Does One Size Fit All?: A Meta-Analysis Examining the Predictive Ability of the Level of Service Inventory (LSI) With Aboriginal Offenders

Holly A. Wilson and Leticia Gutierrez

Criminal Justice and Behavior 2014 41: 196 originally published online 28 August 2013
DOI: 10.1177/0093854813500958

The online version of this article can be found at:
http://cjb.sagepub.com/content/41/2/196

Published by:
SAGE
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

>> Version of Record - Jan 3, 2014

OnlineFirst Version of Record - Aug 28, 2013

What is This?
DOES ONE SIZE FIT ALL?

A Meta-Analysis Examining the Predictive Ability of the Level of Service Inventory (LSI) With Aboriginal Offenders

HOLLY A. WILSON
Ryerson University

LETICIA GUTIERREZ
Public Safety Canada

The application of common risk assessment measures, such as the Level of Service Inventories (LSI), to Aboriginal offenders has been a criticized practice. The belief that Aboriginal offenders have distinct needs has informed the argument that existing risk-need assessments cannot adequately capture their risk. To explore this, the present meta-analysis reviewed 16 samples to test the extent to which LSI scores predict recidivism for Aboriginal compared with non-Aboriginal offenders. In addition, one large sample was used to examine the similarities in recidivism rates per LSI score for Aboriginal and non-Aboriginal offenders. Results indicated that the LSI predicts recidivism for Aboriginal offenders; however, for five of eight subscales, it predicts with less accuracy compared with non-Aboriginal offenders. In addition, the LSI underclassifies low-scoring Aboriginal offenders, but accurately estimates recidivism rates for higher scoring offenders. Implications for research into culturally-specific risk factors and the application of current risk factors to Aboriginal offenders are explored.

Keywords: Aboriginal, risk, prediction, Level of Service Inventory, recidivism, meta-analysis

The applicability of risk assessments to culturally/ethnically diverse offenders (e.g., Black, Hispanic, Aboriginal/Indigenous) has become an increasingly contentious issue. As commonly used risk assessments have primarily been developed using samples of primarily Caucasian male offenders, their use with offender subgroups (e.g., women

AUTHORS’ NOTE: We would like to thank Karl Hanson for his guidance and training on the methodological aspects of this meta-analysis, as well as feedback on the manuscript. Our thanks to Kelly Babchishin and Leslie Helmus for conducting training on analyses. Special thanks to Stephen Wormith and Sarah Hogg for providing access to data for the calibration analyses. Thanks to Jim Bonta for providing data and information on the Level of Service instruments. Also, our thanks to Kimberly Lavoie for her feedback on the manuscript. Finally, a great deal of thanks to those authors who provided additional information from which to code their studies and/or provided us with their data sets (Duyen Luong, Albert Brews, Carrie Tanasichuk, and Jill Rettinger). The views expressed are those of the authors and do not necessarily represent the views of Public Safety Canada. Correspondence concerning this article should be addressed to Holly A. Wilson, Department of Psychology, Ryerson University, Jorgenson Hall, 350 Victoria Street, Toronto, Ontario M5B 2K3, Canada; e-mail: Holly.Wilson@psych.ryerson.ca.

DOI: 10.1177/0093854813500958
© 2013 International Association for Correctional and Forensic Psychology

196
offenders, ethnic minorities) has been criticized (e.g., Allan & Dawson, 2004; Bloom, Owen, & Covington, 2003; Laprairie, 1995; Martel, Brassard, & Jaccoud, 2011). These criticisms are premised on the assumption that these instruments may not be valid measures of risk (or provide appropriate intervention targets) for these subpopulations, as they were not included (sufficiently, if at all) in the development of these tools (Hannah-Moffat & Shaw, 2001; Whiteacre, 2006). Therefore, many scholars and practitioners have called for the development of new risk assessments, consisting of more relevant or “culturally-specific” risk factors that provide a more accurate measure of risk for groups of minority offenders (e.g., Allan & Dawson, 2004; Hann, Harman, & Canfield, 1993; Whiteacre, 2006).

Among these subgroups of offenders are Aboriginal/Indigenous offenders who are mainly overrepresented in the justice systems of Canada, Australia, and New Zealand. In the Canadian context, for example, Aboriginal offenders constitute 3% of the adult Canadian population but account for approximately 20% of admissions to custody and probation (Calverley, 2010). Despite the growing overrepresentation of this group, there has been little research conducted in the area of investigating whether and/or how commonly used risk factors/assessments accurately capture their risk of reoffending compared with how they perform with non-Aboriginal offenders.

**RISK ASSESSMENT WITH OFFENDERS**

The assessment of risk is used at various stages of the criminal justice system. It is most commonly applied to determine the level of supervision that an offender will receive, as well as guide the type and intensity of services that will target an offender’s criminogenic needs. The past 40 years of risk assessment have shown the evolution of the field’s reliance on what have been termed *first-generation* risk assessments (i.e., unstructured professional judgment of risk) to *fourth-generation* risk assessments (e.g., Level of Service/Case Management Inventory [LS/CMI]; Andrews, Bonta, & Wormith, 2004; Service Planning Instrument [SPIn]; Van Dieten & Robinson, 2007), which encompass a more comprehensive actuarial assessment of an offender’s risk to reoffend that also facilitates the development of an intervention plan (Bonta, 1996).

Most current risk assessments rely upon the measurement of static and dynamic factors. Static risk factors are those that are largely historical and unchangeable (e.g., age of first arrest), whereas dynamic factors, also referred to as criminogenic needs, are more easily modified (e.g., antisocial associates). Criminogenic needs are factors that are related to recidivism and can serve as intervention targets for reducing one’s risk to reoffend. Meta-analyses of the general offender literature have substantiated eight risk/need factors which Andrews and Bonta (2010) refer to as the “Central Eight.” These criminogenic risk/need factors are powerful predictors of risk, seven of which also serve as meaningful targets for offender intervention across a variety of offender types. These, predominantly dynamic, factors include the following: criminal history, pro-criminal attitudes, pro-criminal associates, antisocial personality pattern, employment/education, family/marital, substance abuse, and leisure/recreation. Most of the assessments used today commonly include many of the Central Eight risk/need factors in their measures of risk.

**THE LSI INSTRUMENTS**

One of the most popular and widely used risk/needs assessments is the LSI instruments (e.g., LSI-R, LSI-OR, LS/CMI). The various versions of the LSI are used in
approximately 900 criminal justice agencies in North America (e.g., Lowenkamp, Lovins, & Latessa, 2009) with a variety of offender groups inside the institutions (e.g., Lowenkamp, Holsinger, & Latessa, 2001) as well as with offenders on community supervision (e.g., Raynor, 2007). The LSI instruments are founded on a General Personality and Cognitive Social Learning (GPCSL) theory of criminal behavior, which links criminal behavior to an individual’s assessment of the costs and benefits associated with pro-social versus pro-criminal alternatives. Based on this theory, when an individual perceives that the benefits of criminal behavior outweigh the costs (or outweigh the benefits of a pro-social alternative), this behavior is more likely to occur (Andrews & Bonta, 2010). This theory informs risk assessment as, when one conducts an assessment of risk, in essence, one is evaluating the costs and benefits associated with that individual’s criminal behavior. For example, if an individual surrounds himself or herself with peers who are supportive of criminal behavior, this individual is likely to receive/perceive benefits for this type of behavior.

The LSI instruments are theoretically and practically structured according to the GPCSL and the Central Eight risk/need factors. The first section of the LSI assessments, which is the general Level of Service Assessment, consists of 43 items (summed to generate the total risk score) and is organized into subscales that map directly on to the Central Eight (Andrews et al., 2004). The LSI instruments have consistently demonstrated acceptable psychometric properties (e.g., Bonta & Motiuk, 1992) as well as the prediction of both general and violent recidivism (e.g., Gendreau, Goggin & Smith, 2002).

While the GPCSL, and therefore the LSI, is presented as applicable to all offenders (as it intentionally pays little attention to race or gender), given its development using, primarily, male, Caucasian offenders, criticisms have been made concerning the racialized nature of the factors considered under this theory (e.g., Hannah-Moffat, 2013). Critics suggest that although factors typically included in risk assessment tools, such as the Central Eight, do not directly make reference to race, social inequality and racial discrimination contribute to the high-risk status these assessments often recommend for non-White offenders, as they argue that many criminogenic factors are implicated in the lives of marginalized individuals (Hannah-Moffat & Maurutto, 2010; Martel et al., 2011). Therefore, the ability of the GPCSL and the LSI to account for and understand factors related to the criminal behavior of non-White offenders is often questioned.

