RISK ASSESSMENT WITH YOUNG OFFENDERS

A Meta-Analysis of Three Assessment Measures

MARK E. OLVER
University of Saskatchewan

KEIRA C. STOCKDALE
Saskatoon Health Region

J. STEPHEN WORMITH
University of Saskatchewan

The current investigation is a meta-analysis of the predictive accuracy of three well-known forensic instruments used to appraise risk with young offenders: youth adaptations of the Level of Service Inventory and Psychopathy Checklist and the Structured Assessment of Violence Risk for Youth. Through several avenues, 49 potentially suitable published and unpublished studies (across 44 samples representing 8,746 youth) were identified and evaluated for inclusion. Predictive accuracy for general, nonviolent, violent, and sexual recidivism was examined for the three sets of measures. Mean weighted correlations for each of the three measures were significant in the prediction of general, nonviolent, and violent recidivism, with no single instrument demonstrating superior prediction. Separate analyses of specific young offender groups further supported the predictive accuracy of youth adaptations of the Level of Service Inventory among male, female, Aboriginal, and non-Aboriginal youth. Implications regarding the utility of young offender risk measures for enhancing clinical service provision with youth clientele are discussed.

Keywords: risk assessment; young offender; recidivism; prediction; meta-analysis

Assessing risk for general and violent recidivism is a common task for psychologists and other mental health professionals who work with youth involved in the juvenile and criminal justice systems.[1] Such assessments, in turn, serve important functions. One obvious purpose is promotion of public safety (Borum, 2000). For instance, a youth identified as posing a high probability of future harm to others might be an appropriate candidate for a secure custody disposition (to protect the public) and/or intensive treatment (to reduce risk). Youth court justices also routinely make risk assessment referrals, presumably because such evaluations can be helpful in assisting sentencing decisions (e.g., type of disposition, sentence length) or imposing special conditions (Hoge, 1999, 2002). Moreover, risk assessments may be used to formulate recommendations for services (Borum, 2000; Grisso, 1998; Hoge, 2001, 2002; Hoge & Andrews, 1996), and a comprehensive assessment

AUTHORS' NOTE: *We would like to thank (in alphabetical order) Kevin Douglas, Ed Hilterman, Candice Odgers, and Martin Rieger for assisting us by providing data from unpublished, difficult to access, or very recent resources and the two reviewers for their invaluable suggestions and feedback that served to improve the article. The views expressed are those of the authors and do not necessarily represent the views of the University of Saskatchewan or the Saskatoon Health Region.* *Correspondence concerning this article should be addressed to Mark E. Olver, Department of Psychology, University of Saskatchewan, Arts Building Room 154, 9 Campus Dr., Saskatoon, Saskatchewan, Canada S7N 5A5; e-mail: mark.olver@usask.ca.*

CRIMINAL JUSTICE AND BEHAVIOR, Vol. 36 No. 4, April 2009 329-353
DOI: 10.1177/0093854809331457

TABLE 1: Summary of Meta-Analytic Investigations on the Predictive Accuracy of Young Offender Risk Assessment Measures: Results From Aggregate Samples and Among Genders

Meta-Analysis	*Sample*	*Instrument*	k	n	*Mean Weighted Effect Size*	
Edens, Campbell, and Weir (2007)[a]	Both genders	PCL-YV	20	2,787	General	.24
			14	2,067	Violent	.25
			4	654	Sexual	.07
	Males	PCL-YV	19	2,482	General	.25
			14	207	Violent	.26
	Females	PCL-YV	5	1,870	General	.13
			5	228	Violent	.10
Schwalbe (2007)[b]	Both genders	PCL-YV	3	455	General	.70
		YLS/CMI	11	3,265		.64
		NCAR	3	10,534		.60
Schwalbe (2008)[a]	Male	YLS/CMI	4	772	General	.32
		ARNA	4	24,565		.30
		NCAR	3	8,005		.14
	Female	YLS/CMI	3	204	General	.40
		ARNA	4	12,925		.31
		NCAR	3	2,460		.09

Note. PCL-YV = Psychopathy Checklist–Youth Version; YLS/CMI = Youth Level of Service/Case Management Inventory; NCAR = North Carolina Risk Assessment; ARNA = Arizona Risk/Needs Assessment.
a. Effect size reported as *r*.
b. Effect size reported as area under the curve.

can identify relevant risk factors to be targeted for treatment, appropriate intervention strategies, and responsivity issues that may affect the treatment process (e.g., client motivation). Finally, risk assessments may also assist decision making regarding the management of youth in juvenile justice facilities (e.g., evaluating risk of harm to staff or resident youth, identifying possible security restrictions; see Holsinger, Lowenkamp, & Latessa, 2006).

In recent decades, the research and practice of risk assessment has moved away from unstructured clinical judgment and toward more evidence-based, structured approaches (Bonta, 2002). Although this advancement began in risk assessment with adult offenders, an increasing number of measures have been developed for assessing risk for violent and general recidivism in youth (Borum, 2000). These instruments are generally youth-adapted versions of adult measures with modifications based on developmental considerations unique to the young offender population. Research on these measures has varied in methodological quality and the strength of psychometric findings, but the existing findings suggest that youth risk instruments are capable of predicting young offender recidivism with what appears to be at least a comparable degree of accuracy to their adult counterparts (Edens, Campbell, & Weir, 2007; Gendreau, Goggin, & Smith, 2002; Schwalbe, 2007, 2008). Only recently has the young offender risk assessment literature grown to a sufficient magnitude to permit meta-analytic investigation, with three such meta-analyses having been conducted to synthesize the findings (summarized in Table 1). Although these studies have made an important contribution, each investigation has its own set of limitations, and the volume of literature has continued to grow. Three of the most prominent and well-researched young offender measures are described as follows, two of which are youth adaptations of adult instruments.

YOUTH LEVEL OF SERVICE/CASE MANAGEMENT INVENTORY (YLS/CMI) AND OTHER LEVEL OF SERVICE INVENTORY (LSI) ADAPTATIONS

The Revised LSI (Andrews & Bonta, 1995) and the Level of Service/Case Management Inventory (Andrews, Bonta, & Wormith, 2004) are risk and need measures for adults designed to assess risk for general recidivism, identify targets for intervention, and guide future supervision and planning. The LSI was developed from a general personality and social psychological perspective of crime (Andrews & Bonta, 2003), embodied in the Big Four covariates of criminal conduct—antisocial attitudes, antisocial associates, antisocial personality, and a history of antisocial behavior (the constellation is sometimes referred to as the Central Eight, with the inclusion of the needs areas leisure and recreation, family and marital, substance abuse, and employment and education). These covariates are linked to the origin of criminal behavior (and are hence called criminogenic needs), and services directed toward these areas of risk and need might reduce antisocial behavior (Andrews et al., 1990).

Several variants of the family of LSI measures have been developed for youth, including the YLS/CMI (Hoge & Andrews, 2003) and its screening version (YLS/CMI:SV; Hoge & Andrews, 2001), the Level of Service Inventory–Saskatchewan Edition (LSI-SK; Andrews, Bonta, & Wormith, 2001), and the Young Offender LSI (YO-LSI; Shields & Simourd, 1991). The YLS/CMI is described in further detail as a representative illustration of these youth adaptations given the substantial amount of overlap among them. The YLS/CMI is a 42-item clinician rated risk and needs measure developed (as with its adult counterpart) to assess risk for general recidivism, identify targets for intervention, and guide community supervision. Rated on the basis of interview and collateral information (e.g., presentence reports, school records), the YLS/CMI is subdivided into eight criminogenic areas: Prior and Current Offenses/Dispositions, Family Circumstances/Parenting, Education/Employment, Peer Relations, Substance Abuse, Leisure/Recreation, Personality/Behavior, and Attitudes/Orientation. Items are summed to arrive at a risk rating of low (0 to 8), medium (9 to 22), high (23 to 34), or very high (35 to 42). Separate norms exist for male and female young offenders, and approximately half of the youth in the original normative sample are of Aboriginal descent.

Psychometric data on youth-adapted versions of the LSI and the YLS/CMI, in particular, are starting to accumulate. Evaluative research with young offenders has suggested that youth-adapted versions of the LSI are suitable risk and need assessment instruments, and moderate to high predictive accuracy for general and violent recidivism has been reported (e.g., Catchpole & Gretton, 2003; Gossner & Wormith, 2007; Schmidt, Hoge, & Gomes, 2005).

PSYCHOPATHY CHECKLIST–YOUTH VERSION (PCL-YV)

The Hare Psychopathy Checklist–Revised (PCL-R; Hare, 1991, 2003) is a 20-item symptom construct rating scale designed to assess the personality and behavioral features of psychopathy in adults. Inspired originally from the work of Hervey Cleckley (1976), which included a list of 16 attributes descriptive of psychopathy, the PCL-R was developed to enable clinicians and researchers to obtain a valid and reliable diagnosis of this syndrome. Although not originally intended for this purpose, the PCL-R is commonly used in forensic evaluations to assess risk, given its association with violent and general recidivism (e.g., Gendreau et al., 2002).

The PCL-YV (Forth, Kosson, & Hare, 2003) is a downward extension of the PCL-R designed to assess personality characteristics and behaviors of psychopathy in adolescents. It also comprises 20 items with some modification in item wording and content. Factor analytic research on the PCL-YV has supported a four-factor model for the tool (Neumann, Kosson, Forth, & Hare, 2006): Interpersonal (e.g., deceitfulness, manipulation, grandiosity, impression management), Emotional (e.g., callousness, lack of remorse, shallow emotions, failure to accept responsibility), Behavioral (e.g., impulsivity, irresponsibility, stimulation seeking, lack of goals), and Antisocial (e.g., criminal versatility, supervision violations, serious juvenile delinquency, poor anger controls). Each item is rated on a 3-point scale: 0 (*not present*), 1 (*partially present*), 2 (*present*). Although (like its adult counterpart) not originally designed to assess risk, the PCL-YV has been increasingly drawn into clinical practice as a risk assessment measure. This development has been greeted with caution and criticism by researchers and clinicians expressing concerns about potential misuses of the tool and the construct of psychopathy applied to juveniles (Edens, Skeem, Cruise, & Cauffmann, 2001). These concerns notwithstanding, the extant literature, including the results of a meta-analysis (Edens et al., 2007), has shown the PCL measures to have good predictive accuracy for general (r_w = .24) and violent recidivism (r_w = .25).

