

Predicting outcome with the Level of Service Inventory-Revised: The importance of implementation integrity

Anthony W. Flores^{a,*}, Christopher T. Lowenkamp^b,
Alexander M. Holsinger^c, Edward J. Latessa^b

^a Department of Criminal Justice, California State University, Bakersfield, 24 DDH, 9001 Stockdale Highway, Bakersfield, CA 93311–1099, United States

^b Division of Criminal Justice, University of Cincinnati, P. O. Box 210389, Cincinnati, OH 45221–0389, United States

^c Department of Sociology/Criminal Justice and Criminology, University of Missouri, Kansas City, 208 Haag Hall, 5100 Rockhill Hall, Kansas City, MO 64110–2499, United States

Abstract

Correctional agencies face increasing pressure to provide more services for increasing caseloads with fewer resources. The Level of Service Inventory-Revised (LSI-R) is a risk/need assessment instrument that was designed to assist correctional agencies in classifying offenders based upon risk of re-offending, thereby allowing agencies to assign appropriate levels of risk/need and develop intervention/case-plans accordingly. Although predictive validity of the LSI-R has been demonstrated, very little attention has been paid to the effect that staff training and agency experience have on the validity of this comprehensive, dynamic risk/need assessment tool. The use of formalized training and agency experience were found to be important factors that may determine the validity of the risk/need scores that are gleaned from the LSI-R assessment process.

© 2006 Elsevier Ltd. All rights reserved.

Introduction

Although not a new practice, offender assessment and classification has become increasingly important to effective correctional intervention over the last two decades. The United States in particular has experienced unprecedented increases in correctional populations overall and institutional populations in particular (Harrison & Beck, 2004). In addition, despite legislative perceptions of a punitive voting public, considerable public support for the use of rehabilitative interventions also exists (Applegate, Cullen, & Fisher, 1997). The systemic manifestation of such contradictory expecta-

tions—general correctional crowding and the need to provide therapeutic interventions—can be relieved by the use of an effective offender assessment and classification model.

Offender assessment and classification involves developing a typology of offenders (i.e., high-risk; moderate-risk; low-risk). Classification processes resulting in client typologies benefit correctional processes by promoting better informed correctional practices, encompassing a variety of offender management decisions—from housing, to the assignment of security and/or supervision levels, to the determination of intervention strategies offered to the offender. Offender assessment and classification processes are not new correctional phenomena, although most modern-day methods predict future criminal behavior for purposes of informing systemic security and/or intervention decisions.

* Corresponding author. Tel.: +1 661 665 6794; fax: +1 661 665 6955.
E-mail address: aflores14@csub.edu (A.W. Flores).

While a number of offender assessment systems are currently available, they are not equal in terms of function or utility. Typically (for most modern assessment methods) the predictive validity of a particular assessment process is deemed to be of paramount concern. Due to the salience of public safety demands, an effective assessment model must accurately differentiate between high-risk offenders (those who possess substantial amounts of criminogenic or crime-producing needs) and low-risk offenders (those with relatively few criminogenic needs). Further, the incorporation of the risk principle of offender classification dictates that higher risk individuals warrant the majority of correctional attention, including the most intensive levels of both rehabilitative service and supervision (Andrews & Bonta, 2003). Conversely, and arguably as important, is the need to leave lower risk individuals free from intense levels of intervention to avoid interference with the protective factors that are likely present in their environment and within themselves (Hoge, Andrews, & Leschied, 1996).

Actuarial assessment

Bonta (2000) has emphasized the superiority of actuarial (as compared to clinical) offender assessment methods. Actuarial assessment methods base objective risk predictions in the observed outcome behavior of those previously assessed, whereas clinical methods rely on subjective decisions informed by the education, training, and experience of the decisionmaker. Put simply, offender assessment models derived from theoretically based and empirically valid criminogenic predictors that are measured via a standardized, objective, and dynamic instrument are far more accurate than are unstructured clinical assessments of the same factors (Grove, Zald, Lebow, Snitz, & Nelson, 2000). In addition to improved predictive accuracy, actuarial assessment methods also facilitate the creation of offender case plans grounded in empirical evidence (as opposed to clinical intuition). Successful case planning serves to guide an offender's trajectory through correctional service and monitor changes in risk and needs over time. Actuarial assessment, then, can be viewed as a veritable cornerstone for the provision of correctional services.