RISK ASSESSMENTS WITH ABORIGINAL OFFENDERS

Within the larger debate concerning risk assessment with minority offenders, many have specifically questioned the applicability of commonly used risk assessments with Aboriginal/Indigenous offenders, given their overrepresentation in the criminal justice systems of Canada, Australia, and New Zealand. A number of scholars have suggested that, in line with arguments made for minorities generally, the lack of culturally specific factors (e.g., spiritual identity) within these assessments introduces a cultural bias in the estimation of risk with Aboriginal offenders (Laprairie, 1995; Martel et al., 2011). Some have argued that this cultural bias creates a disadvantage for Aboriginal offenders due to the potential incongruences between the currently used risk factors and those that are (truly) related to recidivism for Aboriginal offenders, as, it is stated, the currently used, and often dynamic, risk factors individualize problems that are systemic (Allan & Dawson, 2004; Hannah-Moffat & Maurutto, 2010).
Furthermore, it is argued that the use of these assessments, founded on cultural values/experiences largely representative of a non-Aboriginal society, not only improperly capture the true risk of an Aboriginal offender to reoffend (due to the inadequacy of the factors), they serve to further disadvantage the Aboriginal offender when used for criminal justice decision making (e.g., risk assessment, security classification, program allocation; for example, Allan & Dawson, 2004; Hann et al., 1993; Martel et al., 2011). They argue that the factors Aboriginal offenders are being assessed for make them appear higher risk, which is a reflection of the (in)appropriateness of the factors, and may not represent the actual risk level of the offenders. Or, if they are in fact high risk, we need to gain a better understanding of which factors are (potentially uniquely) criminogenic for Aboriginal offenders so we can target them more effectively through treatment. By assessing Aboriginal offenders on factors that were not developed according to their own historical, cultural, and experiential reality, we run the risk of superimposing a framework of assessment onto a group for whom, for the most part, it does not fit (Day, Howells, & Casey, 2003).

For example, Martel et al. (2011) suggest that current risk assessments (potentially artificially) classify Aboriginal offenders as higher risk and this inappropriate classification of risk level leads to fewer treatment opportunities for Aboriginal offenders, as they are more likely to be placed in maximum security and granted less access to programming. This practice, in turn, serves to isolate these offenders and exclude them (contrary to the “socialization” goals of group programming), which is paradoxically reflective of their sociohistorical experience of colonization. Martel et al. (2011) describe these culturally biased practices in risk management as engendering a “discriminatory effect” for Aboriginal offenders. By having these offenders cycle through the system without adequate programming, this effect, in turn, materializes as an overrepresentation of Aboriginal peoples in the justice system.

The assumption that Aboriginal offenders require different, culturally specific assessments of risk is, in part, attributable to the notion that Aboriginal offenders possess different cultural values/practices compared with non-Aboriginal offenders (e.g., Heckbert & Turkington, 2001; Lapraire, 1995; Martel et al., 2011). These differences are in part due to the historical marginalization and discrimination Aboriginal peoples have experienced either directly (e.g., through the experience of residential schools), or indirectly (e.g., through intergenerational/vicarious trauma). It is argued that these differences are relevant to understanding Aboriginal offenders’ criminal behavior. These historical/cultural differences are what some scholars have assumed are indicative of differential risk factors, which, if targeted, would reduce risk.

There has, however, been little empirical research to-date defining or examining the relationship between culturally specific risk factors for Aboriginal/Indigenous offenders and reoffending. Some have suggested factors such as cultural isolation, experience of residential schools, attitudes toward marginalization, intergenerational/vicarious trauma of colonization, and loss of Aboriginal spirituality/practices are important to understanding Aboriginal offending; however, these factors have only been modestly explored in a largely narrative or qualitative manner (e.g., Ellerby & MacPherson, 2002; Heckbert & Turkington, 2001).

Despite the criticisms of a potential cultural bias, the risk assessments currently used within the justice system are applied with relative consistency to Aboriginal offenders. Individual studies examining the validity of risk assessments with Aboriginal offenders from...
a variety of offender types (e.g., violent, sexual) have shown many common measures to be predictive of recidivism for Aboriginal offenders. Examples of these risk assessments include the following: the Statistical Information on Recidivism Scale (SIR; Nafekh & Motiuk, 2002); the Community Risk/Needs Assessment (CRNA; British Columbia Corrections Branch, 2004); Psychopathy Checklist: Youth Version (PCL: YV; Schmidt, McKinnon, Chattha, & Brownlee, 2006); and the Static-99 (Babchishin, Blais, & Helmus, 2012).

Research examining the applicability of common risk factors for non-Aboriginal offenders to Aboriginal offenders has been sparse, and mainly qualitative in nature. Following Rugge’s (2006) narrative review of risk assessment with male Aboriginal offenders, a meta-analysis was recently conducted examining the predictive ability of the Central Eight risk/need factors with Aboriginal offenders compared with non-Aboriginal offenders (Gutierrez, Wilson, Rugge, & Bonta, 2013). Based on a total of 49 independent samples (n = 57,315 Aboriginal and 204,977 non-Aboriginal offenders), this review found that the Central Eight risk/need factors significantly predicted general and violent recidivism for Aboriginal offenders. However, three of the Central Eight factors predicted recidivism significantly better for non-Aboriginal offenders (i.e., criminal history, alcohol/drug, and antisocial pattern). Given that there is variability in how the Central Eight can be measured, the findings of Gutierrez et al. (2013) raise the question as to how these groups compare when assessed on a single consistent measure of the Central Eight? The present study attempts to investigate this question further using the available empirical literature on LSI instruments with Aboriginal offenders.

**PURPOSE OF PRESENT STUDY**

The present study has two parts. Part 1 is a meta-analysis of the predictive ability of the LSI among Aboriginal and non-Aboriginal offenders. This first part explores two questions: (a) “Does the LSI predict recidivism with Aboriginal offenders?” and (b) “Does the LSI predict recidivism equally for Aboriginal and non-Aboriginal offenders?” Given that differences in discrimination (i.e., whether the LSI scores are higher for recidivists than nonrecidivists for Aboriginal offenders compared with non-Aboriginal offenders) were found, Part 2 of the study is a post hoc exploration of one possible explanation for the observed pattern using a large sample of Aboriginal (n = 1,692) and non-Aboriginal (n = 24,758) offenders from the same jurisdiction (Ontario, Canada). The case-specific data in Part 2 allowed for examination of calibration (i.e., match between expected and observed rates of recidivism) in the two offender groups.

**METHOD (PART 1)**

**SELECTION OF STUDIES**

The search was initially conducted for the Gutierrez et al. (2013) meta-analysis up to August 31, 2010. This consisted of computer searches of PsycINFO, Web of Science, iPortal, Criminal Justice Abstracts, ProQuest, Dissertation Abstracts and the National Criminal Justice Reference System (NCJRS) using key terms restricting the search to Aboriginal offenders. Additional articles were obtained through emails to established researchers in the field of risk assessment and Aboriginal offenders. Some of the researchers provided data sets, without personal identifying information, that were used for their studies,
dissertations, and conference presentations. The period under review was January 1, 1988, to November 30, 2011.

Studies were considered for inclusion in the meta-analysis according to three criteria. The first, similar to that outlined in Gutierrez et al. (2013), required that a study include a sample of offenders who identified as Aboriginal, either by Status or self-report. This included any Aboriginal group or subgroup (nation/tribe) as well as any offender group (i.e., general offender, sexual, violent). Second, the study must have investigated the predictive validity of the Level of Service Inventory (LSI). This included any derivatives of the LSI (e.g., LSI, LSI-OR, LSI-R) and any section of the measure (e.g., total score only or individual items). The final criterion required studies to include a follow-up period with some form of recidivism outcome (e.g., general, violent). Recidivism rates must have been reported specifically for the Aboriginal sample and not simply for an entire sample that included Aboriginal offenders.

Studies had to include sufficient statistical information to calculate an effect size (i.e., Cohen’s $d$) and the recidivism rate. For articles where critical information was not reported (e.g., base rates), attempts were made to contact the original authors for the missing information. Studies were also checked for overlapping samples, with the study providing the largest sample size and longest follow-up being chosen for inclusion. The search yielded 12 usable documents (i.e., published articles, government reports, conference presentations, unpublished dissertations, and data sets) originating from 3 different countries: Canada (10), the United States (1), and Australia (1). The average sample size was 2,144, ranging from 46 to 24,758.

Five data sets were provided to us by researchers. This allowed us to use the raw data to code variables that were published, unpublished or referred to in conference presentations. Raw correlations between predictors and outcomes were calculated and coded regardless of their level of statistical significance.

