Validating the Level of Service Inventory—Revised on a Sample of Federal Probationers

Anthony W. Flores, M.S.
California State University, Bakersfield
Christopher T. Lowenkamp, Ph.D.
Paula Smith, Ph.D.
Edward J. Latessa, Ph.D.
University of Cincinnati

Method
Results
Discussion

THE LAST TWO DECADES have borne witness to a rise in the correctional population so colossal that it was previously inconceivable to practitioners and criminologists alike. As a result, annual exercises in economic calisthenics have become common practice for corrections administrators throughout the U.S. The complexity of such a population boom can only be fully appreciated by realizing that available resources have hardly kept pace. Expectations to do more with less, an emerging mantra among correctional administrators and practitioners, drive corrections professionals in a seemingly unending search for promising technological developments that might help to bridge the service gap created by surging offender populations and waning budgets.

The deleterious effects of this current “crisis in corrections” are becoming so entrenched in local- and state-level practices that some have remarked, “Corrections has become the Pac-Man of government budgets, gobbling up resources as legislators seek to finance competing needs with shrinking tax revenues” (Pierce, 1991 as cited in Cullen, Wright, & Applegate, 1996, p. 70). A discussion of the burdens levied by this crisis would be remiss if it overlooked the presence of these issues at the federal level. In fact, according to a most recent year-in-review report, the federal probation and pretrial services system has recently suffered severe budgetary and, consequently, operational constraints (United States Probation and Pretrial Services, 2005). Specifically identified were operational and service restrictions resulting from unprecedented budget deficiencies. In an attempt to realize some relief through cost-saving practices, the internal administration of federal probation and pretrial services recommended that: a) the agency move toward an evidence-based approach and implement a process to measure client outcome that will generate meaningful agency evaluations and subsequent operation refinements; b) the agency strive to restructure staff workloads and prioritize resources, most notably by identifying offenders eligible for early release; and c) the agency remain resolute in its commitment to provide services of exceptional quality in spite of fewer resources (United States Probation and Pretrial Services).

Actuarial classification systems that yield valid measures of risk and criminogenic need hold
considerable promise for correctional agencies in these regards. First, beyond informing decisions about custody and service provision, initial and reassessment risk/need scores also provide outcome measures useful for evaluating both offender and agency success. Second, because classification tools identify different levels of offender risk, the tools’ corresponding risk categories are inherently useful for restructuring staff workload, prioritizing agency resource expenditure, and identifying low-risk offenders for early release. Finally, actuarial classification systems can do a great deal for agencies concerned about the quality of service provision. Classification tools can improve service quality by promoting resource, custody, supervision, and treatment decisions that are better informed, more accurate, and ultimately more useful (Latessa, 2003–2004).

Classification

Classification systems disaggregate heterogeneous correctional populations into subgroups that maximize between group differences and minimize within group differences. Classification processes create these subgroups on the basis of offender characteristics relevant to correctional outcome, which in turn facilitates and justifies differential service provision. The use of classification devices allows correctional agencies to simultaneously address multiple objectives, including improved predictive accuracy, better informed treatment assignments, more effective supervision, and meaningful outcome analysis (Clements, 1996). While most correctional agencies currently use some type of classification system to guide decision making (Jones, Johnson, Latessa, & Travis, 1999), it is important to note that all classification systems are not created equal.

The approach taken to behavioral prediction can significantly affect the resultant validity. Decisions involving behavioral forecasts are grounded in either a clinical or actuarial approach (Meehl, 1954). Whereas clinical decision making is intuitively based and justified by claims of training, experience, and expertise, actuarial decision making is based on scientific evidence generated from observed behavioral outcomes for risk-similar offenders (Monahan, 1981). Research investigating the accuracy of each method consistently supports the superiority of actuarial classification decisions over clinical practices (Gottfredson & Gottfredson, 1986; Grove, Zald, Lebow, Snitz, & Nelson, 2000).