STRUCTURED ASSESSMENT OF VIOLENCE RISK IN YOUTH (SAVRY)

Finally, the SAVRY (Borum, Bartel, & Forth, 2006) was specifically designed to assess risk of future violence for youth. A reasonable parallel to draw within the adult risk assessment field would be the Historical Clinical Risk–20 (HCR-20; Webster, Douglas, Eaves, & Hart, 1997), a rating scheme designed to assess violence risk drawing on various static and dynamic risk factors linked to violence in adult offenders. The HCR-20 is an example of structured professional judgment (SPJ), in which the professional arrives at an appraisal of violence risk (a summary rating of low, moderate, or high) based on the constellation of item ratings. In contrast to actuarial-based approaches, items are not summed to arrive at a numeric total score (and corresponding risk cutoff), although this approach has been employed to conduct psychometric analyses of the tool. Research supports the predictive accuracy of both HCR-20 numeric scores and SPJ summary ratings for future violence (Douglas, Yeomans, & Boer, 2005).

The SAVRY is a 30-item risk assessment guide consisting of 10 Historical risk factors (e.g., prior violence, poor school achievement), 6 Social/Contextual risk factors (e.g., peer delinquency, lack of support), 8 Individual/Clinical risk factors (e.g., anger management problems, negative attitudes, low empathy), and 6 Protective factors (e.g., prosocial involvement, strong social support). The 24 risk items are rated as low, moderate, or high in terms of their seriousness and relationship to violence for the youth. The 6 Protective factors are rated as either present or absent. As with the HCR-20, the SAVRY draws on SPJ, and the items are not summed to arrive at a risk rating. Rather, the clinician weighs and evaluates the collection and pattern of risk factors to arrive at a global violence risk rating of low, moderate, or high. A unique feature of the SAVRY is its inclusion of protective factors that might mitigate the youth's risk.

Although there is less published research on the SAVRY relative to youth adaptations of the LSI and PCL, the SAVRY has been found to have strong predictive validity for general

and violent recidivism, with moderate to high area under the curve (AUC) values commonly observed for total scores on the tool (e.g., Catchpole & Gretton, 2003; Hilterman, 2007; Meyers & Schmidt, 2008); however, less research has directly examined the SPJ component of the tool (i.e., as it is intended to be used), although psychometric research has also supported the predictive accuracy of the summary risk ratings (e.g., Hilterman, 2007).

Overall, the body of research on standardized forensic assessment measures with youth has been growing steadily, and in the past few years there has been a marked increase in published research and unpublished work (generally theses and dissertations) examining the psychometric properties of these measures. The use of these instruments, however, has created a fair amount of controversy. For instance, some (e.g., Hannah-Moffat & Maurutto, 2003) have voiced concerns that such measures, in particular youth risk and need measures, have yet to demonstrate adequate validity, reliability, and clinical utility with young offenders. Controversy also remains about whether the psychometric properties of risk and need instruments are generalizable to female and Aboriginal youth and whether the use of these instruments is warranted with special populations. Moreover, as alluded to briefly, there have been concerns expressed about the appropriateness of assessing psychopathic features in juveniles. It is argued that some of these attributes are normative (e.g., egocentricity) to some degree and that adolescence is a period of development and transition, including the formation of one's personality (Edens et al., 2001).

SCOPE OF THE PRESENT STUDY

The purpose of the present study is to conduct a meta-analysis of three youth risk assessment measures, specifically to evaluate their predictive validity for various forms of recidivism (general, nonviolent, violent, sexual). We build on prior young offender meta-analyses, which might have focused on a particular instrument or construct such as juvenile psychopathy and the PCL-YV (see Edens et al., 2007) or examined youth and adult offenders together as a whole (e.g., Gendreau et al., 2002). Some prior meta-analyses also have important shortcomings that we attempt to redress. For example, the review of studies for some instruments has been rather narrow (e.g., Schwalbe, 2007, 2008), having neglected to include several unpublished theses and dissertations. There has also been a substantial increase in published and unpublished research since the most recent reviews, which merits inclusion. As such, we attempted to conduct a comprehensive and up-to-date review of the extant young offender risk literature through examining three instruments that have generated the most research, published and unpublished: youth adaptations of the LSI (e.g., YLS/CMI, LSI-SK) and PCL (e.g., PCL-YV) and the SAVRY. We intend to add to the youth meta-analytic risk literature through examining the predictive validity of these tools in specific young offender groups such as female and Aboriginal youth, conducting within-study comparisons in which two or more of the tools are examined, and assessing the generalizability of the predictive accuracy of these tools in Canadian and non-Canadian international samples. Finally, this review excludes specialized instruments designed to assess risk for sexual offending (e.g., Estimate of Risk of Adolescent Sexual Offense Recidivism [ERASOR], Juvenile Sex Offender Assessment Protocol–II [J-SOAP-II], etc.) given that this literature is smaller and the focus of the present study is on general and violent recidivism.

METHOD

SAMPLE IDENTIFICATION

A primary goal is to identify young offender risk assessment studies that examined the predictive accuracy of any of the instruments under investigation (i.e., youth adaptations of the LSI, PCL-YV, and SAVRY). Relevant studies were identified through several avenues. First, well-known academic criminal justice journals were searched for relevant references. Second, the authors searched an electronic database, PsycINFO, for published and unpublished studies using the following search terms: *youth*, *risk assessment*, *recidivism*, *YLS/CMI*, *PCL-YV*, and *SAVRY*. Published studies were retrieved by consulting the original journal source (frequently available online). Unpublished dissertations and theses were accessed through ProQuest, an online database that provides access to full-text master's and doctoral theses. The ProQuest database was also searched using search terms that were limited to the main instrument under study: YLS/CMI, PCL-YV, SAVRY. Three prior meta-analyses (Edens et al., 2007; Gendreau et al., 2002; Schwalbe, 2007) were also consulted to identify potential references that might have been overlooked in the search process. Finally, any additional studies brought to the authors' attention were located, and in some cases (e.g., as with unpublished presentations or foreign language documents) the first author was contacted and the original source or relevant meta-analytic information was obtained.

Studies were examined for their methodological quality and were required to meet the following three criteria for potential inclusion in the current meta-analysis. First, the studies must have included one of the three groups of instruments under study. As there are several versions of the LSI, studies were considered that included the YLS/CMI, YLS/CMI:SV (Hoge & Andrews, 2001), LSI-SK (Andrews et al., 2001), or YO-LSI or, in the case of one study (Nowicka-Sroga, 2004), the LSI–Ontario Revision (Andrews, Bonta, & Wormith, 1995), in which the scoring was adapted to a sample of older adolescent offenders. Similarly, although most of the studies reviewed examined the PCL-YV as developed by Forth et al. (2003), two studies (Brandt, 1993; Forth, Hart, & Hare, 1990) examined modified 18-item versions of the PCL-R in young offender samples. As such, the term *youth adapted* was applied to the LSI and PCL to reflect the adolescent variants of these measures. Finally, the SAVRY appears to have been administered in one form only and is referred to as such.

Our second criterion was that the studies included some measure of recidivism outcome (e.g., arrests, charges, convictions, etc.) in the community after a period of follow-up. We excluded studies that focused solely on institutional offending, as we were primarily interested in examining community recidivism (as opposed to institutional misconduct) and relatively few studies, other than those incorporating the PCL-YV, investigated this outcome. Similarly, studies that only examined postdictive validity (i.e., associations with criminal history) were excluded. If it was uncertain from reading the document whether the study was predictive or actually postdictive, then the study was also excluded. Furthermore, studies examining youth who were not involved in the justice system were excluded. The samples of all studies under consideration included young offenders who had been charged with or convicted of criminal offenses.

The third requirement was that the studies provided sufficient information to code a predictive validity effect size in terms of a common metric (Pearson's r or a point-biserial

r) or facilitate the computation of *r*. For some published work, the original thesis or dissertation was consulted to provide more detailed information either to obtain *r* or to compute it. In some cases, although the study examined the predictive validity of one of the measures under consideration, insufficient information was available to compute *r* or to convert the static reported from the original study into *r*. In such cases, we contacted the original study authors to obtain more specific predictive validity information.[2]

PROCEDURE

A coding protocol was completed for each study included in the analysis, including author and source, sample composition (gender, ethnicity, and age), follow-up time, recidivism base rates, and recidivism predictive accuracy statistics for the three assessment tools. Although some studies provided predictive accuracy information for subscales or components of the different measures (e.g., Factor 1 or 2 score, Leisure/Recreation or Attitudes/Orientation need), this was typically not the case. For this reason, predictive accuracy information was coded only for the total dimensional score on a given instrument. In studies that used more than one outcome measure (e.g., charges and convictions), predictive accuracy information was coded for both outcome criteria.

Criterion measures were coded for four possible types of recidivism: general (i.e., any), nonviolent, violent, and sexual recidivism. Although the operational definition of violent recidivism usually included sex offenses, in studies reporting sex offense recidivism this usually referred to nonsexual violent recidivism. Two additional studies (Schmidt et al., 2005; Schmidt, McKinnon, Chattha, & Brownlee, 2006) also reported predictive accuracy for "serious recidivism" (i.e., violent offenses as well as other serious indictable offenses), and this outcome was coded as violence in the current study.

All studies were originally coded by the primary author and independently recoded by the second author. An overall rate of agreement of 95% was achieved for variables coded (e.g., study and sample characteristics, recidivism rates, predictive accuracy statistics). Discrepancies were resolved through consensus.

EFFECT SIZE CODING

Predictive accuracy statistics were coded in terms of *r*, which in most cases was a point-biserial correlation or r_{pb} (i.e., a correlation between a continuous predictor, such as the score on a risk measure, and a binary criterion variable, such as dichotomous recidivism coded yes–no). When *r* was not reported, the appropriate formula was applied to convert the reported statistic or descriptive information (e.g., mean group differences between recidivists and nonrecidivists) into *r* using formulae provided in Lipsey and Wilson (2001). When only an AUC statistic from Receiver Operating Characteristic (ROC) analyses was provided, the tables provided in Rice and Harris (2005) were used to convert an ROC value into the equivalent r_{pb}.[3] If correlations with continuous recidivism measures (e.g., total convictions) were reported, we chose to compute r_{pb} if sufficient information was available to do so (this was done with Jack, 2000); otherwise the correlation with the continuous outcome measure was retained (done for five studies). In some cases, multiple dependent measures had been coded on a single sample (e.g., separate correlations computed for binary charges and binary convictions). In such cases when more than one effect size

represented a particular outcome measure within a study, a single effect size was created by averaging the two (see Lipsey & Wilson, 2001).