**CODING PROCEDURE**

Section one of the LSI, which focuses on the assessment of the general risk/need factors that contribute to the total score, was the primary focus of this meta-analysis. The predictive validity of the LSI was analyzed at the risk/need item- and subscale-level as well as for the total score. Several of the LSI items (e.g., dissatisfaction with marital or equivalent situation) are scored on a 4-point scale, ranging from 0 to 3. Scores of 0 or 1 represent a very or relatively unsatisfactory situation with a need for improvement, respectively. Scores of 2 or 3 represent a relatively satisfactory or satisfactory situation with little to no need for improvement, respectively. For the purpose of this meta-analysis, scores were grouped with 0 or 1 representing the presence of the risk factor and 2 or 3 representing no risk factor present. Both the subscales and the total score are interval/continuous variables.

Several sample (e.g., majority of sample male) and study (e.g., setting from which offender was released) variables were coded as potential moderators. The source from which the LSI was coded was also examined as a potential proxy for design quality, as the LSI was designed to be coded based on file information as well as an interview (Andrews et al., 2004). However, 11 studies used both file and interview information, whereas only 1 used file only. This did not meet our criteria for moderator analyses (requiring a minimum of 3 studies in each category); therefore, this variable was not included.
The two main dependent variables were general or any recidivism and violent recidivism (including sexual crimes). Recidivism for each outcome variable was measured in different ways in the studies (e.g., new arrest, reincarceration). When more than one outcome was reported in a study, the outcome chosen for calculating the effect size was assigned according to the following ranking: (a) reconviction, (b) reincarceration (excluding technical violations), (c) all inclusive reincarceration, (d) a new arrest or charges, and (e) any other disposition (e.g., breach of conditions, parole violation).

All studies (including the data sets) were coded by the first author. To measure interrater reliability, the second author coded two of the seven studies and one of the five data sets. The interrater agreement of the effect sizes was 100%. The high interrater reliability is not surprising as most effect sizes were calculated using a 2 × 2 table and standard workbook for calculating $d$s. Most discrepancies involved rounding errors, which were corrected for the final rating. In the three reliability studies, Rater 1 identified 635 effect sizes and Rater 2 identified 634 effect sizes.

INDEX OF PREDICTIVE ACCURACY

Research Question 1

The effect size indicator was the standardized mean difference between recidivists and nonrecidivists (Cohen’s $d$) and it was calculated from Hasselblad and Hedges (1995) as reported in Gutierrez et al. (2013). Given that $d$ values are less influenced by recidivism base rates, it was chosen over other common effect size indicators (e.g., $r$) that are more affected by base rates. When interpreting $d$, we followed Cohen’s (1988) guidelines where $d$ values of .20 are considered “small,” .50 “medium,” and .80 as “large.” When the 95% confidence interval (CI) does not contain 0, the $d$ value can be considered statistically significant (i.e., $p < .05$). Nonoverlapping CIs for two predictors indicate that the two predictors are significantly different from one another.

For 2 × 2 tables, the variance of Cohen’s $d$ was estimated using Formula 19 from Sánchez-Meca, Marin-Martínez, and Chacón-Moscoso (2003), with 0.5 added to each cell to account for empty cells during analysis (Fleiss, 1994). When calculating Cohen’s $d$ from other statistics (e.g., means, ROC areas, regression betas), the variance was estimated using Formula 3 from Hasselblad and Hedges (1995).

Research Question 2

To reduce the within-study variability, we restricted our analysis to studies that reported the predictive accuracy of the LSI for both Aboriginal and non-Aboriginal samples. For both samples, Cohen’s $d$ was calculated for each predictor using the same procedure described above. The index used to assess the difference in effectiveness of the items, subscales, and total score was the $d_i$ difference, which was calculated for each predictor within each study by subtracting the $d_i$ of the non-Aboriginal sample from the $d_i$ of the Aboriginal sample. The variance for the $d_i$ difference was calculated according to Ley (1972).2

AGGREGATION OF FINDINGS

Analysis for both research questions was aggregated using the same procedure, despite different effect size measures. When summarizing the findings, weighted mean values were
used (Hasselblad & Hedges, 1995). More weight was given to larger samples by weighting each $d_i$ (or $d_i$ difference for Research Question 2) by the inverse of its variance, which was used to estimate 95% CI.

The results were reported for fixed and random effects. Fixed effect models restrict conclusions to the studies examined in the meta-analysis as it only considers within-study variability. This results in narrow CIs and oftentimes a more liberal interpretation of results (Hedges & Vevea, 1998). Random effects models take into consideration between-study variability and, therefore, produce CIs with a wider range than those of a fixed effect model. This model provides more conservative estimates and allows for generalization of results outside of the observed set of studies (Hedges & Vevea, 1998).

To assess the homogeneity of variance, the $Q$ statistic was utilized (Hedges & Olkin, 1985). The $Q$ statistic is commonly used to test the generalizability of effects across studies and follows a chi-square distribution with $k-1$ degrees of freedom ($k =$ number of studies). $Q$ values higher than the predetermined statistical level of $p < .05$ indicate that there are significant differences among studies. The $Q$ statistic only indicates that heterogeneity exists; however, the degree of heterogeneity can be quantified using the $I^2$ statistic estimated from Huedo-Medina, Sánchez-Meca, Marín-Martínez, and Botella (2006) using Formula (10).3

According to Huedo-Medina et al. (2006), percentages of 25, 50, and 75 indicate small, medium, and large proportions of heterogeneity, respectively. A negative value of $I^2$ was interpreted as 0. Furthermore, for Research Question 1, moderator analysis was conducted to examine the influence of specific study and sample variables on the predictive accuracy of the LSI items, subscales, and total score for Aboriginal offenders. The $Q$-change statistic (also known as $Q$-between or $Q$Δ) was used which tests whether the magnitude of the effect size is significantly associated with a given variable. $Q$-change values higher than $p < .05$ indicate that there is a significant difference in the predictive validity of the risk factor as a function of the moderator variable. These analyses required a minimum of three effect sizes contributing to each category of the moderator (e.g., for Setting, there had to be a minimum of three effect size estimates in the Community category and three in the Custody category).

Outliers and an unusually large sample size were considered to help ensure stable results. Outliers were identified by consideration of the weighted sum of squares ($wss$; a measure of the contribution of each study to the mean weighted $d*$) and any single extreme value of $d_i$ compared with the mean weighted effect size. Outliers were excluded from each category if the single value accounted for more than 50% of the total variance ($Q$). The presence of outliers was not considered if there were fewer than four studies in a category contributing to the mean effect size or the $Q$ was not significant.

Given that Cohen’s $d$ takes into consideration sample size in estimating the weight each study is given to the mean effect size, unusually large sample sizes can influence the pattern of results when few studies are contributing to the overall effect. In the present study, the large sample size ($n = 26,450$) from Wormith and Hogg (2012) was initially a concern for analyses consisting of only three studies (e.g., item level). However, in calculating Cohen’s $d$, the weight allotted to each study is reduced as the split between the two groups (i.e., Aboriginal and non-Aboriginal) deviates from 50%. In this case, the Aboriginal sample ($n = 1,692$) only represents 6% of the entire sample (non-Aboriginal $n = 24,758$) from Wormith and Hogg (2012); therefore, this study ultimately provided no undue influence.
A total of 12 studies (7 reports/articles and 5 data sets) yielded 16 independent Aboriginal samples, which provided 1,186 unique effect sizes. Eight of the studies (66.7%) were peer reviewed (which include peer reviewed journal publications and theses/dissertations) as of November 2011. The majority of the studies were conducted in Canada (83.3%, \( k_i = 10 \)), with the remaining 2 studies from the United States and Australia. The dates of completion for these studies range from 1989 to 2012; however, the majority of studies were completed after 2006. The average follow-up time for the combined sample of Aboriginal and non-Aboriginal offenders was 29.7 months (\( SD = 19.9, k_i = 12 \)).

Table 1 provides a summary description of the Aboriginal and non-Aboriginal samples. The majority of the total sample consisted of general/mixed adult male offenders released from community supervision. The mean age for the sample was 28.7 years (\( SD = 8.0, k_i = 12 \)). The average index sentence length was 13.1 months (\( SD = 4.2, k_i = 4 \)) and the mean length of time ever in prison/hospital was 11.2 months (\( SD = 14.4, k_i = 6 \)). The unweighted recidivism rate for general/any recidivism for the Aboriginal and non-Aboriginal samples was 55.5% and 38.7%, respectively. For violent recidivism, the rates were 32.4% and 31.6%.