One actuarial criminogenic risk/need assessment tool that has been increasingly used in correctional environments is the *Level of Service Inventory-Revised (LSI-R)* (Andrews & Bonta, 1995). The LSI-R is a fifty-four item additive scale that covers ten criminogenic domains (criminal history, education/employment, financial, familial relationships, accommodations, leisure and recre-

ation, companions, alcohol and drug use, emotional health, and attitudes/orientations). Due to its additive nature, higher ranges of scores on the LSI-R are associated with a higher propensity to commit future criminal behavior. The majority of the fifty-four items are measured in a dynamic fashion, meaning they are assessed using the most recent applicable information in an attempt to capture the current state of a particular item or domain. The LSI-R is conducted using a structured interview between the correctional professional and the offender. In addition, collateral information (via police, court, probation, prison, or other files) and information from any correctional professionals having prior knowledge of the offender should also be used when available.

The LSI-R, reliability, and predictive validity

Demonstrated reliability and predictive validity of the LSI-R can be found in a considerable body of research (Andrews, 1982; Andrews & Bonta, 1995; Andrews & Robinson, 1984; Bonta & Andrews, 1993). The LSI-R lends itself well to correctional practices in large part due to its claim of promoting standardized and objective decisions. The achievement of standardized and objective decision making requires that different assessors actually produce similar LSI-R ratings when assessing the same offender. Research endeavors examining the reliability of the LSI-R have shown just this. Specifically, initial research on the LSI supported both intra- and inter-rater reliability (Andrews, 1982). In this research, raters produced very similar LSI decisions across different time intervals just as multiple assessors came to similar LSI informed decisions for the same case. Further, Andrews reported that the total difference between LSI scores from one trained assessor to the next was consistently less than five points.

Initial validation studies completed on the LSI supported the tool's ability to predict outcomes for probationers. In this early research on the LSI, Andrews (1982) found that the LSI predicted supervision success ($r=.35$), in-program outcome status ($r=.47$), and in-program recidivism ($r=.38$). Andrews and Robinson (1984) examined this same initial validation sample over a longer follow-up period and found a strong correlation between the LSI and recidivism ($r=.43$), incarceration ($r=.37$), and self-reported, undetected offenses ($r=.22$). In the *LSI-R User's Manual*, Andrews and Bonta (1995) reviewed the outcomes of several validation studies involving the LSI-R used with a variety of adult correctional populations (including those mentioned above). From this review, Andrews and Bonta concluded that the LSI-R was a valid predictor of

correctional adjustment and outcome for different types of correctional programs (community supervision, residential settings, and incarceration) and for different subgroups within the correctional populations by examining validity within groupings of ethnicity, sex, and age.

In a related and more recent evaluation of the predictive ability of the LSI-R, Lowenkamp, Holsinger, and Latessa (2001) found that the instrument not only accurately predicted recidivism for male offenders ($r=.22$), but also predicted recidivism equally well or better for female offenders ($r=.37$). There is also evidence to support the LSI-R as one of, if not the most, promising risk/need assessment tools currently available. In a meta-analytic examination of the utility and explanatory power of three different offender risk and need assessments, Gendreau, Little, and Goggin (1996) concluded that the LSI-R was the most powerful prediction instrument to date. Similarly, a recent meta-analysis (Gendreau, Goggin, & Smith, 2002) tested the predictive validity of the LSI-R and the Psychopathy Checklist-Revised (PCL-R) for general and violent recidivism. While the analysis found that both instruments predicted recidivism at acceptable levels, the LSI-R outperformed the PCL-R in predicting both types of behavior.

Based on these research efforts, supporters of the LSI-R state that the instrument is indeed a valid risk prediction tool for offenders, as well as a valid needs assessment tool that provides information relevant to standardized and objective programmatic decision making. Although an impressive body of research exists that supports the LSI-R as one of the most promising tools currently available to facilitate consistent and accurate correctional decisions, empirical investigations have yet to study the impact that implementation integrity and agency commitment might have on the tool's efficacy.