In addition to the importance of how predictions are reached, evidence also suggests that what is assessed can greatly affect the accuracy and utility of classification decisions. While the majority of actuarial classification devices rely exclusively on historically-based static risk factors (e.g., criminal history) to predict reoffending, others have evolved in response to recent knowledge advancements in risk prediction. Heeding the findings evidenced in current risk factors research (for example, see Gendreau, Little, & Goggin, 1996), researchers have developed advanced classification tools to augment traditional static-driven prediction models with a battery of predictors said to be dynamic because of their present-day (as opposed to life-history) focus. The resultant combination of both static and dynamic risk factors yields greater predictive accuracy for classification systems. For example, in an examination of male parolee revocation, Brown (2002) found that combined static and dynamic prediction models were significantly better at forecasting revocation than either static or dynamic models alone. Additionally, in a quantitative literature review comparing the averaged correlation between static risk factors and recidivism to the averaged correlation between dynamic risk factors and recidivism, findings supported the greater predictive accuracy of dynamic factors (Gendreau et al.).

Beyond improving the accuracy of predictions, devices that measure risk through a combination of static and dynamic predictors offer considerable utility to correctional agencies. Because dynamic risk factors are characteristics currently present in an offender’s life, they are inherently sensitive to detecting change and measuring progress (Taylor, 2001). Measuring both static and dynamic risk factors thus becomes a way of not only improving risk management but moving toward risk reduction (Latessa, 2003–2004). That being said, it is important to note that risk reduction depends upon correctional agencies matching their service intensity levels according to actuarial-determined offender risk levels (Lowenkamp & Latessa, 2005). Empirical investigations have consistently established that to be effective, correctional agencies must direct a majority of
their resources toward high-risk offenders. Correctional practices informed by this approach are commonly referred to as adhering to the risk principle; they offer evidence-based services entrenched in research findings that have consistently demonstrated efficacy in reducing offending while also avoiding the iatrogenic effects associated with servicing low-risk populations (Andrews, Zinger et al., 1990; Dowden & Andrews, 1999a, 1999b; Lowenkamp & Latessa, 2004a; Lowenkamp & Latessa, 2004b).

The Level of Service Inventory-Revised

One example of an actuarial classification system that measures risk and criminogenic need is the Level of Service Inventory-Revised (LSI-R). The LSI-R measures 54 risk and need factors about 10 criminogenic domains that are designed to inform correctional decisions of custody, supervision, and service provision (Andrews & Bonta, 1995). The theoretically informed predictor domains measured by the LSI-R include criminal history, education/employment, financial situation, family/marital relationships, accommodation, leisure and recreation, companions, alcohol or drug use, emotional/mental health, and attitudes and orientations (Andrews & Bonta).

The LSI-R assessment is administered through a structured interview between the assessor and offender, with the recommendation that supporting documentation be collected from family members, employers, case files, drug tests, and other relevant sources as needed (Andrews & Bonta, 1995). The total risk/need score produced by the LSI-R is indicative of the number of predictor items (out of 54) scored as currently present for the offender. The LSI-R score is then actuarially associated with a likelihood of recidivism that was derived from the observed recidivism rates of previously assessed offenders. Last, domain scores of the LSI-R are used to identify an offender’s most promising treatment targets (Andrews & Bonta).

Because the LSI-R represents a theoretically informed, empirically supported, actuarial-based, and standardized measure of criminogenic risk and need, it boasts considerable potential to improve caseload decisions, resource expenditure, and overall service quality (Andrews & Bonta, 2003; Gendreau et al., 1996).

The LSI-R and Validity

The benefits promised by any classification system must be empirically evaluated against the benefits actually observed in prior research evaluations. Research findings have generated a significant body of evidence that established the LSI-R as a valid predictor of correctional outcome across a variety of measures. Specifically, findings have supported the predictive validity of the LSI-R for institutional infractions (Bonta, 1989), probation failure (Andrews, Kiessling, Robinson, & Mickus, 1986), halfway house failure (Bonta & Motiuk, 1985; Motiuk, Bonta & Andrews, 1986), and parole violations (Bonta & Motiuk, 1990). Current validity research on the LSI-R has also supported the tool as a promising predictor of future offending (Andrews & Bonta, 1995; Goggin, Gendreau, & Gray, 1998). Moreover, empirical analyses reveal that the instrument's accuracy in predicting future offending holds across correctional settings and offender populations (Holsinger, Lowenkamp, & Latessa, 2004; Lowenkamp, Holsinger, & Latessa, 2001).