**PART 1: QUESTION 1**

**Predictive Ability of the LSI for Aboriginal Offenders**

Evaluating the predictive ability of the LSI with Aboriginal offenders involved an examination of LSI items, subscales, and total scores. These analyses were conducted for general/any recidivism (see Table 2) as well as violent recidivism (tables for violent recidivism are available from the first author). Although we report the findings for fixed and random effects analyses in the tables, only the results for random effects analyses are reported in the

---

**TABLE 1: Characteristics of Study and Sample Information for Aboriginal and non-Aboriginal Groups**

<table>
<thead>
<tr>
<th>Characteristics</th>
<th>Aboriginal (n = 21,807)</th>
<th>Non-Aboriginal (n = 42,515)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>%</td>
<td>k/15</td>
</tr>
<tr>
<td>Study characteristics (% of ( k_i ))</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sample type</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Adult</td>
<td>80.0</td>
<td>12</td>
</tr>
<tr>
<td>Juvenile</td>
<td>20.0</td>
<td>3</td>
</tr>
<tr>
<td>Setting released from</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Community</td>
<td>66.7</td>
<td>10</td>
</tr>
<tr>
<td>Custodial/residential</td>
<td>33.3</td>
<td>5</td>
</tr>
<tr>
<td>Sample characteristics (% of ( N ))</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Majority male (75% or more)</td>
<td>67.8</td>
<td>15</td>
</tr>
<tr>
<td>Previous criminal record</td>
<td>45.8</td>
<td>6</td>
</tr>
<tr>
<td>Offender type</td>
<td></td>
<td></td>
</tr>
<tr>
<td>General/mixed</td>
<td>93.3</td>
<td>14</td>
</tr>
<tr>
<td>Violent (only)</td>
<td>7.7</td>
<td>1</td>
</tr>
</tbody>
</table>

*Note. \( k \) represents the number of independent samples from which the information was derived.*
body of the text for subscales and total scores, as random effects provides a more conservative estimate. For the item-level results, fixed and random are also reported in the tables; however, the findings of the fixed analyses will be reported in the text, given that most of those analyses are based on only three studies and therefore generalizability is limited.

### LSI Subscales and Total Scores

For fixed and random effects analyses, each of the eight subscales of the LSI significantly predicted general recidivism for Aboriginal offenders. The mean effect sizes ranged from $d = 0.24$ (95% CI = [0.15, 0.33]) for Family/Marital to $d = 0.60$ (95% CI = [0.53, 0.67]) for Antisocial Pattern (see Table 2). The best predictors were Antisocial Pattern, Criminal History and Employment/Education; however, significant variability, as measured by $Q$ and $I^2$, was observed for all of the subscales except for Employment/Education and Antisocial Pattern. Similarly, for violent recidivism, each of the subscales was significantly predictive. The mean effect sizes ranged from $d = 0.13$ (95% CI = [0.07, 0.18]) for Family/Marital to $d = 0.52$ (95% CI = [0.31, 0.72]) for Criminal History. The best predictors were Criminal History, Antisocial Pattern, and Procriminal Attitude-Orientation; however, each of these (as well as Alcohol/Drug) had significant variability.

For the total scores, the combined LSI Scales total scores (including LSI, LSI-R, LSI-OR, and LSI-SK) significantly predicted general recidivism ($d = 0.62$, 95% CI = [0.46, 0.79]) and violent recidivism ($d = 0.49$, 95% CI = [0.42, 0.56]); however, there was significant variability for the LSI total scores for general recidivism (see Table 2). There was only a sufficient number of studies ($k > 2$) to examine separately the LSI-OR (for general and violent recidivism) and LSI-R (for general recidivism). It was found that these total scores significantly predicted general recidivism and the LSI-OR also predicted violent recidivism.

### Table 2: Predictive Accuracy of LSI Subscales and Total Scores for Aboriginal Sample (General/Any Recidivism)

<table>
<thead>
<tr>
<th>Subscale/Total</th>
<th>Fixed</th>
<th></th>
<th></th>
<th></th>
<th>Random</th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Median</td>
<td>Mean</td>
<td>d</td>
<td>95% CI</td>
<td>Lower</td>
<td>Upper</td>
<td>Q</td>
<td>P%</td>
</tr>
<tr>
<td>Criminal history</td>
<td>.58</td>
<td>.55</td>
<td>.51</td>
<td>.59</td>
<td>.57</td>
<td>.42</td>
<td>.72</td>
<td>93.96*</td>
</tr>
<tr>
<td>Employment/education</td>
<td>.49</td>
<td>.40</td>
<td>.36</td>
<td>.44</td>
<td>.44</td>
<td>.31</td>
<td>.57</td>
<td>69.14*</td>
</tr>
<tr>
<td>Outliers removed</td>
<td>.53</td>
<td>.53</td>
<td>.47</td>
<td>.58</td>
<td>.52</td>
<td>.44</td>
<td>.60</td>
<td>11.89</td>
</tr>
<tr>
<td>Family/marital</td>
<td>.29</td>
<td>.23</td>
<td>.19</td>
<td>.27</td>
<td>.24</td>
<td>.15</td>
<td>.33</td>
<td>31.81*</td>
</tr>
<tr>
<td>Leisure/recreation</td>
<td>.33</td>
<td>.33</td>
<td>.29</td>
<td>.37</td>
<td>.34</td>
<td>.24</td>
<td>.45</td>
<td>40.79*</td>
</tr>
<tr>
<td>Companions</td>
<td>.42</td>
<td>.36</td>
<td>.32</td>
<td>.40</td>
<td>.39</td>
<td>.29</td>
<td>.49</td>
<td>39.14*</td>
</tr>
<tr>
<td>Alcohol/drug</td>
<td>.44</td>
<td>.38</td>
<td>.34</td>
<td>.42</td>
<td>.42</td>
<td>.28</td>
<td>.56</td>
<td>78.71*</td>
</tr>
<tr>
<td>Procriminal attitude-orientation</td>
<td>.43</td>
<td>.32</td>
<td>.28</td>
<td>.36</td>
<td>.39</td>
<td>.27</td>
<td>.51</td>
<td>59.03*</td>
</tr>
<tr>
<td>Antisocial pattern</td>
<td>.62</td>
<td>.61</td>
<td>.56</td>
<td>.66</td>
<td>.60</td>
<td>.53</td>
<td>.67</td>
<td>9.02</td>
</tr>
<tr>
<td>LS scales</td>
<td>.64</td>
<td>.59</td>
<td>.55</td>
<td>.63</td>
<td>.62</td>
<td>.46</td>
<td>.79</td>
<td>146.52*</td>
</tr>
<tr>
<td>LSI-OR</td>
<td>.76</td>
<td>.77</td>
<td>.72</td>
<td>.82</td>
<td>.76</td>
<td>.67</td>
<td>.85</td>
<td>10.83</td>
</tr>
<tr>
<td>LSI-R</td>
<td>.24</td>
<td>.29</td>
<td>.23</td>
<td>.36</td>
<td>.29</td>
<td>.23</td>
<td>.36</td>
<td>1.51</td>
</tr>
</tbody>
</table>

Note. The LS Scales variable and LS subscale scores include effect sizes derived from all LS risk assessment scales (i.e., LSI, LSI-R, LSI-OR, LSI-SK). LSI = Level of Service Inventory.

$p < .05.$

Downloaded from cjb.sagepub.com at PORTLAND STATE UNIV on January 10, 2014
for Aboriginal offenders, with no significant variability for either. It must be noted that the predictive ability of the LSI-R total score \( (d = 0.29, 95\% \text{ CI} = [0.23, 0.36], k = 3) \) was significantly lower than the LSI-OR total score \( (d = 0.77, 95\% \text{ CI} = [0.72, 0.82], k = 3) \) for general recidivism.

**LSI Items**

Analyses conducted at the item level showed that all 43 items significantly predicted general recidivism (for fixed and random effects analyses). The mean effect sizes ranged from the lowest \( d = 0.07 (95\% \text{ CI} = [0.01, 0.12]) \) for Item 18 (i.e., dissatisfaction with marital or equivalent situation) to \( d = 0.69 (95\% \text{ CI} = [0.63, 0.74]) \) for Item 8 (i.e., charge laid, probation breached, or parole suspended during prior community supervision). The best items were Item 8, Item 43b (i.e., three or more address changes in the last year) and Item 7 (i.e., ever punished for institutional misconduct or a behavior report). There was significant variability across studies for 15 of the 43 items.

For violent recidivism, all but five items were found to significantly predict (for fixed effects analyses); for three additional items, the results for only random effects were not significant. The mean effect sizes ranged from \( d = 0.01 (95\% \text{ CI} = [-0.05, 0.07]) \) for Item 13 (i.e., less than regular Grade 12 or equivalent) to \( d = 0.49 (95\% \text{ CI} = [0.43, 0.55]) \) for Item 8. The best items for predicting violent recidivism were Item 8, Item 1 (i.e., any prior youth dispositions or adult convictions), and Item 2 (i.e., two or more prior youth/adult dispositions/convictions). After removing outliers, there was a significant \( Q \) across studies for 14 of the items.