Implementation integrity and agency commitment

It is intuitive that the validity of any classification instrument will break down if the instrument is not properly implemented. Nonetheless, statistical tests of the predictive validity of risk and needs assessment instruments are generally based on the assumption that the tools have indeed been properly implemented. In actuality, it is quite possible that such instruments are applied by untrained personnel or otherwise used in ways for which they were not intended.

As noted above, the LSI-R gathers information largely through a one-on-one interview between the

corrections professional and offender. Although the review of collateral information is a potentially integral part of the LSI-R process, the quality, accuracy, and scope of the information used to complete the assessment ultimately rests with the assessor. The comprehensive approach taken by the LSI-R to measure risk and needs, coupled with the potential for assessor subjectivity in their measure of risk and need items generates a legitimate concern that staff training on the scoring and implementation of the LSI-R may be a key determinant to the tool's success. Similarly, the amount of experience a correctional agency has with the LSI-R assessment process may also have a significant impact on the tool's accuracy, particularly through the maturation of staff competence in collecting quality information, honing of interviewing skills, and development of proper scoring techniques.

Many of the issues inherent to implementation integrity may be resolved through appropriate staff training. Research exploring the effect of staff training on outcome does exist, albeit limited. One example of research investigating the effect that staff training can have on correctional processes found that the use of a classification system (the LSI-R in particular) resulted in a significant reduction in prediction errors as a result of appropriate staff training (Bonta, Bogue, Crowley, & Motiuk, 2001). Beyond espousing existing research on the predictive validity of the LSI-R, the importance of examining implementation integrity, as measured through staff training and length of use, has emerged as a possible key determinant of the LSI-R's ability to predict outcome.

The current research addressed the two issues outlined above. First, the predictive validity of the LSI-R was studied using a sample of offenders drawn from several midwestern residential correctional facilities. Second, the importance of implementation integrity to predictive efficacy was examined by assessing the possibility of differential relationships between the LSI-R and outcome as a function of formal staff training on the LSI-R and/or agency experience using the tool.

Method

Participants

The subjects in this study were adult offenders from several residential correctional facilities (RCFs) located in a large midwestern state. The RCFs in this study provided services and sanctions to offenders convicted of felony offenses and placed under probationary supervision in lieu of incarceration. To be eligible for inclusion

in the study, an offender had to have been assessed using the LSI-R by one of the RCFs.

Procedures

A mailing was sent to each of the RCFs requesting lists of offenders that had been assessed with the LSI-R and subsequently released from the RCF prior to the conclusion of fiscal year 1999. From that mailing, twelve out of a total of eighteen RCF agencies replied with the requested information. A computer-generated random sample was drawn from this list that was then sent back to the RCF agencies along with data collection instruments and instructions on how to use the instruments. From this second mailing, seven facilities provided the data requested and two more agencies provided electronic files containing the information required. Thus, the total number of participants in this study was 2,030 offenders from a total of nine RCF facilities. LSI-R assessments were completed for each offender upon admission to the RCF through personal interviews and the review of collateral information.

Although the LSI-R is comprised of ten subsections, only the composite LSI-R score was used in the analyses. Two measures of implementation integrity were employed. Formal training was a dichotomous measure indicating whether or not RCF staff underwent formal LSI-R training. Formal training was defined as training by a qualified LSI-R trainer who was not employed by the RCF or training by a RCF staff person that had successfully completed a train the trainers curriculum offered by a master trainer of the LSI-R. Initial LSI-R training consists of agency staff attending a two-to-three day session that includes the theoretical underpinnings of classification, the recognized best practices in offender assessment, scoring reliability exercises, and the use of the LSI-R in case planning. The second measure of implementation integrity was also a dichotomous measure that indicated whether or not the RCF had used the LSI-R for three years or more.