There are, however, warranted concerns about the population-specific nature of prediction tools. For instance, Wright, Clear, & Dickson (1986) tested the predictive validity of the Wisconsin model risk assessment on samples of probationers in New York and Ohio. Though the Wisconsin model had demonstrated predictive validity for the sample upon which it was created (Baird, Hines, & Bemus, 1979), it failed to demonstrate predictive validity for either the New York or Ohio sample of probationers (Wright et al.). Specific to the LSI-R, research conducted by Dowdy, Lacy, & Unnithan (2001) failed to support the tool as a predictor of halfway house outcome, two-year recidivism for any crime, or two-year felony recidivism for a sample of halfway house offenders. Taken together, these two findings serve as a reminder to correctional agencies that classification systems must be validated to their specific offender populations.
The literature provided on the validity of the LSI-R has established the tool as a valid predictor of correctional outcome across offender types and settings. The information obtained from the LSI-R can increase the accuracy of important corrections-related decisions (i.e., classification, risk level, criminogenic needs, service provision, intensity of interventions, and program effectiveness). In a similar vein, assessment research has also indicated a lack of universal applicability for prediction instruments. Federal probationers represent a unique correctional population in that they are older, more likely to be Hispanic, and more likely to be drug offenders than their state-supervised counterparts (Glaze & Palla, 2005; United States Probation and Pretrial Services, 2005). Because of this, and because of a current lack of existing research on the validity of the LSI-R for federal probationers, this research investigated the predictive accuracy of the LSI-R for a sample of federal probationers.

Method

Participants

The sample in this study was comprised of 2,107 adult federal probationers. To be eligible for inclusion, a federal probationer had to have been assessed with the LSI-R by a federal probation staff member trained in the administration and scoring of the tool. LSI-R assessment scores for the sample were completed over a two-year period between December of 2001 and December of 2003.

Procedures

In 2001, the southwestern federal probation district that provided these data received a three-day training on the implementation and scoring of the LSI-R. Six months later, follow-up LSI-R training was provided for all staff and immediately followed by a “train the trainers” session for staff that had demonstrated exceptional LSI-R scoring skills. During the follow-up and “train trainers” sessions, administrative staff voiced concern about the tool’s applicability to federal probationers and expressed interest in norming and validating the LSI-R on their offender population. This early discourse between federal probation staff and research consultants about the LSI-R’s psychometric properties served as the impetus for what later matured into a collaborative effort between both parties to provide the agency with aggregated probationer needs reports, normative information for their offender population, and evidence attesting to the LSI-R’s validity for federal probationers.

Save for outcome, the variables of interest in this study were entered into an electronic database maintained by federal probation staff. Once the number of offenders in the database exceeded 2,000, federal probation staff sent a copy to the authors (via electronic mail). Upon receipt, the data were cleaned and then used to generate a data collection sheet that individually listed all sample participants by their name, age, sex, ethnicity, race, and county of committing offense. These data collection sheets were then used to collect outcome data for the sample.

Measures

Although the LSI-R is comprised of ten risk and criminogenic need areas, only the composite LSI-R score was used in the current research. The LSI-R scores used are the result of offender interviews and collateral reviews of file and other offender information as completed by a federal probation staff member. Recidivism data were collected by completing follow-up record checks for each offender in the Federal Bureau of Prisons’ inmate locator database. The measure of recidivism used in the current study was incarceration in the Federal Bureau of Prisons for either a technical violation or new offense that occurred subsequent to the initial LSI-R assessment. Recidivism was coded dichotomously, where a value of 1 indicated the occurrence of subsequent incarceration and a value of 0 indicated that subsequent incarceration had not occurred.

Several demographic variables were also used in the current analyses to both describe the sample of offenders studied and further specify the relationship between the LSI-R and incarceration
through their consideration in a multivariate analysis. The additional variables included in the multivariate analysis were age, sex, and ethnicity. While age was a continuous variable, sex and ethnicity were both coded dichotomously, so that a value of 0 represented the most typical case (male and Hispanic, respectively) and a value of 1 represented a departure from the most typical case (female and non-Hispanic, respectively).

It should be noted that the Federal Bureau of Prisons’ database used to collect recidivism data only allowed for the determination of incarceration in a federal institution subsequent to the initial LSI-R assessment. Alternate measures of outcome, such as commission of a technical violation, time until new commitment, or incarceration under state jurisdiction were unavailable. Though the use of additional outcome measures would admittedly yield more information about the length of time until and type of recidivism, the use of subsequent incarceration is advantageous to the extent that it provides a more conservative test of the LSI-R’s predictive validity.