DATA ANALYTIC STRATEGY

All coded data were entered into SPSS. Effect sizes for each study were transformed into z_r using Fisher's transformation, .5 * ln [(1 + r) / (1 – r)], and multiplied by the inverse variance weight, $w_{zr} = n - 3$. The weighted effect sizes were then summed and divided by the sum of the inverse variance weights to create a mean weighted effect size, Mean ES = Σ wes / Σw. Averaged weighted effect sizes were then transformed back through the inverse of the Z_r transformation (Hedges & Olkin, 1985) to compute a mean weighted correlation, where $r = [e^{2ESzr} - 1] / [e^{2ESzr} + 1]$. This procedure was used to compute mean weighted correlations for all reported effect sizes. In turn, 95% confidence intervals (CIs) were computed for reported effect sizes through the formula, ES +/– √[1 / n – 3] * 1.96.

The studies demonstrated considerable variability in the magnitude of their effects, ranging from very small to quite large. Homogeneity analyses were conducted to examine whether the effect sizes obtained were dispersed around their mean no greater than would be expected from sampling error alone (Lipsey & Wilson, 2001) through computing the Q statistic, in which $Q = (\Sigma w_i ES_i^2) - (\Sigma w_i ES_i)^2 / \Sigma w_i$. The Q statistic is distributed as a chi-square and significance is evaluated on $k - 1$ degrees of freedom. A significant Q indicates that there is significant variability in effect sizes among studies (Lipsey & Wilson, 2001).

RESULTS

STUDY AND SAMPLE CHARACTERISTICS

The search process generated 49 studies potentially suitable for analysis. Of these, 22 examined youth adaptations of the LSI, 28 examined the PCL-YV (or the youth-modified adult version), and 9 examined the SAVRY. Of these studies, 10 examined more than one of the three instruments under review. Effect sizes were coded from 27 published studies, 18 unpublished theses or dissertations, 3 conference presentations, and 1 unpublished government document.[4] In all, 42 studies had nonoverlapping samples. For 3 studies, the predictive validity of three different instruments (YLS/CMI, PCL-YV, SAVRY) was examined within the same sample, but the results for each instrument were published in different articles (Meyers & Schmidt, 2008; Schmidt et al., 2005; Schmidt et al., 2006), and hence the effect sizes were obtained from nonoverlapping samples. Two additional studies, reported in Table 2, used the same samples and data from these three aforementioned studies: a thesis by McKinnon (2004) that examined the predictive validity of the YLS/CMI and PCL-YV in Aboriginal and non-Aboriginal youth and an article by Welsh, Schmidt, McKinnon, Chattha, and Meyers (2008) that compared all three instruments within the same sample. As such, the McKinnon thesis and Welsh et al. article are not included in the primary aggregate analyses; however, they provide unique information involving predictive validity comparisons between the instruments and as a function of ethnicity, and as such they are included in supplemental analyses in lieu of the Schmidt et al. (2005; Schmidt et al., 2006) and Meyers and Schmidt (2008) article (which did not report such analyses).

TABLE 2: Meta-Analysis of Youth Risk Measures: Summary of Studies

					Age	Follow-Up	Recidivism, BR		Youth-Adapted LSI		Youth-Adapted PCL		SAVRY	
Authors	*Origin*	N	*Gender*	*Ethnicity*	*(Years)*	*(Months)*	*Type*	*BR*	r	*ROC*	r	*ROC*	r	*ROC*
Auslander (1998)[U]	United States	124	100[M]	52[AA], 44[C], 4[O]	15	34.3	nv, v, s	34/62[a], 17/41[a], 3/8[a]	—	—	.10[b], .23[b], –.02[b]	**.56, .63, <.50**	—	—
Brandt (1993)[U]	United States	129	100[M]	69[AA], 28[C], 3[O]	—	18 to 24	g, nv, v	.88, —, —	—	—	.19, .15, .16	—	—	—
Campbell (2004)[U]	Canada	212	83[M], 17[F]	83[C], 17[O]	16.2	—	g, nv, v	71, 42, 12	—	—	.06, .13, –.27	.58, **.58,** .41	—	—
Catchpole and Gretton (2003)[P]	Canada	74	85[M], 15[F]	55[C], 30[Ab], 14[O]	15 to 19	12	g, v	58, 23	**.41, .40**	.74, .73	**.48, .40**	.78, .73	**.41, .40**	.74, .73
Corrado, Vincent, Hart, and Cohen (2004)[P]	Canada	161	100[M]	71[C], 18[Ab], 11[O]	17.7	14.5	g, nv, v	76, 60, 31	—	—	**.32, .23, .27**	.68, .63, .65	—	—
Cruise (2000)[U]	United States	105	100[M]	66[C], 21[O], 13[AA]	15.3	—	g	26	—	—	.36	**.71**	—	—
Dodds (1999)[U]	United States	117	100[M]	78[AA], 12[C], 10[O]	16.4	—	g	78	—	—	.22	**.63**	—	—
Dolan and Rennie (2008)[P]	United Kingdom	99	100[M]	84[C], 16[O]	16.2	12	g, v	71, 38	—	—	.15, .09	.60, .54	.28, .23	.69, .64
Douglas, Epstein, and Poythress (2008)[P]	United States	83	100[M]	79[C], 14[AA], 5[O]	15.8	28.8	g, nv, v	45, 39, 23	—	—	**.09, .00, .27**	.55, .50, .66	—	—
Edens and Cahill (2007)[P]	United States	75	100[M]	43[AA], 30[C], 26[O]	15.6	120	g, v	84, 32	—	—	–.01, –.07	.51, .46	—	—
Flores, Travis, and Latessa (2004)[GD]	United States	1,313	79[M], 21[F]	70[C], 30[O]	15.4	24	g	39	.30	**.67**	—	—	—	—
Forth, Hart, and Hare (1990)[P]	Canada	71	100[M]	77[C], 23[Ab]	16.3	—	g, nv, v	79, —, —	—	—	.14, .00, .26	**.58,** —, —	—	—
Gossner and Wormith (2007)[P]	Canada	94	78[M], 22[F]	66[C], 34[Ab]	15.9	6	g	30/52[a]	.40[b]	.73	—	—	—	—
Gretton and Abramowitz (2002)[C]	Canada	176	94[M], 6[F]	75[C], 25[O]	—	12	g, v	—, —	—	—	—	—	**.32,** .25	.68, .67
Gretton, Hare, and Catchpole (2004)[P]	Canada	157	100[M]	79[C] 19, 2[O]	16.1	120	nv, v, s	95, 68, 11	—	—	.19, .32, .11	**.61, .68, .56**	—	—

(continued)

TABLE 2: (continued)

Authors	Origin	N	Gender	Ethnicity	Age (Years)	Follow-Up (Months)	Recidivism, BR		Youth-Adapted LSI		Youth-Adapted PCL		SAVRY	
							Type	BR	r	ROC	r	ROC	r	ROC
Gretton, McBride, Hare, O'Shaughnessy, and Kumka (2001)[P]	Canada	220	100[M]	65[C], 22[Ab], 13[O]	14.7	55	g, v, s	51, 30, 15	—	—	.25, .19, .09	**.64, .61, .55**	—	—
Hilterman (2007)[c]	Spain	85	85[M], 15[F]	—	17.9	13.7	g, v	23, 13	—	—	—	—	**.38, .48**	.72, .78
Ilacqua, Coulson, Lombardo, and Nutbrown (1999)[P]	Canada	164	50[M], 50[F]	—	16 to 17	12	g	73	.35[c]	**.70**	—	—	—	—
Jack (2000)[U]	Canada	146	100[M]	69[C], 17[O], 14[Ab]	15.1	—	nv, v	60, 15	.42, .33	**.74, .69**	.24, .20	**.64, .61**	—	—
Jung and Rawana (1999)[P]	Canada	250	66[M], 34[F]	50[C], 50[Ab]	14.3	6	g	30	.37	**.71**	—	—	—	—
Långström and Grann (2000)[P]	Sweden	46	96[M], 4[F]	—	—	60	g, s	65, 20	—	—	.33, .07	**.69, .54**	—	—
Långström and Grann (2002)[P]	Sweden	81	100[M]	—	18.4	24	v	37	—	—	**.27**	.65	—	—
Livsey (2005)[U]	United States	300	64[M], 36[F]	41[AA], 39[C], 20[O]	14.0	6	g	14	.16	**.59**	—	—	—	—
Lodewijks, Doreleijers, and de Ruiter (2008)	Nether-lands	117	95[M], 5[F]	41[Cb], 28[C], 23[Md] 8[O]	15.3	36	v	20	—	—	—	—	**.27**	.65
Luong (2007)[U]	Canada	193	74[M], 26[F]	64[Ab], 36[C]	15.8	21.1	g	62	.40	.73	—	—	—	—
Marczyk (2002)[U]	United States	72	100[M]	71[AA], 17[O], 13[C]	16.4	—	g	69	.07	**.54**	.04	**.52**	—	—
McEachran (2001)[U]	Canada	108	100[M]	85[C], 11[Ab], 4[O]	15.3	12	nv, v	33, 33	—	—	.41, .46	**.74,** .79	.44, .32	**.76,** .70
McKinnon (2004)[U, EC]	Canada	102	63[M], 37[F]	73[C], 27[Ab]	14.6	—	g, v	47, 27	**.18, .22**	.60, .63	.40, .35	.74, .73	—	—
Meyers and Schmidt (2008)[P]	Canada	121	66[M], 34[F]	69[C], 31[Ab]	14.9	36.9	g, nv, v	49, 21, 28	—	—	—	—	**.45, .32, .46**	.76, .68, .77
Morton (2003)[U]	Canada	77	100[M]	—	15.2	68.1	g, v, s	52, 26, 17	.16, .31, .13	.58, .69, .60	—	—	—	—

(continued)

TABLE 2: (continued)