The measure of recidivism utilized for the current study was future incarceration. Recidivism outcome data were collected using a prisoner database maintained by the State Department of Corrections. Checks were made for subsequent incarcerations in a state facility. As mentioned above, study participants had to have been assessed with the LSI-R and released by the RCF prior to the conclusion of fiscal year 1999. As a result, the length of follow-up varied as a function of RCF release date. The follow-up period ranged between a minimum of one year and a maximum of slightly more than three years. 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.

The database used for offender background checks was limited in that it only allowed for the determination of subsequent incarceration. Other measures of correctional outcome, such as arrest, technical violation, or type of offense, were unavailable. Ideally, these additional measures of recidivism would have been included in the analyses although an important advantage to using incarceration as an outcome measure is the higher degree of confidence that subsequent criminal or antisocial behavior actually occurred. Outcome measures such as arrest and technical violations are coded irrespective of an actual finding of criminal guilt and, as such, may not provide the most reliable indicator that criminal behavior has indeed occurred.

Results

The results for all analyses are reported in two sections. The first section presents descriptive statistics for offender demographics, composite LSI-R score, and the outcome measure. The second section presents measures of association between the composite LSI-R score and future incarceration. Correlations are presented for the overall sample, and then disaggregated across the measures of implementation integrity to examine the importance of staff training and agency experience to the predictive validity of the LSI-R.

Demographic characteristics

Descriptive statistics for offender demographics, LSI-R scores and the outcome measure are presented

Table 1
Sample descriptive statistics and outcome measure (N=2,030)

Variable	M	SD
Composite LSI-R score	M=27.28	6.78
Offender age	M=29.00	7.67
	N	%
<i>Offender sex</i>		
Male	1,654	81.5%
Female	376	18.5%
<i>Offender race</i>		
White	1,132	55.8%
Non-White	898	44.2%
<i>New incarceration</i>		
Yes	1,292	63.6%
No	738	36.4%

in Table 1. In short, the typical offender in this sample can be described as a White male, of twenty-nine years, classified as moderate risk. The measure of outcome revealed that 63.6 percent of the sample experienced a new incarceration during the follow-up period.

Table 2 presents analyses investigating the relationship between the LSI-R composite score and outcome. While the overall correlation between composite LSI-R score and reincarceration is presented, the correlation coefficients are also disaggregated by whether or not the instrument was being used in a trained (versus untrained) environment, the length of experience the agency had with the tool, and for male and female offenders. In addition, the 95 percent confidence intervals are presented for each correlation coefficient.

Overall, a positive and statistically significant ($p < .01$) relationship was observed between the composite LSI-R score and reincarceration. This relationship was in the expected direction, and had a value of .18 (95 percent C.I. = .14 to .22). In short, the likelihood of recidivism (measured as new incarceration in a state facility) was greater for those assessed as higher risk on the LSI-R.

A significant correlation of .21 ($p < .01$) was revealed when examining the relationship between composite LSI-R score and future recidivism for those who were trained in the use of the tool. This correlation was also in the expected (positive) direction and was statistically significant (95 percent C.I. = .16 to .26). A much smaller and insignificant correlation was revealed when examining the composite LSI-R scores generated by untrained correctional professionals. Specifically, a positive but relatively weak correlation of .08 was revealed. While the

size of this correlation is problematic when considering the prediction of future recidivism, it is also important to point out that the 95 percent confidence intervals ranged from $-.02$ to $.18$ (which includes zero). This finding translates into a low probability that untrained professionals utilizing the LSI-R will have any benefit regarding the prediction of future criminal behavior. Further, this finding questions the logic of basing case plans and classification decisions on the information obtained from LSI-R assessments scored by untrained professionals. Additionally, the correlation coefficient for those who were untrained in the use of the tool was not statistically significant. The z-test for differences between correlation coefficients revealed significant differences between the correlations for untrained and trained professionals.