**MODERATOR ANALYSES**

Given that six of the eight LSI subscales and the total score had significant variability in predicting general/any recidivism, moderator analyses were undertaken. Study and sample characteristics were tested to determine whether they could account for some of the variance. Only three potential moderators could be explored for general recidivism (i.e., evaluator role, setting, and majority male sample). There was no significant variability across studies for Employment/Education and Antisocial Pattern; therefore, moderator analyses were not needed for these subscales. There were an insufficient number of effect sizes to conduct moderator analyses with violent recidivism. The strongest moderator was the evaluator role (LSI developer vs. Independent researcher/investigator) for all six subscales as well as the total score. Significantly greater predictive validity was found when the studies were conducted by a LSI developer. For Setting, significantly greater predictive validity (for Family/Marital, Procriminal Attitudes, and LSI Scales) was found for samples in a Custody setting. There was no significant \( Q\text{-change} \) for any of the subscales or total score when examined by majority male sample (i.e., whether the sample was mostly male vs. mixed).

**PART 1: QUESTION 2**

Differences in the Predictive Validity of the LSI Between Aboriginal and non-Aboriginal Offenders

Given that the analyses for Question 1 established that the LSI items, subscales, and total scores significantly predict general/any and violent recidivism for Aboriginal offenders,
this second question examines whether there is a difference in their predictive ability between Aboriginal and non-Aboriginal offenders. Using only studies that had an Aboriginal and non-Aboriginal sample, this was done by subtracting the non-Aboriginal mean weighted from the mean weighted of the Aboriginal sample. The resultant statistic is the weighted mean difference score (ddiff). Table 3 provides the ddiff for fixed and random effects analyses for general/any recidivism. Negative values indicate that the d was larger for the non-Aboriginal group (e.g., (dAB = 0.25) – (dnon-AB = 0.45) = −0.20).

### LSI Subscales and Total Scores

Using the more conservative random effects results, five of the eight LSI subscales predicted general/any recidivism significantly better for non-Aboriginal offenders than Aboriginal offenders. The results of the fixed effects analyses estimated that seven of the eight subscales predicted recidivism significantly better for non-Aboriginal offenders (see Table 3). Only the Leisure/Recreation subscale showed a consistent nonsignificant difference in predictive ability between the two groups (ddiff = −0.03, 95% CI = [−0.08, 0.02]). For the total scores (i.e., LSI Scales, LSI-OR and LSI-R), all predicted significantly better for the non-Aboriginal than Aboriginal offenders; these results were consistent for both fixed and random effects analyses. There was only significant variability for the Criminal History (Q = 35.5, df = 8, p < .05) and Family/Marital subscales (Q = 22.6, df = 8, p < .05).

For violent recidivism, the results for fixed and random effects analyses were mixed. The results for random effects showed that all subscales did not significantly differ in their predictive ability for Aboriginal compared with non-Aboriginal offenders; however, the results

### Table 3: Cohen’s d-Difference (ddiff) for LSI Subscales and Total Scores (General/Any Recidivism)

<table>
<thead>
<tr>
<th>Subscale/Total</th>
<th>Fixed Median</th>
<th>Mean ddiff</th>
<th>95% CI Lower</th>
<th>95% CI Upper</th>
<th>Random Median</th>
<th>Mean ddiff</th>
<th>95% CI Lower</th>
<th>95% CI Upper</th>
<th>Q</th>
<th>I² %</th>
<th>k</th>
<th>n</th>
</tr>
</thead>
<tbody>
<tr>
<td>Criminal history</td>
<td>−.36</td>
<td>−.18</td>
<td>−.23</td>
<td>−.14</td>
<td>−.31</td>
<td>−.44</td>
<td>−.19</td>
<td>35.46*</td>
<td>77.44</td>
<td>9</td>
<td>57,444</td>
<td></td>
</tr>
<tr>
<td>Employment/education</td>
<td>−.06</td>
<td>−.08</td>
<td>−.12</td>
<td>−.03</td>
<td>−.08</td>
<td>−.16</td>
<td>−.01</td>
<td>13.52</td>
<td>40.84</td>
<td>9</td>
<td>57,444</td>
<td></td>
</tr>
<tr>
<td>Family/marital</td>
<td>−.04</td>
<td>−.05</td>
<td>−.10</td>
<td>−.01</td>
<td>−.04</td>
<td>−.14</td>
<td>.06</td>
<td>22.59*</td>
<td>64.58</td>
<td>9</td>
<td>57,444</td>
<td></td>
</tr>
<tr>
<td>Leisure/recreation</td>
<td>−.05</td>
<td>−.04</td>
<td>−.09</td>
<td>.00</td>
<td>−.06</td>
<td>−.15</td>
<td>.03</td>
<td>19.19*</td>
<td>58.31</td>
<td>9</td>
<td>57,391</td>
<td></td>
</tr>
<tr>
<td>Outlier removed</td>
<td>−.05</td>
<td>−.03</td>
<td>−.08</td>
<td>.02</td>
<td>−.03</td>
<td>−.08</td>
<td>.02</td>
<td>4.79</td>
<td>0</td>
<td>8</td>
<td>57,198</td>
<td></td>
</tr>
<tr>
<td>Companions</td>
<td>−.19</td>
<td>−.13</td>
<td>−.17</td>
<td>−.08</td>
<td>−.13</td>
<td>−.20</td>
<td>−.06</td>
<td>12.05</td>
<td>33.62</td>
<td>9</td>
<td>57,443</td>
<td></td>
</tr>
<tr>
<td>Alcohol/drug</td>
<td>−.12</td>
<td>−.11</td>
<td>−.16</td>
<td>−.07</td>
<td>−.12</td>
<td>−.19</td>
<td>−.04</td>
<td>13.64</td>
<td>41.37</td>
<td>9</td>
<td>57,444</td>
<td></td>
</tr>
<tr>
<td>Procriminal attitude-orientation</td>
<td>−.12</td>
<td>−.07</td>
<td>−.12</td>
<td>−.03</td>
<td>−.09</td>
<td>−.16</td>
<td>−.02</td>
<td>12.67</td>
<td>36.86</td>
<td>9</td>
<td>57,414</td>
<td></td>
</tr>
<tr>
<td>Antisocial pattern</td>
<td>−.26</td>
<td>−.11</td>
<td>−.18</td>
<td>−.03</td>
<td>−.29</td>
<td>−.52</td>
<td>−.07</td>
<td>24.65*</td>
<td>79.72</td>
<td>6</td>
<td>29,503</td>
<td></td>
</tr>
<tr>
<td>Outlier removed</td>
<td>−.12</td>
<td>−.08</td>
<td>−.15</td>
<td>−.01</td>
<td>−.16</td>
<td>−.32</td>
<td>.00</td>
<td>8.79</td>
<td>54.49</td>
<td>5</td>
<td>29,310</td>
<td></td>
</tr>
<tr>
<td>LS scales</td>
<td>−.18</td>
<td>−.15</td>
<td>−.19</td>
<td>−.11</td>
<td>−.20</td>
<td>−.31</td>
<td>−.10</td>
<td>33.06*</td>
<td>69.75</td>
<td>11</td>
<td>58,719</td>
<td></td>
</tr>
<tr>
<td>Outlier removed</td>
<td>−.18</td>
<td>−.14</td>
<td>−.18</td>
<td>−.09</td>
<td>−.15</td>
<td>−.22</td>
<td>−.07</td>
<td>15.55</td>
<td>42.11</td>
<td>10</td>
<td>58,510</td>
<td></td>
</tr>
<tr>
<td>LSI-OR</td>
<td>−.18</td>
<td>−.21</td>
<td>−.28</td>
<td>−.14</td>
<td>−.32</td>
<td>−.55</td>
<td>−.10</td>
<td>16.41*</td>
<td>75.63</td>
<td>5</td>
<td>29,017</td>
<td></td>
</tr>
<tr>
<td>Outlier removed</td>
<td>−.18</td>
<td>−.18</td>
<td>−.26</td>
<td>−.11</td>
<td>−.18</td>
<td>−.26</td>
<td>−.11</td>
<td>1.05</td>
<td>0</td>
<td>4</td>
<td>28,808</td>
<td></td>
</tr>
<tr>
<td>LSI-R</td>
<td>−.12</td>
<td>−.12</td>
<td>−.18</td>
<td>−.06</td>
<td>−.12</td>
<td>−.18</td>
<td>−.06</td>
<td>0.96</td>
<td>0</td>
<td>3</td>
<td>28,225</td>
<td></td>
</tr>
</tbody>
</table>

Note. The LS Scales variable and LS subscale scores include effect sizes derived from all LS risk assessment scales (i.e., LSI, LSI-R, LSI-OR, LSI-SK). LSI = Level of Service Inventory.
•*p < .05.
for fixed effects revealed that all of the subscales predicted significantly better for non-Aboriginal offenders. Given that there were few studies contributing to the results for violent recidivism \((k = 3)\) and there was significant variability for five of the eight subscales (e.g., \(I^2 = 72\%-96\%\)), interpretation of these results warrants some caution. For the total score, only the LSI Scales (consisting of the LSI and LSI-OR total scores combined) could be explored. As was found for the subscales, the results for fixed and random were mixed and there was a significant and large proportion variability \((I^2 = 90.4\%)\).