Authors	Origin	N	Gender	Ethnicity	Age (Years)	Follow-Up (Months)	Recidivism, BR		Youth-Adapted LSI		Youth-Adapted PCL		SAVRY	
							Type	BR	r	ROC	r	ROC	r	ROC
Nowicka-Sroga (2004)[U]	Canada	154	100[M]	100[C]	16 to 19	45.6	g, nv, v	51, 47, 21	.46, .43, .31	.77, .75, .72	—	—	—	—
Odgers, Reppucci, and Moretti (2006)[P]	Canada	114	100[F]	40[C], 53[AA], 3[O], 4[Ab]	16.2		g, nv, v	50, 28, 22	—	—	.04, –.03, .05	**.52,** <**.50, .53**	—	—
O'Neill, Lidz, and Heilbrun (2003)[P]	United States	64	100[M]	52[AA], 28[O], 20[C]	16.0	12	g	—	—	—	.33	—	—	—
Onifade et al. (2008)[P]	United States	328	73[M], 27[F]	43[C], 31[AA], 26[O]	14.7	12	g	26	.19	.62	—	—	—	—
Parsons (2005)[U]	Canada	95	100[M]	—	16.1	36	g, nv, v	63, —, 31	—	—	.25, .25, .11	**.64, .64, .56**	—	—
Rector, Wormith, and Banka (2007)[C]	Canada	872	76[M], 24[F]	72[Ab], 28[C]	—	16.5	g	49	.38	**.72**	—	—	—	—
Ridenour, Marchant, and Dean (2001)[P]	United States	80	100[M]	81[C], 18[AA], 1[O]	14 to 18	12	g, nv, v	—	—	—	.62, .59, .48	—	—	—
Rieger, Stadtland, Freisleder, and Nedophil (2008)[P]	Germany	83	88[M], 12[F]	100[C]	16.5	70.8	g, v	71, 38	—	—	**.43, .37**	.75, .71	**.33, .24**	.69, .64
Rowe (2002)[U]	Canada	408	80[M], 20[F]	—	15.3	24.8	g, v	63, 23	.40, .25	**.73, .64**	.41, .27	**.74, .66**	—	—
Salekin (2008)[P]	United States	130	71[M], 29[F]	62[O], 39[AA], 7[C]	14.9	36 to 45	g, v	64, 41	—	—	.27, .24	**.66, .64**	—	—
Schmidt, Hoge, and Gomes (2005)[P]	Canada	104	63[M], 37[F]	71[C], 29[Ab]	14.6	35.8	g, v	46, 29	.19, .26	.61, .67	—	—	—	—
Schmidt, McKinnon, Chattha, and Brownlee (2006)[P]	Canada	127	63[M], 37[F]	69[C], 31[Ab]	14.9	35.5	g, nv, v	46, 20, 26	—	—	.36, .10, .32	.71, .56, .72	—	—
Skowron (2004)[U]	Canada	220	100[M]	—	14.2	48	g, nv, v, s	71, 17, 36, 18	.38, .08, .27, .25	.72, .54, .66, .66	—	—	—	—
Stockdale (2008)[U]	Canada	62	35[M], 65[F]	80[Ab], 20[C]	15.7	25.8	g, v	61, 32	.51, .45	.79, .78	.44, .47	.76, .76	—	—
Thompson and Pope (2005)[P]	Australia	174	100[M]	—	16.6	17	g	40	.28	.67	—	—	—	—

(continued)

TABLE 2: (continued)

Authors	Origin	N	Gender	Ethnicity	Age (Years)	Follow-Up (Months)	Recidivism, BR		Youth-Adapted LSI		Youth-Adapted PCL		SAVRY	
							Type	BR	r	ROC	r	ROC	r	ROC
Upperton and Thompson (2007)[P]	Australia	113	88[M], 12[F]	—	16.2	16.6	g	51	.43	.75	—	—	—	—
van de Ven (2004)[U]	Canada	750	75[M], 25[F]	64[C], 13[AA], 13[O], 8[Unk], 2[Ab]	12 to 19	—	g, v	12/28[a], 6/14[a]	.27[b], .20[b]	—, —	—	—	—	—
Viljoen et al. (2008)[p]	United States	169	100[M]	83[C], 8[AA], 8[O]	13.4	79	g, v, s	43, 13, 8	—	—	—	—	**.15, .15, .06**	.58, .58, .53
Welsh, Schmidt, McKinnon, Chattha, and Meyers (2008)[P,W/in]	Canada	105	64[M], 36[F]	71[C], 29[Ab]	14.6	35.8	g, nv, v	46, 20, 26	.18, .00, .21	.60, .50, .64	.40, .14, .33	.74, .60, .73	.46, .08, .45	.77, .55, .81

Note. LSI = Level of Service Inventory; PCL = Psychopathy Checklist; SAVRY = Structured Assessment of Violence Risk in Youth; BR = base rate (all values are percentages rounded to even numbers); ROC = Receiver Operating Characteristic; P = published study; C = conference presentation; U = unpublished thesis or dissertation; GD = unpublished government document; W/in = included in within-group comparisons only; EC = included in ethnicity comparisons only; *N* = sample size of youth included in predictive validity analyses in a given study; M = male; F = female (all values are percentages rounded to even numbers); C = Caucasian; AA = African American; Ab = Aboriginal; Cb = Caribbean; Md = Mediterranean; O = Other; Unk = Unknown (all values are percentages rounded to even numbers); g = general recidivism; nv = nonviolent recidivism; v = violent recidivism; s = sexual recidivism.

a. Separate base rates are presented for convictions and charges, respectively. Predictive validity statistics (youth-adapted LSI, youth-adapted PCL, and SAVRY columns): *r* and ROC values in bold are estimated values obtained from Rice and Harris (2005) tables.

b. Separate correlations were computed each for charges and convictions, and the value presented is the averaged *r* for the two outcomes.

c. Separate correlations for male and female youth were reported only, and the value is the averaged *r* for the two groups (each of which had the same *n* and recidivism base rate).

Finally, two studies by Långström and Grann investigated the predictive validity of the PCL-YV in the same sample, although one article examines predictive accuracy for violence (Långström & Grann, 2002) whereas the other examines predictive accuracy for sexual and general recidivism (Långström & Grann, 2000). As such, both studies were used for separate analyses.

The 44 independent samples included a total of 8,746 young offenders, of which 82.6% were male and 17.4% female, with a mean age of 15.7 years (SD = 1.36).[5] Across the 35 studies reporting ethnicity information (N = 7,264), 56.9% of youth were Caucasian, 16.4% Aboriginal, 11.8% African American, and 14.9% other ethnic descent. Youth were followed up an average of 29.1 months (SD = 24.8) postrelease (k = 35). Base rates of recidivism were 50.2% (k = 34) for general recidivism, 46.4% (k = 11) for nonviolent recidivism, 28.4% (k = 24) for violent recidivism, and 13.3% (k = 7) for sexual recidivism.

META-ANALYTIC RESULTS

Table 2 provides a summary of the 49 studies for potential inclusion in the meta-analysis. The studies are arranged alphabetically by first author and include the origin (i.e., country in which the study was conducted), sample size (n) of youth for whom recidivism information was obtained (this value being the weight used in the effect size calculation), proportion of sample by gender and ethnicity, mean age, length of follow-up (months), recidivism base rates for each criterion measure, and the predictive accuracy statistic for each of the three assessment measures for each criterion reported. Both the correlation coefficient (usually r_{pb}) and ROC statistic (AUC) are included in the table (studies that used only continuous recidivism measures have no ROC computed). Aggregate analyses incorporated the correlation coefficient as the only metric.

Outliers were identified by obtaining the mean unweighted correlation and its standard deviation for each of the effect sizes. Two studies that examined the predictive accuracy of the PCL-YV had outliers as demonstrated by the fact that their reported effect sizes were more than two standard deviations above or below the mean correlations for a particular effect size. These outliers included Campbell's (2004) negative correlation for violence (r = –.27) and Ridenour, Marchant, and Dean's (2001) reported predictive validity correlations for general (r = .62) and nonviolent (r = .59) recidivism. These two studies were excluded from any aggregate analyses. All other effect sizes were within two standard deviations from the mean, resulting in 44 studies included in the meta-analysis, of which 42 studies with nonoverlapping effect sizes were included in the primary aggregate analyses. The largest remaining number of studies was obtained for the youth adaptations of the LSI (k = 22) and PCL (k = 27), with a smaller collection of studies for the SAVRY (k = 9). Separate analyses were initially computed between the YLS/CMI proper and the other youth adaptations of the LSI (e.g., LSI-SK); however, given their substantial overlap in item content and the minimal differences observed in the predictive accuracy between the two sets of instruments, we chose to aggregate the findings across all youth variants of the LSI. In addition, SAVRY effect sizes were derived from the total numeric score on the tool rather than the summary risk rating.

The results for the first aggregate analyses across the three instruments are presented in Table 3. Several themes are worth noting. First, all three measures significantly predicted general, nonviolent, and violent recidivism (i.e., given that none of the 95% CIs fell below

TABLE 3: Predictive Accuracy of Selected Youth Risk Measures

Measure	k	n	*Unweighted Mean* r	r_w	*95% Confidence Interval (Lower, Upper)*	*Q* r_w
Youth-adapted Level of Service Inventory						
General	19	5,722	.32	.32	.29, .34	48.95**
Nonviolent	3	520	.31	.29	.20, .37	17.47**
Violent	9	1,995	.29	.26	.21, .30	9.05
Sexual	2	187	.19	.20	.06, .35	0.68
Youth-adapted Psychopathy Checklist						
General	20	2,335	.25	.28	.24, .32	45.17**
Nonviolent	11	1,316	.15	.16	.11, .22	19.60*
Violent	20	2,547	.25	.25	.21, .29	34.39**
Sexual	4	547	.07	.07	–.01, .16	1.11
Structured Assessment of Violence Risk in Youth						
General	7	807	.33	.32	.28, .35	9.44
Nonviolent	2	229	.38	.38	.24, .51	1.19
Violent	9	1,032	.31	.30	.24, .36	14.41
Sexual	1	169	.06	.06	–.09, .21	—

*Q statistic significant at $p < .05$. **Q statistic significant at $p < .01$.

zero) and demonstrated comparable degrees of predictive accuracy as demonstrated by their overlapping CIs within a given recidivism category. This being said, youth versions of the LSI and PCL seemed to predict general recidivism (mean $r_w = .32$ and .28, respectively) somewhat better than violence (mean $r_w = .26$ and .25, respectively), whereas the SAVRY yielded more comparable predictive accuracy for both outcomes (mean $r_w = .32$ and .30, respectively). Although the youth PCL had several studies that examined nonviolent recidivism, there were very few for youth adaptations of the LSI or the SAVRY. There were also comparatively few studies examining the predictive accuracy of the three measures for sexual recidivism, which tended to be lowest for this outcome, although the YLS/CMI significantly predicted sexual recidivism across two studies (mean $r_w = .20$).[6]

It is important to note that there was a substantial amount of variability in the magnitude of several of the effect sizes for both youth versions of the LSI and PCL. The youth LSI had a very large Q statistic (48.95, $p < .01$) for the prediction of general recidivism across 19 studies, although there was reasonable homogeneity among the studies in the prediction of violence across 9 studies (Q = 9.05, *ns*). Significant heterogeneity was also observed in the effect sizes of the PCL in the prediction of both violent (Q = 34.39, $p < .01$) and general (Q = 45.17, $p < .01$) recidivism. The SAVRY, however, did not demonstrate significant variability in its effect sizes across its smaller number of studies. We revisit this issue of effect size variability later in the article.