Results

Descriptives

Descriptive statistics for offender demographics, LSI-R scores and incarceration are presented in Table 1. An examination of these data reveals that the typical federal probationer in this sample was a Hispanic male, of 37 years, classified as low/moderate risk on the LSI-R ($M = 14.08, SD = 7.81$). The descriptive results also revealed a 26.1 percent base rate of incarceration for the sample, indicating that nearly three out of four federal probationers had not recidivated prior to the completion of follow-up record checks for the sample.

Validation

The predictive validity of the LSI-R for federal probationers was examined by conducting three separate analyses. The first of these involved calculating a predictive validity estimate of the relationship between the LSI-R and incarceration for the sample of 2,107 federal probationers. This analysis supported the LSI-R as a significant predictor of subsequent incarceration ($r = .283, p < .01$).

The second test in this research investigated the validity of the LSI-R and incarceration using receiver operating characteristic (ROC) analysis. The correlation coefficient calculated in the first analysis of this research represents the validity estimate most commonly reported in existing LSI-R research endeavors. However, because the magnitude of a correlation coefficient is dependent on the percentage of the sample identified as a recidivist by the risk tool (selection ratio) and percentage of actual recidivists in that sample (base rate), it can be said that existing research has yet to yield a statistic that would permit an unbiased comparison of the LSI-R’s predictive strength across samples and as compared to other prediction tools. To address this deficiency, ROC analysis was chosen as a second means of examining the LSI-R’s predictive validity in this study. The ROC method produces a statistic that, because it is unaffected by sample-specific base rates and selection ratios, is remarkably useful in comparing the utility and strength of a prediction instrument across different samples and vis-à-vis other prediction tools (Mossman, 1994a).

Statistics derived from ROC curves represent the ratio of true positives to false positives present in a given sample (i.e., as identified from predictions made by the risk instrument and observations for the outcome variable). The ROC analysis completed in this research produced an area equal to .689 ($p < .01$) to describe the relationship between the LSI-R and subsequent incarceration. Simply put, there was a 68.9 percent chance that a randomly selected recidivist had a higher score on the LSI-R than did a randomly selected nonrecidivist (Rice & Harris, 1995).

The analyses performed thus far have revealed that the LSI-R was a valid and robust predictor of
subsequent incarceration for federal probationers. To further specify this relationship, the final analysis in this study estimated a multivariate logistic regression model that examined the relationship between the LSI-R and incarceration while simultaneously considering the effects of age, sex, and ethnicity. The results of this multivariate analysis are reported in Table 2. An examination of the logistic regression model reveals that the LSI-R continued to be a significant predictor of incarceration, even when offender age, sex, and ethnicity were controlled. Moreover, a review of the values for Exponent(B) shows that the LSI-R was the strongest predictor of incarceration among the variables included in the multivariate model. This conclusion is also reached through an examination of the R values estimated for each of the predictor variables. The R value of .250 for the LSI-R is more than twice the R values estimated for the other significant predictors of incarceration included in the model (age and sex). Finally, the nature of the relationship observed between incarceration and the control variables of age and sex is also worth noting. The logistic regression analyses revealed a significant and negative R value for each variable, indicating female federal probationers and older federal probationers were less likely to recidivate than were their male and younger counterparts.

Discussion

This research sought to make two important contributions to the knowledge base of recidivism prediction. First, because the utilization of valid and efficient risk/needs assessment tools implies a certain level of improvement in the ability to manage offender caseloads and classify correctional populations, their use is becoming increasingly prevalent throughout North America. As the popularity of such prediction tools increases, so too, does the diversity of the offender population to which the tools will be applied. Consequently, it was deemed important to examine the LSI-R’s efficacy to predict long-term outcome for a sample of federal probationers, a correctional population previously overlooked in existing LSI-R validation studies. Second, this study used ROC methods to calculate an unbiased measure of predictive accuracy appropriate for comparisons across samples and between different tools, a measure absent in existing research, but critical for knowledge advancement.

Results from the predictive validity analyses were encouraging and provided evidence that the LSI-R was a valid and robust predictor of subsequent incarceration for this sample of federal probationers. Additionally, the multivariate analysis conducted in this research found that the LSI-R remained a valid predictor of subsequent incarceration when the effects of age, sex, and ethnicity were controlled. Taken together, these results make a strong case for the generalizability of the LSI-R to diverse offender populations. The findings of these, as well as previous, analyses further demonstrated that a theoretically informed and empirically refined actuarial measure of client risk and criminogenic need has much to offer correctional practice. When these findings are considered in the context of existing research, use of the LSI-R to inform correctional decisions of supervision and service provision appears to epitomize evidence-based practices.