**LSI Items**

The results for fixed effect analyses at the item level for general/any recidivism revealed that nearly half \((43.2\%, n = 19)\) of the items predicted recidivism significantly better for non-Aboriginal offenders compared with Aboriginal offenders. A total of 35 items had a mean \(d_{adj}\) in the negative direction (i.e., predicting better for non-Aboriginals) and eight items were in the positive direction. For the majority of items \((52.3\%, n = 23)\), however, there were no significant differences between groups. It must be noted that for 15 of the 23 items, the direction of the effect showed that these items were predicting recidivism better for non-Aboriginal offenders; however, this difference did not reach statistical significance. Only two items (i.e., Item 30: alcohol problem currently; Item 34: alcohol/drugs contributed to employment/school problems) predicted general/any recidivism significantly better for Aboriginal offenders than non-Aboriginal offenders. This finding was consistent for fixed and random effects analyses. Given that the majority of the results for the item-level analyses \((70.4\%)\) are based on only three studies and there was significant variability for approximately \(40\% (n = 17)\) of the items, these findings should be interpreted with some caution.

For violent recidivism, the results of the fixed effect analyses showed that 32 items (of the 41 items that had a sufficient number of studies to be analyzed) were predicting recidivism significantly better for non-Aboriginal offenders than Aboriginal offenders. These results were consistent between fixed and random for only 11 of the items. For the items showing no significant differences between groups for fixed effect analyses \((n = 9)\), eight of those items were in the direction of predicting violent recidivism better for non-Aboriginal offenders. None of the items predicted recidivism significantly better for Aboriginal offenders. Similar to the findings with general recidivism, caution in interpreting and generalizing these results is warranted given that the majority of analyses for these items \((90.2\%, n = 37)\) was based on only three studies and there was significant variability within nearly half of the items \((48.8\%, n = 20)\).

**PART 2: EXAMINATION OF CALIBRATION**

The meta-analytic findings showed that the ability of the LSI to discriminate between recidivists and nonrecidivists was significantly better for non-Aboriginal offenders than Aboriginal offenders for five of the eight LSI subscales. However, to more fully assess predictive validity requires a consideration of both discrimination (i.e., the ability of the LSI to distinguish between recidivists and nonrecidivists), as well as calibration (i.e., the ability of the LSI to estimate absolute recidivism rates). Given that Cohen’s \(d\) only provides information on the magnitude and direction of a relationship, we considered the possibility that the discrepancy in discrimination could in part be related to issues with calibration. We therefore sought the opportunity to compare the expected recidivism rate as assessed by the
LSI with the observed recidivism rate of Aboriginal offenders (i.e., calibration). This would provide information on the potential under- or overclassification of the LSI for Aboriginal offenders compared with non-Aboriginal offenders.

To preliminarily explore the issue of calibration, we conducted secondary analyses with a large, representative sample of offenders. Using the data from Wormith and Hogg (2012) on 1,692 Aboriginal offenders and 24,758 non-Aboriginal offenders assessed using the LSI-OR (reanalysis of Table 6 in Appendix C of their report), we conducted a logistic regression to explore the possibility of an interaction effect between Aboriginal status (coded as 1 for Aboriginal and 0 for non-Aboriginal offenders) and the LSI-OR total score. An interaction was tested to see whether the shape of the distribution of mean predicted recidivism by LSI-OR total score for Aboriginal offenders differed from the expected shape (i.e., the non-Aboriginal distribution). Recidivism was used as the dichotomous outcome and was coded as present (1) if the offender was reconvicted for any offense between 2004 (when they were admitted to probation/conditional sentence or released from a custodial sentence) and January 2009. This provided a minimum 4-year follow-up period.

For the Aboriginal group, 75.3% were male and 64.4% of the sample was given a conditional sentence or a term of probation. The mean age of the sample was 35.7 (SD = 10.1) at the date of data extraction. For the non-Aboriginal group, 82.2% of the sample was male and 82.4% were given a conditional sentence or probation. The mean age of the non-Aboriginal sample was 38.1 (SD = 11.8). The base rate of general recidivism was 56.9% and 33.1% for the Aboriginal and non-Aboriginal offender groups, respectively. The average LSI-OR total score was 20.7 (SD = 9.6, range = 0-43) for the Aboriginal offenders and 11.8 (SD = 8.4, range = 0-42) for the non-Aboriginal offenders. For a complete description of the Aboriginal and non-Aboriginal samples, see Wormith and Hogg (2012).

Logistic Regression

To test the shape of the distribution of recidivism by LSI-OR total score for each group first required testing whether LSI-OR total score and Aboriginal status predicted recidivism. In the first model, a test of the LSI-OR total score revealed that it significantly predicted recidivism ($\beta = 0.117$, Wald = 4,187.78, $p < .001$; see Table 4). In the second model, Aboriginal status was added as a predictor and found to significantly predict recidivism incrementally to the LSI-OR total score ($\beta = 0.150$, Wald = 6.62, $p < .05$). To assess whether the shape of the distribution of mean predicted recidivism by LSI-OR total score was different for Aboriginal offenders compared with the non-Aboriginal sample, in the third model, we tested the interaction between LSI-OR total score and Aboriginal status for predicting recidivism and it was significant ($\beta = -0.029$, Wald = 21.25, $p < .05$); however, the results of the Hosmer-Lemeshow goodness-of-fit test revealed that this predicted model was a poor fit to the data, $\chi^2 (8) = 39.75$, $p < .001$ (see Table 4 for additional goodness-of-fit indices).

To account for the additional curvature of the data (i.e., not a fully logistic shape), the predicted model required a polynomial expansion; therefore, a second-order polynomial expansion (i.e., a square function) was applied to the LSI-OR total score as well as the interaction term. The results of the Hosmer-Lemeshow test revealed that the resultant model was a good fit to the data, $\chi^2 (8) = 10.55$, $p = .229$. The final model showed that LSI-OR total score predicted recidivism ($\beta = 0.151$, Wald = 525.01, $p < .01$), Aboriginal status was
significant ($\beta = 1.052$, Wald = 19.41, $p < .01$), and the interaction was also significant ($\beta = -0.082$, Wald = 10.21, $p < .01$, see Figure 1). The full model is presented in Table 4.

Figure 1 shows the lines-of-best-fit generated from the final (expanded) logistic regression model for the mean predicted probability of recidivism by LSI-OR total score for Aboriginal and non-Aboriginal groups. To illustrate how well the predicted model fits the observed data, the absolute recidivism rates for the Aboriginal and non-Aboriginal offenders for each total score are provided in grey. The majority of the Aboriginal recidivists (65.3%) scored in the high to very high-risk categories (i.e., LSI-OR total score of 20+), 26.6% of the sample were medium risk (LSI-OR total score of 11-19), and only 8.1% of the sample fell into the very low to low risk category (i.e., LSI-OR total score of 0-10). For the non-Aboriginal offenders, 37.5% of the sample were in the high- to very-high-risk categories, 35.6% in the medium-risk category, and 27% were in the very-low- to low-risk categories.

As can be seen in Figure 1, the mean predicted probability of recidivism increased for both groups as LSI-OR total scores increased, as predicted. However, low-scoring Aboriginal offenders tended to have higher mean predicted probabilities of recidivism than expected (i.e., compared with low-scoring non-Aboriginal offenders), resulting in an under-classification of low-scoring Aboriginal offenders. As the total scores increased beyond the high-risk category (a total score of approximately 19+), the LSI-OR appeared to classify...
Aboriginal recidivists with similar accuracy as non-Aboriginal recidivists, illustrating the significant interaction that was found between LSI-OR total score and Aboriginal status.