WITHIN-STUDY COMPARISONS

The issue of within-study variability is a significant concern when comparing the magnitude of effect sizes across different instruments. Put simply, studies using different samples differ in important ways, and arguably the most valid paradigm for examining and interpreting possible differences in the predictive efficacy of different measures would be

TABLE 4: Within-Study Comparisons of Youth Risk Measures for General and Violent Recidivism

Measure	r_w	*95% Confidence Interval (Lower, Upper)*	$Q\ r_w$	*Measure (Lower, Upper)*	r_w	*95% Confidence Interval (Lower, Upper)*	$Q\ r_w$
General recidivism				Violent recidivism			
$n = 721$, $k = 5$				$n = 795$, $k = 5$			
YLS/CMI	.35	.27, .43	13.49*	YLS/CMI	.29	.22, .36	5.36
PCL-YV	.39	.31, .46	10.82*	PCL-YV	.29	.22, .36	4.85
$n = 361$, $k = 4$				$n = 469$, $k = 5$			
PCL-YV	.36	.26, .47	7.12*	PCL-YV	.33	.24, .42	9.21*
SAVRY	.37	.26, .49	2.52	SAVRY	.33	.24, .42	4.37
$n = 179$, $k = 2$				$n = 179$, $k = 2$			
YLS/CMI	.28	.13, .43	2.74	YLS/CMI	.29	.14, .44	1.77
SAVRY	.44	.29, .59	0.15	SAVRY	.43	.28, .58	0.18

Note. YLS/CMI = Youth Level of Service/Case Management Inventory; PCL-YV = Psychopathy Checklist–Youth Version; SAVRY = Structured Assessment of Violence Risk in Youth.
*Q statistic significant at $p < .05$. **Q statistic significant at $p < .01$.

to examine their performance within the same sample of youth. As such, studies that involved administering two or more of the measures within the same sample were selected, and within-study comparisons were conducted among the three instruments. Comparisons are limited to the prediction of general recidivism and violence, given that so few studies examined the prediction of other outcomes. The results are reported in Table 4.[7]

Five studies examined the predictive accuracy of the PCL-YV and the YLS/CMI within the same sample (four studies included both general and violent recidivism, whereas two studies examined only one of these outcomes). As with the main aggregate analyses, no clear winner emerged. Both the YLS/CMI and the PCL-YV significantly predicted general recidivism, and there was significant variability in their effect sizes across studies. The two measures significantly predicted violent recidivism, with the effect size somewhat smaller in magnitude but demonstrating no significant variability across studies.

A similar trend emerged in comparisons of the PCL-YV and SAVRY, with each measure significantly predicting general and violent recidivism with a comparable degree of accuracy. As in prior analyses, the PCL-YV demonstrated significant variability in its effect sizes whereas the SAVRY did not.

Finally, YLS/CMI and SAVRY comparisons were conducted on two studies that administered both instruments. Although there appear to be substantive differences between the two measures in predictive accuracy for both outcomes, it is difficult to generalize from these findings given the small number of studies available for analysis. Although one study found identical predictive accuracy between the two measures for both outcomes (Catchpole & Gretton, 2003), in another the SAVRY demonstrated stronger prediction (Welsh et al., 2008).

GENDER AND ETHNICITY COMPARISONS

A further important issue is the generalizability of prediction across samples that are diverse with respect to gender, culture, and ethnicity, that is, whether such risk measures can predict recidivism in male and female youth as well as those who are of varying cultural and ethnic backgrounds (e.g., Aboriginal youth). A collection of studies reported separate predictive validity information specifically for such groups of interest. The few

TABLE 5: Gender and Ethnicity Within-Study Comparisons on Youth Adaptations of the Level of Service Inventory for General and Violent Recidivism

Measure	k	n	r_w	*95% Confidence Interval (Lower, Upper)*	*Q* r_w
General recidivism					
Males	9	2,968	.33	.29, .36	20.32**
Females	9	992	.36	.29, .42	13.05*
Aboriginal	5	860	.35	.28, .41	2.22
Non-Aboriginal	5	462	.32	.23, .41	7.10
Violent recidivism					
Males	4	974	.23	.17, .30	4.67
Females	4	350	.24	.13, .34	2.95

*Q statistic significant at $p < .05$. **Q statistic significant at $p < .01$.

studies that examined such differences were primarily limited to youth variants of the LSI (e.g., YLS/CMI, LSI-SK) within a Canadian sample. As such, sufficient data for meta-analytic comparisons were available only for male–female and Aboriginal–non-Aboriginal within-group comparisons on the LSI.[8] Results are reported in Table 5.

The youth variants of the LSI continued to significantly predict general recidivism and to do so at comparable degrees of magnitude for male, female, Aboriginal, and non-Aboriginal youth. The magnitude of prediction was comparable to the findings reported across the range of studies in the primary aggregate analyses. In the four studies examining the prediction of violence as a function of gender, youth adaptations of the LSI also predicted comparably (albeit to a smaller degree as in the main analyses) for both genders. As only two studies (McKinnon, 2004; Stockdale, 2008) made Aboriginal–non-Aboriginal comparisons based on the YLS/CMI in the prediction of violence, further analyses were not conducted.

INTERNATIONAL COMPARISONS

As some of these measures (youth LSI, PCL-YV) have been developed and normed primarily on Canadian samples, international comparisons were conducted to examine the predictive accuracy for Canadian and non-Canadian studies (e.g., United States, United Kingdom, Sweden, Australia, Germany, Spain). The results are presented in Table 6.

The statistical significance of the difference between the weighted effect sizes for a particular outcome was examined via Fisher's z test for independent r values, in which $z = (z_{r1} - z_{r2})/\sqrt{([1 / \Sigma N_1 - 3] + [1 / \Sigma N_2 - 3])}$. Fairly marked differences emerged in the predictive accuracy of youth versions of the LSI and PCL among Canadian and non-Canadian studies. The youth version of the PCL had a significantly greater magnitude of prediction for general recidivism in Canadian studies ($r = .34$, $k = 8$) than in countries conducted outside of Canada ($r = .20$, $k = 12$), $z = 3.60$, $p < .001$. Significant differences were also observed in the predictive validity of the youth PCL for violence (Canadian, $r = .28$, $k = 11$; non-Canadian, $r = .18$, $k = 9$), $z = 2.53$, $p < .05$. It is interesting that when predictive accuracy was examined within a given international designation (i.e., within vs. outside Canada), the variability in the effect sizes among the studies frequently decreased to nonsignificance (as demonstrated by a nonsignificant Q), suggesting that differences contributed to between-study variability.

TABLE 6: Comparisons on Selected Youth Risk Measures by Region (Canada vs. Other Countries)

Measure	k	n	r_w	*95% Confidence Interval (Lower, Upper)*	*Q* r_w	*Scale* M	*Scale* SD	*Base Rate*	*Mean* n	*Mean Follow-Up*
Youth-adapted LSI General recidivism only										
Canada	13	3,422	.35	.32, .39	21.60*	17.8	8.0	55.1	263.3	26.8
Other countries	6	2,300	.26	.22, .30	13.54*	16.2	7.4	37.0	503.3	14.0
Youth-adapted PCL General recidivism										
Canada	8	1,218	.34	.29, .40	11.69	21.5	8.7	61.3	152.3	29.1
Other countries	12	1,117	.20	.14, .26	20.41*	20.2	6.2	61.4	93.1	48.9
Violent recidivism										
Canada	11	1,629	.28	.23, .33	15.28	21.4	8.1	30.5	148.1	25.1
Other countries	9	918	.18	.12, .25	12.96	22.6	6.8	34.9	102.0	47.0
SAVRY General recidivism										
Canada	3	371	.38	.28, .48	8.74*	24.0	9.1	53.5	123.7	20.3
Other countries	4	436	.26	.17, .36	4.18	21.5	7.9	50.2	109.0	43.9
Violent recidivism										
Canada	4	479	.34	.25, .43	4.41	24.5	8.8	28.1	119.8	18.2
Other countries	5	553	.26	.18, .35	7.74	21.4	7.6	22.4	110.6	42.3

Note. LSI = Level of Service Inventory; PCL = Psychopathy Checklist; SAVRY = Structured Assessment of Violence Risk in Youth. Youth-adapted LSI descriptive statistics conducted for studies including the Youth Level of Service/Case Management Inventory only (Canadian $k = 7$, non-Canadian $k = 4$). For descriptive statistics computed on remaining measures, the *k* values were as follows: Youth-adapted PCL: general recidivism Canadian $k = 8$, non-Canadian $k = 9$; violent recidivism Canadian $k = 9$, non-Canadian $k = 7$; SAVRY: general recidivism Canadian $k = 3$, non-Canadian $k = 3$; violent recidivism Canadian $k = 4$, non-Canadian $k = 4$. All *t* test comparisons for scale mean, scale *SD*, base rate, mean *n*, and mean follow-up between sets of Canadian and non-Canadian studies are nonsignificant.
*Q statistic significant at $p < .05$.

A similar trend emerged for youth variants of the LSI. As only Canadian studies examined the predictive accuracy of the youth LSI for violence, international differences were examined for general recidivism. The youth LSI also demonstrated significantly greater predictive accuracy among Canadian studies ($r = .35$, $k = 13$) than in non-Canadian studies ($r = .26$, $k = 6$), $z = 3.65$, $p < .001$. Although significant variability in the magnitude of effect continued to be observed across the 13 Canadian studies, the magnitude of the *Q* statistic decreased substantially. Finally, the observed differences fell short of significance when Canadian and non-Canadian studies were compared in the predictive validity of the SAVRY for both general ($r = .38$, $k = 3$ and $r = .26$, $k = 4$, respectively; $z = 1.87$, $p = .06$) and violent recidivism ($r = .34$, $k = 4$ and $r = .26$, $k = 5$, respectively), $z = 1.38$, *ns*.

