARNA-cr	.26	.23
					ARNA-cr	.24	.25
					ARNA-pp	.25	.26
					ARNA-cr	.25	.27
					Orange	.26	.03
Sharkey, Furlong, Jimerson, and O'Brien (2003)	159	33.0	.51	.51	SBARA	.41	.53
Sharkey (2003)	378	35.0	.18	.16	YLS/CMI-AA	.29	—
Thompson and Pope (2005)	174	0.00	.40	NR	NCCD-VA	.29	—
Weibush, Wagner, and Ehrlich (1999)	1,260	21.6	.36	.27		.24	.20

Note. YLS/CMI = Youth Level of Service/Case Management Inventory; YO-LSI = Young Offender Level of Service Inventory; NCCD = National Council on Crime and Delinquency-Missouri; ARNA = Arizona Risk/Needs Assessment; SECAPS = Secure Care Psychosocial Screening; NCAR = North Carolina Assessment of Risk; JRM = Joint Risk Matrix; SBARA = Santa Barbara Assets and Risks Assessment.

TABLE 2: Distribution of Categorical Moderators and Average Effect Size by Gender

	Male			Female		
	<i>n</i>	<i>r</i>	95%CI	<i>n</i>	<i>r</i>	95%CI
Publication status						
Unpublished	8	.30	.27-.35	8	.30	.21-.49
Published	17	.27	.11-.43	16	.27	.01-.50
Sampling frame						
Probation	21	.26	.13-.44	20	.27	.11-.48
High risk	4	.30	— ^a	4	.20	.02-.43
Cross-validation						
No	6	.30	.25-.37	6	.30	.20-.51
Yes	19	.26	.12-.43	18	.27	.08-.44
Recidivism						
Referral	15	.28 ^a	.15-.40	15	.28 ^b	.12-.42
Adjudication	6	.17 ^a	.02-.56	5	.13 ^b	-.09-.72
Instrument type						
Second generation	15	.26	.11-.41	15	.27	.07-.42
Third generation	10	.32	.21-.42	9	.34	.14-.57

Note. Values that share superscripts are statistically different ($p < .05$).

a. 95%CI could not be calculated due to negative residual variance after corrections for sampling error were taken into account by the random effects model (Hunter & Schmidt, 2004).

TABLE 3: Distribution of Interval-Level Moderators and Association With Effect Size by Gender

	Male ($k = 25$)			Female ($k = 24$)		
	<i>k</i>	<i>M</i>	<i>r</i>	<i>k</i>	<i>M</i>	<i>r</i>
Sample size	25	2002	-.24	24	942	.09
Base rate	23	0.40	.89	22	0.32	.83
Length of follow-up (months)	25	13	-.39	24	13	-.38

Table 4 compares the average effect sizes across gender for the full sample of effect sizes and for risk assessment instruments which were studied in multiple samples. Across all studies, average effect sizes are virtually identical for male and female offenders ($r = .26$ and $r = .27$, respectively). Ninety-five percent confidence intervals vary widely from small effects ($r = .11$ and $r = .09$, respectively) to moderate effects ($r = .41$ and $r = .45$, respectively). Wide confidence intervals precluded statistically significant gender differences for both the YLS/CMI ($r = .32$ and $r = .40$) and the NCAR ($r = .14$ and $r = .09$). As with the overall sample, average predictive validity estimates were virtually identical for the Arizona Risk/Needs Assessment Instrument (ARNA; $r = .30$ and $r = .31$).

Table 5 presents results for the multivariate analysis of moderator effects in which four models are shown. Whereas Models 1 and 2 present the direct effects model and the interaction effects model using listwise deletion ($k = 37$), Models 3 and 4 replicate the analysis using full data from multiple imputation ($k = 49$). Across all models, more than 80% of the variance is explained. Models 1 and 3 show that male gender predicts lower predictive validity estimates when methodological moderators are statistically controlled. In addition, base rate and type of sample are also statistically significant in both models, whereas length of follow-up is statistically significant in the imputed data set only. In the final analysis, all moderator variables were interacted with gender with only the gender-by-base rate interaction achieving statistical

TABLE 4: Average Effect Sizes for the Full Sample and Discrete Risk Assessment Instruments by Gender