DISCUSSION

As the Level of Service risk assessment measure was developed from a risk-need perspective, the argument that Aboriginal offenders have a distinct set of experiences and needs has put into question its applicability with this offender group. The purpose of this review was to evaluate whether the LSI predicted recidivism for Aboriginal offenders and, if so, how it compares with its accuracy with non-Aboriginal offenders. Part 1 focused on the discrimination of the LSI and demonstrated, through meta-analysis, that the LSI risk assessments significantly predict recidivism for Aboriginal offenders (Question 1); however, it demonstrated weaker predictive accuracy for five of eight subscales compared with its strength with non-Aboriginal offenders (Question 2). Part 2 examined the calibration of the LSI and found that it is better calibrated with medium to high scoring Aboriginal offenders than those who score low.

For general offenses, the LSI, in its entirety, significantly discriminated between Aboriginal recidivists and nonrecidivists, with $d$s ranging from small to medium for subscales and very small to medium for items. The LSI also significantly predicted violent recidivism at the subscale- and total score-level, though with less accuracy than general
offending as has been found in prior research (Olver, Stockdale, & Wormith, 2012). These results are consistent with previous discrimination literature demonstrating the applicability of the LSI with other offender groups, including women (Smith, Cullen, & Latessa, 2009), youth offenders (Olver, Stockdale, & Wormith, 2009), and other minority groups (Olver et al., 2012). As the LSI is structured around the Central Eight risk/need factors, the validity estimates of the eight LSI subscales also replicate the findings of Gutierrez et al. (2013) demonstrating that the Central Eight significantly predict recidivism with Aboriginal offenders; though this is not surprising, as all but one study in this review was used in the Gutierrez meta-analysis. These results also appear to provide indirect support for the generalizability of the GPCSL model (Andrews & Bonta, 2010) to Aboriginal offenders.

The results for general recidivism were also relatively stable within the moderator analyses. The findings were only consistently moderated by author association, as it was found that studies involving LSI developers produced larger effect sizes. Some argue that these larger effect sizes reflect a bias held by developers associated with, for example, failing to publish negative results, ongoing changes in study design (e.g., length of follow-up) to facilitate positive results, or expectancy biases by involved researchers (Blair, Marcus, & Boccaccini, 2008; Fanelli, 2009). However, others (most notably risk assessment developers) argue that these higher estimates are associated with greater methodological rigor and result in more precise findings (e.g., Harris, Rice, & Quinsey, 2010). To address this issue, a recent study found that the variation in the size of predictive validity for the LSI, with larger estimates associated with evaluator involvement, were in fact associated with quality implementation practices (e.g., greater assessment training, use of quality assurance protocol) rather than illegitimate behavior, such as selective reporting or willful manipulation (Andrews et al., 2011).

Another discrepancy was found in the magnitude of the effect sizes for the LSI-R ($d = .29, 95% CI = [.23, .36]$) and LSI-OR ($d = .76, 95% CI = [.67, .85]$) total scores for Aboriginal offenders. When comparing the LSI-R and LSI-OR within the $d$-difference analysis (Table 4), however, the disparity between the effect sizes disappears, as both measures predict recidivism equally poorer for Aboriginal offenders compared with non-Aboriginal offenders. An examination of the two studies using the LSI-R indicated that the three unique samples used were non-Canadian. While an in-depth exploration of the discrepancy is beyond the scope of the article, it could be that because the LSI-R was developed using Canadian offenders, it does not work as well for other groups and, therefore, produces lower effect sizes than expected.

Whereas the results for Question 1 in Part 1 provide some support for the use of the LSI with Aboriginal offenders, the caution with which it should be used is highlighted when comparing its predictive accuracy for Aboriginal and non-Aboriginal offenders. Although all subscales and the total LSI score predicted violent recidivism similarly for both groups using the more conservative random effects analysis, the total LSI score and five subscales (i.e., Criminal History, Education/Employment, Companions, Alcohol/Drugs, and Procriminal Attitude) predicted general recidivism with significantly less accuracy for Aboriginal offenders. In fact, four of the five subscales and the total LSI score had nonsignificant estimates of variance ($Q$), demonstrating a general consensus in the data. These results indicate that while the LSI predicts recidivism with Aboriginal offenders, it does so with less accuracy than with non-Aboriginal offenders. These issues with discrimination may be, in part, explained by the issues found in calibration.
An examination of Figure 1 shows that the LSI-OR appears to be underclassifying low-scoring Aboriginal offenders, as this sample reoffended at higher rates than expected compared with the low-scoring non-Aboriginal offenders. However, for the high-medium and high scorers, differences between the samples were small, indicating that the LSI-OR appears to be better calibrated with Aboriginal offenders with higher total scores than those with low scores. This calibration issue with low-scoring offenders could account for why several LSI subscales discriminated with less accuracy. The lack of precision with which five of the LSI subscales predicted recidivism for Aboriginal offenders in Question two may be explained by the higher than expected recidivism among low-scoring Aboriginal offenders.

With respect to the initial question posed in the title of this article, does one size fit all? The answer appears to be no, yet for unexpected reasons. The phrasing of this question, which aimed to capture whether the LSI works (i.e., predicts recidivism as well as it does with non-Aboriginals) with Aboriginal offenders, relies on the assumption that an offender group is homogeneous, regardless of level of risk. Perhaps the question is now better interpreted as whether the LSI, which works for some Aboriginal offenders, works for all Aboriginal offenders. With this question answered, the next question is why doesn’t it work for all Aboriginal offenders?

POSSIBLE EXPLANATIONS FOR THE CALIBRATION ISSUE

The solution to the issue of calibration, and in turn discrimination, is contingent upon the causal sources attributed to the problems of prediction. We propose four possible explanations. The first attributes the issues with calibration to racial discrimination within the criminal justice system. It has been well-established that Aboriginal offenders, compared with their non-Aboriginal counterparts, have been subjected to, for example, longer periods of incarceration, lower parole grant rates, and overconviction practices (Canada, Office of the Correctional Investigator, 2011; Law Reform Commission of Western Australia, 2006). These discriminatory practices could lead to (perceived) higher base rates of recidivism, making Aboriginal offenders appear higher risk and rendering our common risk assessment tools less predictive with this population. This explanation would support the conclusion that these low risk Aboriginal offenders are simply more likely to be caught and processed for crime than non-Aboriginal offenders, rather than more likely to truly commit more crime. Thus, these higher base rates of recidivism for low-scoring offenders would be merely an artifact of detection and does not reflect true recidivism. One potential way to examine this effect would be to match low-scoring Aboriginal offenders with low-scoring non-Aboriginal offenders and follow-up several years later by examining their self-reported crime and the number and nature of convictions.

The second potential explanation is that the same risk factors apply to both Aboriginal and non-Aboriginal offenders, but Aboriginal offenders simply have a greater number of them. As results from this meta-analysis and the previous meta-analysis examining the Central Eight with Aboriginal offenders (Gutierrez et al., 2013) have demonstrated, the standard risk constructs (e.g., education/employment, family/marital) and specific risk factors (e.g., having less than a Grade 10 education) significantly predict Aboriginal offending. However, given that Aboriginal offenders may have more risk factors in general (as indicated by their higher average LSI-OR score), the current set of risk factors in the LSI does
not adequately distinguish between recidivists and nonrecidivists. In other words, offenders who possess a particular risk factor may have a similar recidivism rate as offenders who do not possess that same risk factor, simply because they have several other risk factors. Thus, the low-scoring Aboriginal offenders in our calibration study may have a higher number of measured (by the LSI-OR) and unmeasured risk factors present within the Central Eight constructs that would result in higher average recidivism rates than expected. It may be that historical marginalization has contributed significantly to this increase in risk factors (e.g., Martel et al., 2011); however, for the purpose of risk assessment, this does not distinguish Aboriginal offenders as qualitatively different from non-Aboriginal offenders. A possible way to examine this effect would be to interview low-scoring Aboriginal offenders and ask them a greater number and variety of questions within existing constructs (i.e., the Central Eight) and examine whether an increase in risk items, and more specific risk items, provides a more accurate prediction of recidivism.