Table 6 also presents the descriptive statistics for the risk measures (mean and standard deviations), recidivism base rates, follow-up time, and sample means for each collection of Canadian and non-Canadian studies with respect to a given recidivism outcome across the three measures.[9] Given the variability in scale totals and standard deviations among the various youth adaptations of the LSI, only descriptive statistics for studies specifically examining the YLS/CMI are presented. All *t* test comparisons between the sets of studies on these dependent variables were nonsignificant.

DISCUSSION

The present study was a meta-analysis of three assessment measures commonly used in risk assessment with young offenders: youth adaptations of the LSI (e.g., YLS/CMI, LSI-SK) and the PCL (e.g., PCL-YV) and the SAVRY. In all, 49 studies were initially identified for possible inclusion, and 44 were retained for the primary analyses. All three measures significantly predicted general, nonviolent, and violent recidivism with comparable degrees of accuracy. With mean weighted correlations ranging from $r = .28$ to .32 for general recidivism, $r = .16$ to .38 for nonviolent recidivism, and $r = .25$ to .30 for violent recidivism, the magnitude of prediction for the three measures was comparable to prediction findings for their adult counterparts (e.g., Gendreau et al., 2002). This trend also extended to within-study comparisons, which sought to examine the relative predictive accuracy of the three measures when they had been administered and examined within the same study and sample. Although at times, the number of studies available for within-study comparisons between two or more given instruments was limited, the substantive findings did not support the superior predictive accuracy of one measure over another for any particular recidivism outcome. The relatively short mean follow-up time (28.4 months) for the young offender studies is also noteworthy, especially when compared to the adult literature in which the mean follow-up time in meta-analytic investigations is about 5 years (e.g., Bonta, Law, & Hanson, 1998; Hanson & Bussière, 1998). Although the majority of studies have tended to focus on juvenile recidivism, which may account for the shorter follow-up times, some studies have examined the long-term predictive validity of these tools (e.g., Gretton, Hare, & Catchpole, 2004). Although the prediction of adult criminal recidivism is important and interesting, some have argued (Douglas & Kropp, 2002), and we concur, that the ultimate purpose of risk assessment should be the *prevention* as opposed to the *prediction* of criminal recidivism. As such, we would argue that some of the most productive inroads in the field of young offender risk assessment might be found in assessing risk and preventing recidivism through treatment, effective case management, and supervision, so as to prevent young offenders from becoming adult offenders. In addition, given that adolescence is a period of substantial development and transition, more frequent and repeated assessments are warranted.

The weakest predictive accuracy was observed with respect to sexual recidivism for all three instruments, although this was not unexpected given that none of these measures was specifically designed to assess risk for sexual violence. Results obtained for the SAVRY and PCL-YV appeared to be somewhat lower compared to specialized tools such as the J-SOAP-II or the ERASOR (e.g., Morton, 2003; Skowron, 2004; Viljoen et al., 2008), although the results from this literature have also been mixed, perhaps in part because of low sexual recidivism base rates. It is interesting that the YLS/CMI fared reasonably in predicting this outcome relative to specialized measures (e.g., Morton, 2003; Skowron, 2004); however, given the small collection of studies, it may be premature to apply too much weight to this finding.

It is interesting that the three instruments performed comparably overall in predicting different forms of recidivism, given that they were designed for different purposes. Perhaps this is none too surprising given their content-related overlap and high degree of convergence, at least among their adult counterparts (e.g., Wormith, Olver, Stevenson, & Girard, 2007). The youth LSI was developed to assess risk for general recidivism and identify

general risk and need areas to be targeted for treatment, the PCL-YV is a diagnostic tool designed to provide a valid and reliable measurement of psychopathic features in juveniles, and the SAVRY was developed as a SPJ tool specifically to assess risk for violence. The decision to incorporate one of these instruments when conducting risk assessments with youth may matter more in terms of other potential uses of a given measure (e.g., treatment planning, level of community supervision) than any benefits to be gained in predictive accuracy. Put simply, a youth scoring high on a youth adaptation of the LSI, the PCL-YV, or the SAVRY is likely to pose substantial recidivism risk, general or violent. However, the different measures have clinical applications beyond mere recidivism prediction that may warrant consideration in their use. For instance, risk for violence may inform planning for violence-specific treatment, whereas a high level of psychopathic features may have important responsivity implications for programming and/or supervision.

Another question with clinical relevance is whether such measures are indeed valid with specific young offender groups, such as female offenders or Aboriginal youth (Hannah-Moffat & Maurutto, 2003). Although the dominant demographic across the studies was Caucasian males, the majority of the samples were ethnically diverse, and many included female offenders. Although female offenders tended to occupy a noticeable minority (about 15% overall, which reflects the proportion of women across many correctional organizations), few studies investigated differences in the predictive accuracy of any of the measures specifically within female youth or ethnic minorities. Nevertheless, where possible we investigated the predictive accuracy of the measures among young offender groups diverse with respect to gender, culture, and ethnicity. At the time of the current study, predictive validity information for specific gender and ethnic groups sufficient for meta-analysis was available only for youth adaptations of the LSI. It is important that the instrument continued to predict recidivism across male, female, Aboriginal, and non-Aboriginal youth at a magnitude comparable to that in the main analyses. As such, the results support the predictive validity of youth adaptations of the LSI with female and Aboriginal youth.

In terms of the validity of the PCL-YV and SAVRY with female offenders and ethnic minorities (e.g., Aboriginal youth), few studies have reported the results of separate male–female or Caucasian–minority analyses. Edens et al. (2007) presented separate male and female results for the PCL-YV in which the results from all of the studies with male youth ($k = 19$) were compared to all studies with female youth ($k = 5$). Since the Edens et al. meta-analysis, we have found only one unique study (Stockdale, 2008) reporting separate results by gender, with the PCL-YV significantly predicting violent ($r = .33$) and general ($r = .34$) recidivism among female youth ($N = 40$). Part of the concern it seems is that heterogeneous young offender samples tend to include only a small number of female youth, making analyses among specific genders potentially untenable (e.g., Salekin, 2008). It is tempting to conclude that the PCL-YV and SAVRY have predictive validity for violent and general recidivism among female youth, given the heterogeneity of samples with respect to gender and also in light of the findings for youth adaptations of the LSI. Although this may not be an unreasonable conclusion to draw, in light of the mixed findings for the PCL-YV with females (e.g., Odgers, 2006; Odgers, Reppucci, & Moretti, 2005), it is an empirical question, and more research on the validity and utility of these tools with female youth and ethnic minorities is warranted.

In addition to gender and ethnicity considerations, a further practical question concerns the generalizability of predictive validity findings across jurisdictions. We examined international

differences in the prediction of the tools by comparing the magnitude of the predictive validity coefficients between Canadian and non-Canadian studies. Two interesting findings emerged. First, two out of the three measures had significantly stronger predictive accuracy in Canadian studies, although it is important to note that all three measures continued to demonstrate significant predictive accuracy in non-Canadian jurisdictions. Second, the significant variability that existed in the magnitude of the effect sizes diminished substantially when effect sizes were computed by country, suggesting that "international" differences contributed to the variability across studies.

Although the observed differences are tempered by the fact that in some cases a very small number of studies formed the basis of the comparisons, an attempt to explain the discrepancies in predictive accuracy is warranted. There are various possible explanations. First, the youth adaptations of the LSI and the PCL-YV (along with their adult counterparts) have been developed and normed on Canadian samples, and the international differences may be reflective of the scales' origins. It is not uncommon that measures demonstrate some decrease in the magnitude of their predictive accuracy coefficients when cross-validated in other samples. However, given that the vast majority of Canadian studies employed samples that were entirely independent from the instruments' respective development samples, we believe this is unlikely to account for such differences.

It is also possible that there were systematic differences between Canadian and non-Canadian studies, although we found only minimal evidence in this regard. Follow-up analyses did not reveal any significant differences between Canadian and non-Canadian studies in base rates of recidivism, length of follow-up, or mean scores on the three instruments, although there was a tendency for non-Canadian studies to have smaller standard deviations (hence, less variability), which might have served to attenuate the magnitude of predictive validity correlations. Other possible considerations might be international differences in justice systems or perhaps the issue of exporting a scale to a different culture. The international differences in predictive accuracy, whether they are "real" or artifact, remains a conundrum awaiting explanation. Further avenues of investigation might include examining the contributions of legal or correctional system variables (e.g., custody vs. referred youth), training in instrument use and fidelity of administration, heterogeneity within samples, instrument ratings made by raters within a research setting versus "real-world" assessment contexts, and possible differences owing to the accuracy in coding the recidivism dependent variable (e.g., in Canada there is the national Canadian Police Information Centre system, which is used in many, but not all, recidivism studies).

In this regard, there are some limitations in the current study that are worth noting. First, we did not fully investigate the variability that we obtained for many of the effect sizes on most of the instruments. Future, more detailed research should investigate possible sources. Second, and in a similar vein, we were unable to aggregate the results by gender and ethnicity for all of the instruments under investigation. Such analyses are essential for the acceptance of the examined instruments by correctional agencies. Third, the studies incorporated into the current review relied on official sources of criminal recidivism; although official records might have increased reliability, they invariably underestimate the true base rates of recidivism. Finally, we were interested in the performance of the aggregate measures and therefore did not consider the predictive accuracy for individual scale components, such as individual criminogenic needs (e.g., antisocial attitudes) or facets of juvenile psychopathy,

nor did we examine different models of juvenile psychopathy (e.g., three-factor vs. four-factor models). The availability of predictive validity information for scale components was highly variable; however, as the volume of literature continues to grow, future research might provide a more nuanced examination of these assessment measures.

In conclusion, the meta-analytic findings support the predictive efficacy of three forensic youth measures for general and violent recidivism. Although we would hardly expect the current study to quell the controversy that comes with clinical applications of these tools with this clientele, we submit that a conscientious, ethical, appropriate, and standardized administration of these tools can be part of effective clinical service provision. Although there have been understandable apprehensions expressed about the potentially harmful effects of youth risk measures with this vulnerable client group (Edens et al., 2001; Hannah-Moffat & Maurutto, 2003), the appropriate use of such measures can actually minimize the intrusiveness of the justice system (e.g., use of diversionary approaches for lower risk youth), identify targets (criminogenic) for service delivery (e.g., community and treatment resources), and inform sentencing options (e.g., the Intensive Rehabilitative Custody and Supervision provisions in Section 42 of Canada's Youth Criminal Justice Act [2002] as an alternative to an adult sentence for serious offenses) to improve criminal justice outcomes for young offenders.