	Male				Female			
	<i>k</i>	<i>r</i>	95%CI	Failsafe <i>n</i>	<i>k</i>	<i>r</i>	95%CI	Failsafe <i>n</i>
Full Sample	25	.26	.11-.41	9,653	24	.27	.09-.45	3,884
YLS/CMI	4	.319	— ^a	240	3	.403	.17-.64	46
ARNA	4	.304	.14-.46	4,320	4	.307	.11-.50	1,856
NCAR	3	.138	.09-.19	1,782	3	.094	.05-.14	627

Note. YLS/CMI = Youth Level of Service/Case Management Inventory; ARNA = Arizona Risk/Needs Assessment; NCAR = North Carolina Assessment of Risk.

a. 95%CI could not be calculated due to negative residual variance after corrections for sampling error were taken into account by the random effects model (Hunter & Schmidt, 2004).

significance. The interaction term in Model 2 ($p < .05$) and in Model 4 ($p < .10$) shows that the effect of base rates on the predictive validity estimates of female offenders is larger than the effect of base rates on the predictive validity estimates of male offenders. Using Model 4 parameter estimates and holding all variables at their mean, the model predicted effect size is $r = .23$ for male offenders and $r = .25$ for female offenders. When recidivism base rates are low, say 15%, male and female effect sizes are indistinguishable ($r = .07$ and $r = .09$, respectively). When recidivism base rates are near the maximum found in this study, say 70%, differences between male and female effect sizes are comparably larger ($r = .43$ and $r = .59$, respectively).

DISCUSSION

Results of this study support the use of risk assessment instruments with both male and female offenders. The meta-analysis showed that risk assessment predictive validity did not vary appreciably by gender. Indeed, only when sample base rates of recidivism approach 70%, as might be seen in some residential settings that serve high risk youths, did model predicted gender differences become meaningful. Thus it appears that gender-specific risk assessments should not be required for most jurisdictions and programs that implement these decision aids. Interestingly, gender differences observed in the multivariate analysis were in an unexpected direction. An argument against the use of gender neutral risk assessment instruments implies that these instruments may be less effective with females than with males, in the sense that risk assessment instrument should overpredict recidivism for female offenders. This study supports the opposite conclusion—that risk assessment instruments are more effective with female offenders than with male offenders. Although the magnitude of these differences is small, this finding nevertheless informs the debate about the association of risk with recidivism across gender.

In general, we can conclude that the design of most risk assessment instruments, in conjunction with the cumulative risk property, leads to risk classifications with similar levels of predictive validity for male and female offenders. As statistical prediction devices, actuarial risk assessments do not assume an underlying causal process related to recidivism. Rather, they count risk factors irrespective of the specific factors that may or may not be present for an individual case. It appears that as constructed, we can infer that most risk assessment instruments measure an array of risk factors sufficient to identify risk for girls as well as for boys. However, these results do not imply that male and female risk patterns are necessarily invariant. It is possible, and indeed likely, that boys and girls of similar risk levels will require different intervention packages targeted to unique risk factors to reduce

TABLE 5: Multivariate Analysis of Moderators

	Listwise Deletion		Multiple Imputation	
	Model 1 (<i>k</i> = 37)	Model 2 (<i>k</i> = 37)	Model 3 (<i>k</i> = 49)	Model 4 (<i>k</i> = 49)
Intercept	.067	.030	.138	.099
Recidivism	.037	.033	.047	.043
Publication status	.012	.010	.012	.008
Second generation	-.007	-.005	-.028	-.034
Log(<i>N</i>)	.011	.008	.012	.012
Cross-validated	-.049	-.045	-.057	-.056
High risk sample	-.320*	-.362*	-.163*	-.161*
Base rate	.774***	.972***	.725***	.895***
Length of follow-up	-.006	-.005	-.008*	-.008*
Male	-.067**	.018	-.066***	.007
Male × Base Rate	—	-.270*	—	-.234 [†]
<i>R</i> ²	.87	.89	.82	.83

[†]*p* < .10. **p* < .05. ***p* < .01. ****p* < .001.

recidivism risk (Bloom, Owen, Deschenes, & Rosenbaum, 2002; Goodkind, 2005; Hipwell & Loeber, 2006).