In addition to further establishing the predictive validity of the LSI-R, these analyses contributed to existing research by using ROC methods to calculate an index of predictive accuracy that is independent of sample base rates and selection ratios. ROC methods are important in prediction research to facilitate comparisons of predictive strength across samples and, more importantly, across prediction instruments. The ROC area of .689 generated in this research was moderate to large in magnitude (Rice & Harris, 1995), and indicated that a randomly selected recidivist would have a higher LSI-R score than would a randomly selected nonrecidivist 68.9 percent of the time. Beyond the importance of this finding for the accuracy of the LSI-R, it is hoped that the ROC analyses reported here prompt future prediction efforts (on all risk/need assessment tools) to consider the inherent link between comparisons of predictive accuracy and knowledge advancement.

Optimistic conclusions aside, there were several limitations present in this research. First, the outcome measure used did not permit analyses with respect to length of time until failure, failure
due to technical violation, type of crime committed at failure, or failure under state jurisdiction. Additionally, these analyses examined the predictive accuracy of the LSI-R for a large sample of federal probationers, but did not investigate the tool’s applicability to offender subgroups. Certainly, existing research can benefit from future efforts that might further specify the relationship between the LSI-R and outcome across offender sex, ethnicity, and race.

References
### Table 1: Descriptives for the Sample (N = 2,107)

<table>
<thead>
<tr>
<th>Variable</th>
<th>M</th>
<th>SD</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age</td>
<td>37.27</td>
<td>11.47</td>
</tr>
<tr>
<td>LSI-r Score</td>
<td>14.08</td>
<td>7.81</td>
</tr>
<tr>
<td><strong>N</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>%</strong></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Sex</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Male</td>
<td>1,551</td>
<td>73.6</td>
</tr>
<tr>
<td>Female</td>
<td>556</td>
<td>26.4</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Ethnicity</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Hispanic</td>
<td>1,369</td>
<td>65.0</td>
</tr>
<tr>
<td>non-Hispanic</td>
<td>738</td>
<td>35.0</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Recidivism</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>No Reincarceration</td>
<td>1,557</td>
<td>73.9</td>
</tr>
<tr>
<td>Rincarceration</td>
<td>550</td>
<td>26.1</td>
</tr>
</tbody>
</table>
### Table 2: Logistic Regression Model Predicting Incarceration for the Sample (N = 2,107)

<table>
<thead>
<tr>
<th>Variable</th>
<th>B</th>
<th>SE</th>
<th>Wald</th>
<th>DF</th>
<th>Sig</th>
<th>R</th>
<th>Exp(B)</th>
<th>Lower</th>
<th>Upper</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age</td>
<td>-0.030</td>
<td>.005</td>
<td>34.8265</td>
<td>1</td>
<td>.000</td>
<td>-.117</td>
<td>.970</td>
<td>.961</td>
<td>.980</td>
</tr>
<tr>
<td>Sex</td>
<td>-0.815</td>
<td>.134</td>
<td>37.0412</td>
<td>1</td>
<td>.000</td>
<td>-.121</td>
<td>.443</td>
<td>.341</td>
<td>.575</td>
</tr>
<tr>
<td>Ethnicity</td>
<td>-0.036</td>
<td>.112</td>
<td>.1040</td>
<td>1</td>
<td>.747</td>
<td>.000</td>
<td>.964</td>
<td>.774</td>
<td>1.202</td>
</tr>
<tr>
<td>LSI-R</td>
<td>0.085</td>
<td>.007</td>
<td>152.2118</td>
<td>1</td>
<td>.000</td>
<td>.250</td>
<td>1.088</td>
<td>1.074</td>
<td>1.103</td>
</tr>
<tr>
<td>Consultant</td>
<td>-1.034</td>
<td>.211</td>
<td>23.9882</td>
<td>1</td>
<td>.000</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

95% CI for Exp(B)

Note: -2 log likelihood = 2175.887; $\chi^2$(4) = 236.335; p < .001; pseudo R2 = .156.


**Validating the Level of Service Inventory—Revised on a Sample of Federal Probationers**


**Validation of the Risk and Resiliency Assessment Tool for Juveniles in the Los Angeles County probation System**