The current systematic review reveals to us specific gaps in the literature and leads to the following suggestions. Although many of the studies we reviewed were multiethnic and included both males and females, separate predictive accuracy findings were rarely reported for these specific groups. As such, we would strongly encourage researchers to report findings for specific gender and ethnic groups, statistical power permitting. We would also respectfully encourage researchers to report predictive accuracy findings using a standard and easily interpretable metric such as a point-biserial correlation, phi coefficient, or AUC statistic to facilitate the aggregation of findings through meta-analytic techniques.

NOTES

1. The terms *youth* (Canada) and *juvenile* (United States), as defined by national (Canada) or state (United States) legislations, are not to be confused with the term *youthful*, as often used in U.S. jurisdictions (up to age 25).

2. We did not obtain effect size information from secondary sources (i.e., published meta-analyses reporting effect size information) when either the original source was unavailable or the reported data could not be readily converted to *r*. For this reason, two studies from Edens, Campbell, and Weir (2007) were not included (Parks [2004] and Vincent et al. [2005], which was submitted for publication at the time but is now published as Vincent, Odgers, McCormick, and Corrado [2008]), as the necessary information could not be obtained from the original authors.

3. The point-biserial correlation and area under the curve (AUC) are proportional but not completely equivalent (i.e., there is not a 100% correspondence between the two). As such, there is some error involved in estimating an AUC value from r_{pb} and vice versa. This is reflected in the summary of studies in Table 2, in which case the reader will notice that sometimes the estimated AUC for a given correlation will be the same value as the actual AUC for a correlation coefficient of a different value.

4. Although some of the unpublished master's and doctoral theses were later published, at times the published document did not include sufficient information to code or compute a predictive validity effect size. In such cases, the original source is cited instead.

5. The sample characteristics for Schmidt, McKinnon, Chattha, and Brownlee (2006) are included in the broad sample analyses, given that this sample is the largest (N = 127) out of the three articles. See Table 2 for sample descriptives for Schmidt, Hoge, and Gomes (2005) and Meyers and Schmidt (2008).

6. In computing the effect size for sexual recidivism with the Youth Level of Service/Case Management Inventory (YLS/CMI) in the Skowron dissertation, predictive validity statistics for sexual recidivism were computed only for the 110 youth who were sex offenders. Predictive validity statistics for general, nonviolent, and violent recidivism, however, were conducted on the entire sample (N = 220).

7. The Welsh, Schmidt, McKinnon, Chattha, and Meyers (2008) article is used for the within-study comparisons in lieu of the Schmidt et al. (2005; Schmidt et al., 2006) and Meyers and Schmidt (2008) studies, which reported the different instruments using slightly different *N* values from the same sample.

8. The McKinnon (2004) master's thesis is used for Aboriginal–non-Aboriginal comparisons on the YLS/CMI, given that the Schmidt et al. (2005) article did not report these findings. We use the published results from Schmidt et al. for the male–female comparisons.

9. Complete information for these descriptive statistics was not available for all studies. For instance, in some cases, mean scores and/or standard deviations were not consistently reported in the predictive validity studies, as was the case with recidivism base rates (generally with continuous recidivism measures) or follow-up times. The descriptive statistics were computed using all the information available to the study authors.

REFERENCES

*References marked with an asterisk indicate studies included in the meta-analysis.

Andrews, D. A., & Bonta, J. (1995). *Level of Supervision-Revised (LSI-R): An offender assessment system. User's guide*. Toronto: Multi-Health Systems.

Andrews, D. A., & Bonta, J. (2003). *The psychology of criminal conduct* (3rd ed.). Cincinnati, OH: Anderson.

Andrews, D. A., Bonta, J., & Wormith, J. S. (1995). *Level of Service Inventory–Ontario Revision (LSI-OR): Interview and scoring guide*. Toronto: Ontario Ministry of the Solicitor General and Correctional Services.

Andrews, D. A., Bonta, J., & Wormith, J. S. (2001). *Level of Service Inventory–Saskatchewan Youth Edition*. Toronto: Multi-Health Systems.

Andrews, D. A., Bonta, J., & Wormith, J. S. (2004). *Level of Supervision/Case Management Inventory (LS/CMI): An offender assessment system. User's guide*. Toronto: Multi-Health Systems.

Andrews, D. A., Zinger, I., Hoge, R. D., Bonta, J., Gendreau, P., & Cullen, F. T. (1990). Does correctional treatment work? A clinically-relevant and psychologically informed meta-analysis. *Criminology, 28*, 369-404.

*Auslander, B. A. (1998). *An exploratory study investigating variables in relation to juvenile sexual reoffending*. Unpublished doctoral dissertation, Florida State University, Tallahassee.

Bonta, J. (2002). Offender risk assessment: Guidelines for selection and use. *Criminal Justice and Behavior, 29,* 355-379

Bonta, J., Law, M., & Hanson, K. (1998). The prediction of criminal and violent recidivism among mentally disordered offenders: A meta-analysis. *Psychological Bulletin, 123*(2), 123-142.

Borum, R. (2000). Assessing violence risk among youth. *Journal of Clinical Psychology, 56,* 1263-1288.

Borum, R., Bartel, P., & Forth, A. (2006). *Manual for the Structured Assessment of Violence Risk in Youth (SAVRY).* Odessa FL: Psychological Assessment Resources.

*Brandt, J. R. (1993). *Assessment of psychopathy in a population of incarcerated adolescent offenders*. Unpublished doctoral dissertation, Florida State University, Tallahassee.

*Campbell, M. (2004). *The relative contribution of psychopathy and traditional risk factors in predicting young offender recidivism*. Unpublished doctoral dissertation, Dalhousie University, Halifax, Nova Scotia, Canada.

*Catchpole, R. E., & Gretton, H. M. (2003). The predictive validity of risk assessment with violent young offenders. *Criminal Justice and Behavior, 30,* 688-708.

Cleckley, H. (1976). *The mask of sanity* (5th ed.). St. Louis, MO: C. V. Mosby.

*Corrado, R. R., Vincent, G. M., Hart, S. D., & Cohen, I. M. (2004). Predictive validity of the Psychopathy Checklist: Youth version for general and violent recidivism. *Behavioral Sciences and the Law, 22,* 5-22.

*Cruise, K. R. (2000). *Measurement of adolescent psychopathy: Construct and predictive validity in two samples of juvenile offenders*. Unpublished doctoral dissertation, University of North Texas, Denton.

*Dodds, L. D. (1999). *Legislative exclusion statutes and psychopathy: Balancing rehabilitation ideals and retribution goals with juvenile sentencing*. Unpublished doctoral dissertation, Medical College of Pennsylvania and Hahnemann University, Philadelphia.

*Dolan, M. C., & Rennie, C. E. (2008). The Structured Assessment of Violence Risk in Youth as a predictor of recidivism in a united Kingdom cohort of adolescent offenders with conduct disorder. *Psychological Assessment, 20,* 35-46.

*Douglas, K. S., Epstein, M. E., & Poythress, N. G. (2008). Criminal recidivism among juvenile offenders: Testing the incremental and predictive validity of three measures of psychopathic features. *Law and Human Behavior, 32,* 423-438.

Douglas, K. S., & Kropp, P. R. (2002). A prevention-based paradigm for violence risk assessment: Clinical and research applications. *Criminal Justice and Behavior, 29*, 617-658.

Douglas, K. S., Yeomans, M., & Boer, D. P. (2005). Comparative validity analysis of multiple measures of violence risk in a sample of criminal offenders. *Criminal Justice and Behavior, 32,* 479-510.

*Edens, J. F., & Cahill, M. A. (2007). Psychopathy in adolescence and criminal recidivism in young adulthood: Longitudinal results from a multiethnic sample of youthful offenders. *Assessment, 14,* 57-64.

Edens, J. F., Campbell, J. S., & Weir, J. M. (2007). Youth psychopathy and criminal recidivism: A meta-analysis of the Psychopathy Checklist measures. *Law and Human Behavior, 31,* 53-75.

Edens, J. F., Skeem, J. L., Cruise, K. R., & Cauffmann, E. (2001). Assessment of "juvenile psychopathy" and its association with violence: A critical review. *Behavioral Sciences and the Law, 19,* 53-80.

*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.

*Forth, A. E., Hart, S. D., & Hare, R. D. (1990). Assessment of psychopathy in male young offenders. *Psychological Assessment: A Journal of Consulting and Clinical Psychology, 2*, 342-344.

Forth, A. E., Kosson, D. S., & Hare, R. D. (2003). *Psychopathy Checklist: Youth Version (PCL:YV).* Toronto: Multi-Health Systems.

Gendreau, P., Goggin, C., & Smith, P. (2002). Is the PCL-R really the "unparalleled" measure of offender risk? A lesson in knowledge cumulation. *Criminal Justice and Behavior, 29,* 397-426.

*Gossner, D., & Wormith, J. S. (2007). The prediction of recidivism among young offenders in Saskatchewan. *Canadian Journal of Police & Security Services, 5,* 70-82.

*Gretton, H., & Abramowitz, C. (2002, March). *SAVRY: Contribution of items and scales to clinical risk judgments and criminal outcomes.* Paper presented at the biennial conference of the American Psychology and Law Society, Austin, TX.

*Gretton, H. M., Hare, R., & Catchpole, R. E. H. (2004). Psychopathy and offending from adolescence to adulthood: A 10-year follow-up. *Journal of Consulting and Clinical Psychology, 72*, 636-645.

*Gretton, H. M., McBride, M., Hare, R. D., O'Shaughnessy, R., & Kumka, G. (2001). Psychopathy and recidivism in adolescent sex offenders. *Criminal Justice and Behavior, 28,* 427-449.

Grisso, T. (1998). *Forensic evaluations of juveniles.* Sarasota, FL: Personal Resource Press.

Hannah-Moffat, K., & Maurutto, P. (2003). *Youth risk/need assessment: An overview of issues and practices.* Ottawa: Department of Justice Canada, Research and Statistics Division.

Hanson, R. K., & Bussière, M. (1998). Predicting relapse: A meta-analysis of sexual offender recidivism studies. *Journal of Consulting and Clinical Psychology, 66,* 348-362.

Hare, R. D. (1991). *The Psychopathy Checklist Revised*. Toronto: Multi-Health Systems.

Hare, R. D. (2003). *The Psychopathy Checklist Revised* (2nd ed.). Toronto: Multi-Health Systems.