Despite the widespread similarity in risk assessment predictive validity across gender, several individual studies found substantial gender differences (Schwalbe et al., 2004, 2006, 2007; Sharkey, Furlong, Jimerson, & O'Brien, 2003). Rather than casting doubt on the credibility of those studies, results of the present analysis rule out a provocative hypothesis—that widespread gender differences in risk assessment predictive validity account for their findings. Instead, these studies provide a window into the presence of gendered decision-making practices in the juvenile justice system. An alternative explanation is that gender-related sampling biases and intervention effects may be present in individual jurisdictions. As shown by Schwalbe et al. (2007), juvenile court decision-making practices can have a substantial impact on the observed psychometric properties of risk assessment instruments.

The strong effect of base rates on effect sizes observed in this study appears to be a statistical artifact. The challenge of accurately classifying risk of critical outcomes like recidivism increases when base rates vary from 50%. A visual inspection of the distributions of effect sizes suggests that many of the largest effect sizes for both males and females are clustered around samples that had base rates approximating 50%. As the large majority of study samples had base rates well below 50%, their predictive validity estimates were correspondingly suppressed, resulting in the strong positive correlation between base rates and effect sizes. Nonparametric effect size measures such as the AUC are more robust to variations in base rates. Therefore, this finding may not have been observed had other effect size measures been available.

Two study limitations should be noted. First, the author coded all research reports. Thus, it was not possible to establish the reliability of the coding scheme. Second, the sample size was small. This may be due to the file-drawer phenomenon and may also be due to the state of the literature on risk assessment. In Schwalbe's (2007) meta-analysis, 11 of 28 studies did not report predictive validity estimates separately by gender. Although some of these studies used samples that were too small to make reliable gender comparisons, others used large samples yet still did not report gender comparisons. It can also be noted that of the research reports included in the present meta-analysis, four studies involving research by a single author contributed 20 unique effect sizes (41%). Follow-up analysis with these studies (not reported here)

showed that overall effect sizes were smaller than for the full sample but that findings related to gender were the same. It can be stated with some degree of confidence, then, that reporting predictive validity by gender is not yet standard practice in risk assessment research.

Within the bounds of these limitations, the present study suggests two implications for policy and research. First, because risk assessment predictive validity appears, on average, to be similar for boys and girls, this study supports the use of risk assessment instruments in varied juvenile justice agencies with male and female offenders. Indeed, risk assessment classifications of risk for recidivism may contribute meaningfully to judicial decisions and agency practices related to sanctioning severity and level of care for male and for female offenders. As a first step toward implementing risk assessment for justice-involved youths, results of this study suggest starting first with a well-validated assessment instrument that would be applied to youths of both genders. However, this study does not lend evidence to the irrelevance of gender to risk assessment predictive validity. Rather, the second implication of this study is that attention to gender in risk assessment research should continue. Gender differences, when they appear, can open a window into the presence of gender biases in juvenile justice decision-making. By routinely testing for gender differences, risk assessment validation studies can expose gender biases that can become the focal point for ongoing research and policy interventions. In this way, risk assessment instruments, and the research that supports them, can serve to increase, rather than undermine, gender equity in the juvenile justice system.