When considering the effect of experience using the LSI-R, differences in the strength of the relationship between the composite score and future incarceration appeared. When the tool had been in place for less than three years, a correlation of .14 between the composite score and future incarceration was revealed (95 percent C.I. = .08 to .20). When the tool had been in place for three or more years, a correlation of .25 emerged (95 percent C.I. = .19 to .30). Both of the correlation coefficients were statistically significant ($p < .01$). While the composite LSI-R score may help predict future recidivism in both situations (less than three years' experience versus three or more years' experience), the relationship becomes markedly larger over time. This is likely due to agency experience with the tool, and the extent to which the tool becomes a part of the correctional landscape, thereby becoming a larger part of the agency's decision-making protocol. The z-test of differences between correlation coefficients also revealed significant differences in the tool's predictive validity across agency length of use.

In order to further compare the differences between trained and untrained professionals using the LSI-R, and those that had less than three years experience with the tool to those that had three or more, the common language (CL) effect size indicator (McGraw & Wong, 1992) was used. The CL effect size indicator revealed the relative utility of one measure of risk over another. The CL statistic converts an effect size into the probability that an estimated predictor sampled at random from the distribution of one risk measure will be greater than that sampled from the distribution of another (see Gendreau et al., 2002). The CL effect size indicator was .9171 when comparing trained to untrained professionals using the LSI-R, and .8245 when comparing those with three or more years of experience to those with less than three years' experience using the tool.

Table 2
Relationship between LSI-R and outcome, and implementation variables

Group	r	N	Lower 95 % C.L.	Upper 95 % C.L.
All	.18**	2,030	.14	.22
Trained ^d	.21**	1,635	.16	.26
Untrained	.08	395	-.02	.18
Less than three years ^b	.14**	921	.08	.20
More than three years	.25**	1,109	.19	.30
Male offenders ^c	.17**	1,654	.12	.22
Female offenders	.25**	376	.15	.34

^a The z-test for the difference between correlations predicting reincarceration for the trained and untrained agencies is 2.3646, $p = 0.0090$.

^b The z-test for the difference between correlations predicting reincarceration for the agencies using the LSI-R for more than three years and less than three years is 2.5642, $p = 0.0052$.

^c The z-test for the difference between correlations predicting reincarceration for the males and females is 1.3858, $p = 0.0829$.

* $p < .05$.

** $p < .01$.

Consistent with other research (Kirkpatrick, 1999; Lowenkamp et al., 2001; Rettinger, 1998), the composite LSI-R score correlated significantly with recidivism for both male and female offenders (.17 and .25 respectively), though somewhat stronger for female offenders.

Discussion

Offender classification is without question an essential component of correctional practices. In an effort to make necessary changes as quickly as possible, correctional agencies all too often implement strategies without fully supporting them, at times due mainly to resource and administrative constraints. The current research investigated the predictive ability of the LSI-R for a sample of midwestern residential correctional agencies. In addition, the importance of implementation integrity was also investigated by examining the importance of staff training and experience with the instrument.

The initial findings provided some support for the predictive validity of the LSI-R. When the relationship between LSI-R score and future incarceration was examined in the aggregate (for all nine RCFs simultaneously), the results indicated a significant but relatively weak relationship. Yet, when the relationship between the LSI-R and outcome was examined separately for agencies that underwent formal LSI-R training, the magnitude of the relationship was similar to previous validation studies. By contrast, the validity of the LSI-R appeared to suffer when the tool was implemented by staff untrained in its use. Furthermore, the relationship between LSI-R and outcome was much stronger for agencies that had been using the LSI-R for more than three years prior to when the sample was drawn. This likely indicates that some time is necessary for an agency to adequately train staff and ensure that the assessment is being conducted properly.

Correctional agencies implementing standardized classification tools in general, and the LSI-R specifically, will do well to make a long-term commitment to the assessment process and utilize formal staff training. These findings support the predictive validity of the LSI-R for agencies attending to strategies reflective of implementation integrity. Selecting an assessment instrument that accurately identifies offenders most likely to re-offend is an important part of successful correctional intervention. Faced with limited resources and ever-expanding clientele, correctional agencies have much to gain from valid classification instruments. Correctional agencies must use the information obtained from valid classification systems to guide the allocation of what are often scarce resources. Most notably, correctional agencies should use the

information obtained from risk and need assessment tools to identify the aggregate programming needs of their clients, and accordingly, invest in those therapeutic interventions that most appropriately target the criminogenic needs of their client population. Offenders assessed as moderate and high-risk must become the primary recipients of service while offenders assessed as low-risk must be weeded out of intensive correctional service so as to not waste correctional dollars. Correctional agencies that implement assessment and classification practices to facilitate decisions regarding offender risk and need are better equipped to appropriately service clients, save time and money, and ultimately achieve public safety.