Hedges, L. V., & Olkin, I. (1985). *Statistical methods for meta-analysis.* Orlando, FL: Academic Press.

*Hilterman, E. (2007, June). *Use of the SAVRY by clinicians and its relation with recidivism by juveniles in Catelonia, Spain.* Paper presented at the International Association of Forensic Mental Health Services, Montreal, Quebec, Canada.

Hoge, R. D. (1999). An expanded role for psychological assessments in juvenile justice systems. *Criminal Justice and Behavior, 26,* 251-266.

Hoge, R. D. (2001). *The juvenile offender: Theory, research, and applications.* Boston: Kluwer.

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

Hoge, R. D., & Andrews, D. A. (1996). *Assessing the youthful offender: Issues and techniques.* New York: Plenum.

Hoge, R. D., & Andrews, D. A. (2001). *The Youth Level of Service/Case Management Inventory (YLS/CMI) Screening Version.* Ottawa, Ontario, Canada: Carleton University.

Hoge, R. D., & Andrews, D. A. (2003). *Youth Level of Service/Case Management Inventory (YLS/CMI) user's manual.* Toronto: Multi-Health Systems.

Holsinger, A. M., Lowenkamp, C. T., & Latessa, E. J. (2006). Predicting institutional misconduct using the Youth Level of Service/Case Management Inventory. *American Journal of Criminal Justice, 30,* 267-284.

*Ilacqua, 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.

*Jack, L. A. (2000). *Psychopathy, risk/need factors, and psychiatric symptoms in high-risk youth: Relationships between variables and their link to recidivism*. Unpublished doctoral dissertation, Simon Fraser University, Burnaby, British Columbia, Canada.

*Jung, S., & Rawana, E. P. (1999). Risk and need assessment of juvenile offenders. *Criminal Justice and Behavior, 26,* 69-89.

*Långström, N., & Grann, M. (2000). Risk for criminal recidivism among young sex offenders. *Journal of Interpersonal Violence, 15,* 855-871.

*Långström, N., & Grann, M. (2002). Psychopathy and violent recidivism among young criminal offenders. *Acta Psychiatrica Scandinavica, 106,* 86-92.

Lipsey, M. W., & Wilson, D. B. (2001). *Practical meta-analysis. Applied social research methods series* (Vol. 49). Thousand Oaks, CA: Sage.

*Livsey, S. E. (2005). *Is the YLS a valid predictor of juvenile recidivism?* Unpublished doctoral dissertation, Michigan State University, East Lansing.

*Lodewijks, H. P. B., Doreleijers, T. A. H., & de Ruiter, C. (2008). SAVRY risk assessment in violent Dutch adolescents. *Criminal Justice and Behavior, 35,* 696-709.

*Luong, D. (2007). *Risk assessment and community management: The relationship between implementation quality and recidivism*. Unpublished master's thesis, University of Saskatchewan, Saskatoon, Saskatchewan, Canada.

*Marczyk, G. (2002). *Predicting juvenile recidivism and validating juvenile risk factors in an urban environment*. Unpublished doctoral dissertation, Medical College of Pennsylvania and Hahnemann University, Philadelphia.

*McEachran, A. (2001). *The predictive validity of the PCL:YV and the SAVRY in a population of adolescent offenders*. Unpublished master's thesis, Simon Fraser University, Burnaby, British Columbia, Canada.

*McKinnon, L. J. (2004). *Predicting risk of violence in a young offender population: The predictive validity of the PCL:YV and the YLS/CMI*. Unpublished master's thesis, Lakehead University, Thunder Bay, Ontario, Canada.

*Meyers, J. R., & Schmidt, F. (2008). Predictive validity of the Structured Assessment for Violence Risk in Youth (SAVRY) with juvenile offenders. *Criminal Justice and Behavior, 35,* 344-355.

*Morton, K. E. (2003). *Psychometric properties of four risk assessment measures with male adolescent sexual offenders*. Unpublished master's thesis, Carleton University, Ottawa, Ontario, Canada.

Neumann, C. S., Kosson, D. S., Forth, A. E., & Hare, R. D. (2006). Factor structure of the Psychopathy Checklist: Youth Version (PCL:YV) in incarcerated adolescents. *Psychological Assessment, 18,* 142-154.

*Nowicka-Sroga, M. (2004). *The LSI-OR: A recidivism follow-up study within a sample of male young offenders*. Unpublished doctoral dissertation, University of Ottawa, Ottawa, Ontario, Canada.

*Odgers, C. L. (2006). [Updated PCL-YV outcome data in a sample of female young offenders]. Unpublished raw data.

Odgers, C. L., Reppucci, N. D., & Moretti, M. M. (2005). Nipping psychopathy in the bud: An examination of the convergent, predictive, and theoretical utility of the PCL-YV among adolescent girls. *Behavioral Sciences and the Law, 23,* 743-763.

*O'Neill, M. L., Lidz, V., & Heilbrun, K. (2003). Adolescents with psychopathic characteristics in a substance abusing cohort: Treatment process and outcomes. *Law and Human Behavior, 27,* 299-313.

*Onifade, E., Davidson, W., Campbell, C., Turke, G., Malinowski, J., & Turner, K. (2008). Predicting recidivism in probationers with the Youth Level of Service Case Management Inventory (YLS/CMI). *Criminal Justice and Behavior, 35,* 474-483.

Parks, G. A. (2004). *Juvenile sex offender recidivism: Typological differences in risk assessment*. Unpublished doctoral dissertation, Walden University, Minneapolis, MN.

*Parsons, E. M. (2005). *Validity of the PCL:YV for predicting recidivism among incarcerated male young offenders*. Unpublished master's thesis, University of Alberta, Edmonton, Alberta, Canada.

*Rector, B., Wormith, J. S., & Banka, D. (2007, June). Predictive validity of the LSI-SK Youth Edition. In D. A. Andrews (Chair), *Level of Service Inventory: Risk/need assessment of female, young and Aboriginal offenders*. Symposium conducted at the First North American Correctional and Criminal Justice Psychology Conference, Ottawa, Ontario, Canada.

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

*Ridenour, T. A., Marchant, G. J., & Dean, R. S. (2001). Is the Revised Psychopathy Checklist clinically useful for adolescents? *Journal of Psychoeducational Assessment, 19,* 227-238.

*Rieger, M., Stadtland, C., Freisleder, F. J., & Nedophil, N. (2008). Psychiatrische Beurteilung des Gewaltrisikos im Jugendalter [Psychiatric assessment of violence risk in adolescence]. *Nervenarzt*. Epub ahead of print September 29, 2008. DOI: 10.1007/s00115-008-2574-3.

*Rowe, R. C. (2002). *Predictors of criminal offending: Evaluating measures of risk/needs, psychopathy, and disruptive behavior disorders*. Unpublished doctoral dissertation, Carleton University, Ottawa, Ontario, Canada.

*Salekin, R. (2008). Psychopathy and recidivism from mid-adolescence to young adulthood: Cumulating legal problems and limiting life opportunities. *Journal of Abnormal Psychology, 117,* 386-395.

*Schmidt, F., Hoge, R. D., & Gomes, L. (2005). Reliability and validity analyses of the Youth Level of Service/Case Management Inventory. *Criminal Justice and Behavior, 32,* 329-344.

*Schmidt, F., McKinnon, L., Chattha, H. K., & Brownlee, K. (2006). Concurrent and predictive validity of the Psychopathy Checklist: Youth Version across gender and ethnicity. *Psychological Assessment, 18,* 393-401.

Schwalbe, C. S. (2007). Risk assessment for juvenile offenders: A meta-analysis. *Law and Human Behavior, 31*, 449-462.

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

Shields, I. W., & Simourd, D. J. (1991). Predicting predatory behavior in a population of young offenders. *Criminal Justice and Behavior, 18,* 180-194.

*Skowron, C. (2004). *Differentiation and predictive factors in adolescent sexual offending*. Unpublished doctoral dissertation, Carleton University, Ottawa, Ontario, Canada.

*Stockdale, K. C. (2008). *The validity and reliability of the Violence Risk Scale–Youth Version (VRS-YV)*. Unpublished doctoral dissertation, University of Saskatchewan, Saskatoon, Saskatchewan, Canada.

*Thompson, A. P., & Pope, Z. (2005). Assessing juvenile sex offenders: Preliminary data for the Australian adaptation of the Youth Level of Service/Case Management Inventory (Hoge & Andrews, 1995). *Australian Psychologist, 40,* 207-214.

*Upperton, R. A., & Thompson, A. P. (2007). Predicting juvenile offender recidivism: Risk-need assessment and juvenile justice officers. *Psychiatry, Psychology, and Law, 14,* 138-146.

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

*Viljoen, J. L., Scalora, M., Cuadra, L., Bader, S., Chavez, V., Ullman, D., et al. (2008). Assessing risk for violence in adolescents who have sexually offended: A comparison of the J-SOAP-II, J-SORRAT-II, and SAVRY. *Criminal Justice and Behavior, 35,* 5-23.

Vincent, G. M., Odgers, C. L., McCormick, A. V., & Corrado, R. R. (2008). The PCL: YV and recidivism in male and female juveniles: A follow-up into young adulthood. *International Journal of Law and Psychiatry, 31,* 287-296.

Webster, C. D., Douglas, K. S., Eaves, D., & Hart, S. D. (1997). *HCR-20: Assessing the risk for violence* (*Version 2*). Vancouver, British Columbia, Canada: Simon Fraser University, Mental Health, Law, and Policy Institute.

*Welsh, J. L., Schmidt, F., McKinnon, L., Chattha, H. K., & Meyers, J. R. (2008). A comparative study of adolescent risk assessment instruments. *Assessment, 15,* 104-115.

Wormith, J. S., Olver, M. E., Stevenson, H., & Girard, L. (2007). The long term prediction of offender recidivism using diagnostic, personality, and risk/need approaches to offender assessment. *Psychological Services, 4,* 287-305.

Youth Criminal Justice Act, S.C. 2002, c. 1 (2002).

Mark E. Olver is an assistant professor in the Department of Psychology at the University of Saskatchewan. He also works clinically with youth and adult offenders and supervises a young offender practicum.

Keira C. Stockdale is a clinician on the Young Offender Team in the Saskatoon Health Region who provides assessment and treatment services to young offenders.

J. Stephen Wormith is a professor in the Department of Psychology at the University of Saskatchewan, where he teaches courses in offender assessment, intervention, and meta-analytic techniques. He also works clinically providing forensic psychological assessments and expert testimony for the court.