The third possible explanation for the undercalibration is that the LSI items may not be accounting for the unique experiences of Aboriginal people. As discussed in the introduction, scholars have argued that to be effective and avoid a culture-bias, risk assessment measures must consider variables that account for Aboriginal culture to more accurately capture their risk (e.g., LaPrairie, 1995; Martel et al., 2011). Though the Central Eight, as assessed by Gutierrez et al. (2013) and by an examination of the LSI subscales within this study, have been demonstrated to significantly predict recidivism with Aboriginal offenders, perhaps the items within these factors need to be operationalized differently to account for factors more related to Aboriginal offenders. Thus, perhaps making the existing items more culturally specific, on which some currently low-scoring Aboriginal offenders could score high, would increase the predictive ability of the LSI measures. For example, within the LSI subscale Family/Marital, the item evaluating whether the offender’s family members have a criminal record includes the individual’s spouse or biological family members (or, in some cases, nonbiological parental figures, such as adoptive parents). However, the definition of family and the role of community may differ within Aboriginal culture (Canada, Royal Commission on Aboriginal Peoples, 1996); therefore, the manner in which this family/marital item is coded may not accurately capture people who exert a nonbiological “familial” influence in the offender’s life, thus making the way it is currently being measured inadequate.

Finally, other scholars have argued for the consideration of entirely new culturally specific variables that also take into account experiences of marginalization. For example, Heckbert and Turkington (2001) have suggested that cultural or spiritual isolation is a prominent issue for Aboriginal peoples and that it plays a significant role in the healing and successful reintegration of offenders back into the community. Other examples of culturally specific factors that have been raised in the literature include loss of native language (Ellerby & McPherson, 2002; Mann, 2009), impact of residential schools (Mann, 2009), and lack/loss of pride in heritage (Heckbert & Turkington, 2001). Unfortunately, as demonstrated by Gutierrez et al. (2013), little research has empirically tested the relevance of these potential risk factors as they relate to recidivism; therefore, our knowledge regarding their utility in risk prediction is limited. It may be that low-scoring Aboriginal offenders score high on culturally specific items that are not currently captured in the LSI, which would in turn account for their underclassification. Future research into the predictive ability of these culturally specific factors is needed.
LIMITATIONS

Although there were sufficient studies to examine the predictive accuracy of the LSI, conclusions drawn from the item-level and violent recidivism analysis regarding differences in predictive accuracy should be interpreted with caution, as they were primarily derived from only three studies. Caution is also warranted when interpreting the subscales/items with significant variability, as this also indicates less stability in the results.

We were also limited in the variables we could examine in the moderator analyses. One variable of particular interest was the role living on/off a reserve may play in relation to the predictive ability of risk factors. It has been argued that an individual’s environment shapes the expression of Aboriginal culture (Archambeault, 2003; Canada, Royal Commission on Aboriginal Peoples, 1996), and it could be that individuals living off a reserve are more likely to identify with non-Aboriginal based risk factors. Therefore, this variable could help determine whether the accuracy of LSI for Aboriginal offenders is moderated by the offender’s proximity to a predominantly non-Aboriginal society. Similarly, we were interested in the use of the LSI with different tribes/nations (e.g., First Nations Aboriginal offenders) as well as a variety of Aboriginal groups (e.g., Métis, Inuit); however, those variables were not reported with sufficient frequency to examine their role in predicting recidivism. As nearly all studies derived from Canada, we were also unable to examine the potentially moderating role of country and criminal justice system. Therefore, generalizability of these results is primarily limited to Canada.

IMPLICATIONS FOR PRACTICE

There are considerable and meaningful practical implications for the findings of the present study. The first is that the LSI can provide valid recidivism estimates for most Aboriginal offenders. Arguments have been made suggesting that risk assessments developed on non-Aboriginal offenders cannot accurately be used with Aboriginal offenders (Martel et al., 2011; Maynard et al., 1999); however, this does not hold true for the LSI risk assessment measures. This finding, however, does not preclude the benefits of improving the accuracy of the LSI for this offender group. This may be done by addressing the difference in LSI classification for offenders scoring in the low risk range (8%-10% of the Aboriginal offenders in the Ontario sample).

If this underclassification is a result of any of the proposed explanations, much more research is needed. However, an argument can be made for applying the LSI as usual in the interim with the knowledge that low-scoring Aboriginal offenders will be underclassified. It is up to the jurisdictions using this assessment measure to balance the benefits and consequences of its use. The clear benefit is that the LSI significantly predicts recidivism with this group and that this underclassification will likely affect only a small proportion of Aboriginal offenders. However, the direction of this inaccurate classification appears to be toward increased recidivism for these offenders, which may pose a greater concern for certain jurisdictions.

The need for future research examining why these offenders are being underclassified is highlighted by the considerable implications associated with each potential explanation. If this underclassification is due to discriminatory practices and these low-scoring Aboriginal offenders are not reoffending at higher rates than similarly scoring non-Aboriginal offenders, then treating this underclassification as legitimate would lead to more sanctions (e.g., probation conditions) and, in turn, further discrimination. If, however, this underclassification with the LSI is due to (a) a greater presence of existing measured and unmeasured...
risk factors for Aboriginal offenders, (b) existing items failing to consider Aboriginal culture, or (c) the LSI failing to incorporate culturally specific items, then the LSI is legitimately underclassifying these low-scoring offenders.

The solution to this is contingent on the purpose of the risk assessment scale’s use. The mere pursuit of an accurate recidivism probability estimate would require only the renorming of the LSI to provide separate probabilities for Aboriginal offenders. However, as mentioned, with the development of different classes of risk assessment scales, the field has moved beyond second- or third-generation risk assessments, that focus almost solely on estimates of recidivism, to the use of fourth-generation risk assessments, intending to guide treatment and case planning in addition to risk prediction (Andrews, Bonta, & Wormith, 2006). Servicing these low-scoring offenders at the “right level” and with the appropriate targets requires an understanding of why they are reoffending at higher rates than expected. For example, Gutierrez et al. (2013) found that internalizing emotional problems and history of victimization (as defined by the LSI) showed similar predictive accuracy to the Central Eight for Aboriginal offenders, contrary to results in previous meta-analyses indicating that these variables had only minimal associations with recidivism (Gendreau et al., 1996). If these factors are more closely tied to recidivism for Aboriginal offenders than non-Aboriginal offenders, and may account for the lack of predictive accuracy for low-scoring Aboriginal offenders, perhaps targeting the emotional impact of previous victimization (e.g., abuse in foster care, LaPrairie, 1995; involvement in residential schools, Public Safety Canada, 2009) above and beyond interventions focusing on the Central Eight could provide greater assistance (and, in turn, risk reduction) to Aboriginal offenders than it does with non-Aboriginal offenders.

CONCLUSION

Despite the lower predictive validity of several subscales, the usefulness of the Central Eight with Aboriginal offenders should not be ignored. As Gutierrez et al. (2013) has shown, the Central Eight risk/need factors, which are clearly represented in the LSI risk assessment subscales, are significant predictors of recidivism with Aboriginal offenders and could, therefore, serve as effective treatment targets. Replications of the LSI calibration with Aboriginal offenders, in addition to more research into the cause(s) of this discrepancy, are undoubtedly needed given the potential implications for policy and practice. However, it could be that Aboriginal offenders scoring low on the LSI assessments do, in fact, more closely resemble medium-scoring offenders. Consequently, with the substantiation provided by replication, it may be that low-scoring Aboriginal offenders could benefit from greater treatment opportunities than would be afforded to them if they continued to be classified as low risk. The renorming of the LSI without additional information explaining the underclassification would impede these potentially useful treatment opportunities and, therefore, cannot be supported. As such, action should be grounded in further research into what works best with Aboriginal offenders.

NOTES

1. Specific search terms are available from the first author.
2. Using the formula, $Var(d_{diff}) = s^2 + s^2 - 2r_{xy}sxsy$, where $s$ is the standard deviation of the $d_i$ from the Aboriginal sample, $s$ is the standard deviation of the $d_i$ from the non-Aboriginal sample, and $r_{xy}$ is a correlation coefficient estimating the relationship between the average effect size for the Aboriginal sample and the non-Aboriginal sample. The correlation coefficient was derived from the risk factor with the greatest number of unique effect sizes (i.e., Total Level of Service Inventory [LSI] score), thereby providing the most stable estimate.
3. $I^2 = \left( \frac{Q - df}{Q} \times 100 \right)$.

4. $k_i$ denotes study information. For additional study information, contact the first author.

5. $k_i$ denotes independent sample information.

6. For a copy of the item-level tables, please contact the second author.

7. For a copy of the moderator table, please contact the first author.

REFERENCES

Asterisks denote studies in the meta-analysis from which effect sizes were derived.


**Holly A. Wilson**, MA, is a doctoral student in the clinical psychology graduate program at Ryerson University. She completed her master’s degree in forensic psychology at Carleton University. Her research interests include Aboriginal offenders, youth delinquency and interventions, and risk assessment.

**Leticia Gutierrez**, MA, is a research analyst in the Corrections Research Unit of the Department of Public Safety and a doctoral candidate in the forensic psychology program at Carleton University. Her areas of research include Aboriginal offenders, offender rehabilitation, community supervision, and problem-solving courts.