REFERENCES

- Note. Asterisks identify studies that were included in the meta-analysis.
- Allison, P. D. (2002). *Missing data*. Thousand Oaks, CA: Sage.
- *Baker, K., Jones, S., Roberts, C., & Merrington, S. (2003). *The evaluation of the validity and reliability of the Youth Justice Board's assessment for young offenders: Findings from the first two years of ASSET*. Oxford, UK: Centre for Criminological Research, University of Oxford.
- Bloom, B., Owen, B., Deschenes, E. P., & Rosenbaum, J. (2002). Moving toward justice for female juvenile offenders in the new millennium: Modeling gender-specific policies and programs. *Journal of Contemporary Criminal Justice*, 18, 37-56.
- Bonta, J. (1996). Risk-needs: Assessment and treatment. In A. T. Harland (Ed.), *Choosing correctional options that work: Defining the demand and evaluating the supply* (pp. 18-32). Thousand Oaks, CA: Sage.
- Chesney-Lind, M., & Sheldon, R. G. (1998). *Girls, delinquency, and juvenile justice* (2nd ed.). Belmont, CA: West/Wadsworth.
- Cottle, C. C., Lee, R. L., & Heilbrun, K. (2001). The prediction of criminal recidivism in juveniles: A meta-analysis. *Criminal Justice and Behavior*, 34, 367-394.
- Dawes, R. M., Faust, D., & Meehl, P. E. (1989). Clinical vs. actuarial judgment. *Science*, 243, 1668-1674.
- Downie, N. M., & Heath, R. W. (1983). *Basic statistical methods* (5th ed.). New York: Harper & Row.
- Ferguson, J. L. (2002). Putting the "what works" research into practice: An organizational perspective. *Criminal Justice and Behavior*, 29, 472-492.
- *Flores, A. W., Travis, L. F., & Latessa, E. J. (2004). *Case classification for juvenile corrections: An assessment of the Youth Level of Service/Case Management Inventory (YLS/CMI), final report*. Washington, DC: National Institute of Justice.
- Fraser, M. W., Kirby, L. D., & Smokowski, P. R. (2004). Risk and resilience in childhood. In M. W. Fraser (Ed.), *Risk and resilience in childhood: An ecological perspective* (2nd ed., pp. 13-66). Washington, DC: NASW Press.
- Funk, S. J. (1999). Risk assessment for juveniles on probation: A focus on gender. *Criminal Justice and Behavior*, 26, 44-68.
- Gaarder, E., Rodriguez, N., & Zatz, M. (2004). Criers, liars, and manipulators: Probation officers' views of girls. *Justice Quarterly*, 21, 547-578.
- Gavazzi, S. M. (2006). Gender, ethnicity, and the family environment: Contributions to assessment efforts within the realm of juvenile justice. *Family Relations*, 55, 190-199.
- Gavazzi, S. M., Yarcheck, C. M., & Chesney-Lind, M. (2006). Global risk indicators and the role of gender in a juvenile detention sample. *Criminal Justice and Behavior*, 33, 597-612.
- Goodkind, S. (2005). Gender-specific services in the juvenile justice system: A critical examination. *Affilia: Journal of Women and Social Work*, 20, 52-70.
- Goodkind, S., Ng, I., & Sarri, R. C. (2006). The impact of sexual abuse in the lives of young women involved or at risk of involvement with the juvenile justice system. *Violence Against Women*, 12, 456-477.