These results demonstrate the importance of implementation integrity (measured here as formal staff training and amount of time spent using the tool) to the predictive validity of risk and need classification tools. The utilization of risk and need assessment tools in name only, irrespective of implementation integrity, results in compromised validity. Basing correctional decisions on assessments rendering inaccurate predictions hampers efforts toward effective correctional intervention, and ultimately, jeopardizes public safety.

References

- Andrews, D. A. (1982). *The Level of Supervision Inventory (LSI): The first follow-up*. Toronto, Ontario, Canada: Ontario Ministry of Correctional Services.
- Andrews, D. A., & Bonta, J. (1995). *LSI-R: The Level of Service Inventory-Revised*. Toronto, Ontario, Canada: Multi-Health Systems, Inc.
- Andrews, D. A., & Bonta, J. (2003). *The psychology of criminal conduct*. Cincinnati, OH: Anderson.
- Andrews, D. A., & Robinson, D. (1984). *The Level of Supervision Inventory: Second report*. Toronto, Ontario, Canada: Ontario Ministry of Correctional Services.
- Applegate, B., Cullen, F. T., & Fisher, B. S. (1997). Public support for correctional treatment: The continuing appeal of the rehabilitative ideal. *Prison Journal*, 77, 237–258.
- Bonta, J. (2000, August). Offender assessment: General issues and considerations. *Forum on Corrections Research*, 14–18.
- Bonta, J., & Andrews, D. A. (1993). The Level of Supervision Inventory: An overview. *IARCA Journal*, 5(4), 6–8.
- Bonta, J., Bogue, B., Crowley, M., & Motiuk, L. (2001). Implementing offender classification systems: Lessons learned. In G. A. Bernfeld, D. P. Farrington A.W. Leschied (Eds.), *Offender rehabilitation in practice: Implementing and evaluating effective program* (pp. 227–245). Chichester, NY: J. Wiley and Sons.
- Gendreau, P., Goggin, C., & Smith, P. (2002). Is the PCL-R really the “unparalleled” measure of offender risk? A lesson in knowledge cumulation. *Criminal Justice and Behavior*, 29(4), 397–426.
- Gendreau, P., Little, T., & Goggin, C. (1996). A meta-analysis of the predictors of adult offender recidivism: What works! *Criminology*, 34, 575–607.
- Grove, W. M., Zald, D. H., Lebow, B. S., Snitz, B. E., & Nelson, C. (2000). Clinical versus mechanical prediction: A meta-analysis. *Psychological Assessment*, 12(1), 19–30.

- Harrison, P. M., & Beck, A. J. (2004). *Prisoners in 2003 (Bureau of Justice Statistics Bulletin)*. Washington, DC: U.S. Department of Justice.
- Hoge, R. D., Andrews, D. A., & Leschied, A. W. (1996). An investigation of risk and protective factors in a sample of youthful offenders. *Journal of Child Psychology and Psychiatry*, 37(4), 419–424.
- Kirkpatrick, B. L. (1999). Exploratory research of female risk prediction and the LSI-R. *Corrections Compendium: The National Journal for Corrections*, 24, 1–17.
- Lowenkamp, C. T., Holsinger, A. M., & Latessa, E. J. (2001). Risk/need assessment, offender classification, and the role of child abuse. *Criminal Justice and Behavior*, 28(5), 543–563.
- McGraw, K. O., & Wong, S. P. (1992). A common language effect size statistic. *Psychological Bulletin*, 111, 361–365.
- Rettinger, L. J. (1998). *A recidivism follow-up study investigating risk and need within a sample of provincially sentenced women*. Unpublished doctoral dissertation, Carleton University, Ottawa, Canada.