- Gottfredson, D. M., & Tonry, M. (Eds.). (1987). *Prediction and classification: Criminal justice decision making*. Chicago: The Chicago University Press.
- Griffin, P., & Bozynski, M. (2003). *National overviews: State juvenile justice profiles*. Retrieved November 5, 2003, from <http://www.ncjj.org/stateprofiles/>
- Grove, W. M., & Meehl, P. E. (1996). Comparative efficiency of informal (subjective, impressionistic) and formal (mechanical, algorithmic) prediction procedures: The clinical-statistical controversy. *Psychology, Public Policy, and Law*, 2, 293-323.
- Hall, S. M., & Brannick, M. T. (2002). Comparison of two random-effects methods of meta-analysis. *Journal of Applied Psychology*, 28, 377-389.
- Hipwell, A. E., & Loeber, R. (2006). Do we know which interventions are effective for disruptive and delinquent girls? *Clinical Child and Family Psychology Review*, 9, 221-255.
- Hoge, R. D. (2002). Standardized instruments for assessing risk and need in youthful offenders. *Criminal Justice and Behavior*, 29, 380-396.
- Hoge, R. D., & Andrews, D. A. (2003). *The youth level of service/case management inventory (YCS/CMI): Intake manual and item scoring key*. Carleton University, Ottawa, Ontario, Canada.
- Howell, J. C. (1995). *Guide for implementing the comprehensive strategy for serious, violent, and chronic juvenile offenders*. Washington, DC: Office of Juvenile Justice and Delinquency Prevention.
- Howell, J. C. (2003). *Preventing & reducing juvenile delinquency: A comprehensive framework*. Thousand Oaks, CA: Sage.
- Huff, D., & Prell, L. (n.d.). *The proposed Iowa juvenile court intake risk assessment*. Retrieved January 30, 2007, from <http://www.state.ia.us/government/dhr/cjpp/images/pdf/before99/riskrpt6.pdf>
- Hunter, J. E., & Schmidt, F. (2004). *Methods of meta-analysis* (2nd ed.). Thousand Oaks, CA: Sage.
- Hunter, J. E., & Schmidt, F. L. (2000). Fixed effects vs. random effects meta-analysis models: Implications for cumulative research knowledge. *International Journal of Selection and Research*, 8, 275-292.
- *Iacqua, G. E., Coulson, G. E., Lombardo, D., & Nutbrown, V. (1999). Predictive validity of the Young Offender Level of Service Inventory for criminal recidivism of male and female young offenders. *Psychological Reports*, 84, 1214-1218.
- *Johnson, K., Wagner, D., & Matthews, T. (2002). *Missouri juvenile risk assessment revalidation report*. Madison, WI: National Council on Crime and Delinquency.
- *Jung, S., & Rawana, E. P. (1999). Risk and need assessment of juvenile offenders. *Criminal Justice and Behavior*, 26, 69-89.
- *Krysiak, 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, 71-81.
- *LeCroy, C. W., Krysiak, J., & Palumbo, D. (1998). *Empirical validation of the Arizona risk/needs instrument and assessment process*. Tucson, AZ: LeCroy & Milligan Associates.
- Leiber, M. J., & Mack, K. Y. (2003). The individual and joint effects of race, gender, and family status on juvenile justice decision-making. *Journal of Research in Crime and Delinquency*, 40, 34-70.
- Lipsey, M. W. (2003). Those confounded moderators in meta-analysis: Good, bad, and ugly. *The Annals of the American Academy of Political and Social Science*, 587, 69-81.
- Lipsey, M. W., & Derzon, J. H. (1998). Predictors of violent or serious delinquency in adolescence and early adulthood. In R. L. D. P. Farrington (Ed.), *Serious & violent juvenile offenders* (pp. 86-105). Thousand Oaks, CA: Sage.
- MacDonald, J. M., & Chesney-Lind, M. (2001). Gender bias and juvenile justice revisited: A multiyear analysis. *Crime and Delinquency*, 47, 173-195.
- McGuire, M. D., & Kuhn, K. E. (2003). Gender and the likelihood of being securely detained for contempt. *Juvenile & Family Court Journal*, 54, 17-32.
- Mears, D. P., Ploeger, M., & Warr, M. (1998). Explaining the gender gap in delinquency: Peer influence and moral evaluations of behavior. *Journal of Research in Crime and Delinquency*, 35, 251-266.
- Miller, J., & Lin, J. (2007). Applying a generic juvenile risk assessment instrument to a local context: Some practical and theoretical lessons. *Crime & Delinquency*, 53, 552-580.
- *NCCD. (2000). *Alameda County placement risk assessment validation: Draft of final report*. Oakland, CA: Author.
- Ogders, C. L., Moretti, M. M., & Reppucci, N. D. (2005). Examining the science and practice of violence risk assessment with female adolescents. *Law and Human Behavior*, 29, 7-27.
- Piquero, N. L., Gover, A. R., MacDonald, J. M., & Piquero, A. R. (2005). The influence of delinquent peers on delinquency: Does gender matter? *Youth & Society*, 36, 251-275.
- *Putnins, A. L. (2005). Assessing recidivism risk among young offenders. *Australian and New Zealand Journal of Criminology*, 38, 324-339.
- Rice, M. E., & Harris, G. T. (1995). Violent recidivism: Assessing predictive validity. *Journal of Consulting and Clinical Psychology*, 63, 737-748.
- Rosenthal, R. (1991). *Meta-analytic procedures for social research* (2nd ed.). Newbury Park, CA: Sage.
- Rosenthal, R., & DiMatteo, M. R. (2001). Meta-analysis: Recent developments in quantitative methods for literature reviews. *Annual Review of Psychology*, 52, 59-82.
- *Rowe, R. C. (2002). *Predictors of criminal offending: Evaluating measures of risk/needs, psychopathy, and disruptive behaviour disorders*. Unpublished doctoral dissertation, Carleton University, Ottawa, Ontario, Canada.
- SAS Institute, Inc. (2007). *SAS 9.1*. Cary, NC: Author.
- *Schmidt, F., Hoge, R. D., & Gomez, L. (2005). Reliability and Validity Analyses of the Youth Level of Service/Case Management Inventory. *Criminal Justice and Behavior*, 32, 329-344.

- Schwalbe, C. S. (2007). A meta analysis of juvenile justice risk assessment predictive validity. *Law and Human Behavior, 31*, 449-462.
- *Schwalbe, C. S. (in press). Risk assessment stability: A revalidation study of the Arizona Risk/Needs Assessment Instrument. *Research on Social Work Practice*.
- *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*, 348-361.
- *Schwalbe, C. S., Fraser, M. W., Day, S. H., & Arnold, E. M. (2004). North Carolina Assessment of Risk (NCAR): Reliability and Predictive Validity with Juvenile Offenders. *Journal of Offender Rehabilitation, 40*, 1-22.
- Schafer, J. L. and Graham, J. W. (2002). Missing data: Our view of the state of the art. *Psychological Methods, 7*, 147-177.
- *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*, 305-324.
- Schwalbe, C. S., Hatcher, S. S., & Maschi, T. M. (in press). The effects of treatment needs and prior social services utilization on juvenile court decision making. *Social Work Research*.
- *Sharkey, J. D. (2003). *Examining the relationship between risk and protective factors and juvenile recidivism for male and female probationers*. Doctoral dissertation, University of California at Santa Barbara.
- *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*, 467-494.
- Simkins, S., & Katz, S. (2002). Criminalizing abused girls. *Violence Against Women, 8*, 1474-1499.
- Snyder, H. N., & Sickmund, M. (2006). *Juvenile offenders and victims: 2006 National Report*. Washington, DC: U.S. Department of Justice, Office of Justice Programs, Office of Juvenile Justice and Delinquency Prevention.
- Steel, P. D., & Kammeyer-Mueller, J. D. (2002). Comparing meta-analytic moderator estimation techniques under realistic conditions. *Journal of Applied Psychology, 87*, 96-111.
- Stouthamer-Loeber, M., Loeber, R., Homish, D. L., & Wei, E. (2001). Maltreatment of boys and the development of disruptive and delinquent behavior. *Development and Psychopathology, 13*, 941-955.
- *Thompson, A. P., & Pope, Z. (2005). Assessing juvenile offenders: Preliminary data for the Australian Adaptation of the Youth Level of Service/Case Management Inventory (Hoge & Andrews, 1995). *Australian Psychologist, 40*, 207-214.
- Towberman, D. B. (1992). A national survey of juvenile risk assessment. *Family & Juvenile Court Journal, 43*, 61-67.
- Wiebush, R. G. (2002). *Graduated sanctions for juvenile offenders: A program model and planning guide*. Reno, NV: Juvenile Sanctions Center: National Council of Juvenile and Family Court Judges.
- *Wiebush, R., Wagner, D., & Ehrlich, J. (1999). *Development of an empirically-based risk assessment instrument: For the Virginia Department of Juvenile Justice final report*. Madison, WI: National Council on Crime and Delinquency.

Craig Schwalbe is an assistant professor at the Columbia University School of Social Work. His research examines juvenile justice approaches to the assessment and treatment of youths at high risk for chronic delinquency.



Criminal Justice and Behavior (CJB) is pleased to offer the opportunity to earn continuing education credits online for this article. Please follow the access instructions below. The cost is \$9 per unit, and a credit card number is required for payment.

Access for IACFP members:

- Go to the IACFP homepage (www.ia4cfp.org) and log in to the site using your email address
- Click on the Publications heading in the sidebar on the left-hand side of the page or on the Publications icon at the bottom of the page
- Click on the *CJB* cover image located at the bottom of the page. You will be prompted for your user name and password to log in to the journal
- Once you have logged in to *CJB*, select the appropriate issue and article for which you wish to take the test, then click on the link in the article to be directed to the test (*note: only marked articles are eligible for continuing education credits*)

Access for institutional subscribers:

- Log in through your library or university portal to access *CJB*
- Select the appropriate issue and article, then click on the link in the article to be directed to the test (*note: only marked articles are eligible for continuing education credits*)

If you are not a member of the IACFP and your institution does not subscribe to *CJB*, individual articles (and corresponding test links within the articles) are available on a pay-per-view basis by visiting <http://cjb.sagepub.com>. However, membership in the IACFP provides you with unlimited access to *CJB* and many additional membership benefits! Learn more about becoming a member at www.ia4cfp.org.