

**RELIABILITY AND VALIDITY  
STUDY  
OF THE LSI-R  
RISK ASSESSMENT  
INSTRUMENT**

James Austin, Ph.D.  
Dana Coleman  
Johnette Peyton, M.S.  
Kelly Dedel Johnson, Ph.D.

*Final Report*

*Submitted to  
The Pennsylvania Board of Probation and Parole*

*January 9, 2003*

The Institute on Crime, Justice and Corrections  
at The George Washington University  
10 G Street, NE, Ste. 610 Washington, DC 20002  
(202) 408-6800 ph (202) 408-6818 fax

## **ACKNOWLEDGMENTS**

We wish to express our gratitude to the Pennsylvania Commission on Crime and Delinquency for sponsoring this project. In particular, our PCCD project manager, Lonisha Morgan, helped us to complete this project in a timely and thoughtful manner. This study would not have been possible without the dedication and hard work of the PBPP Research and Statistics Division. Special recognition is due to Jim Alibrio, Director of Research and Statistics, and his dedicated staff that made significant contributions to the design and data collection efforts required for this study.

This study was funded by the Pennsylvania Commission on Crime and Delinquency and is in response to DCSI Program Purpose Area #11. Points of view or opinions stated in this document are those of the authors and do not necessarily represent the official position or policies of the Pennsylvania Commission on Crime and Delinquency or the Pennsylvania Board of Probation and Parole.

## **EXECUTIVE SUMMARY**

The Commonwealth of Pennsylvania uses sentencing guidelines to help determine whether an offender will be incarcerated (disposition) and for how long (duration). Sentencing guidelines are used to determine a minimum and maximum sentence for offenders committed to county jails or state correctional institutions. State sentenced prisoners are eligible for parole consideration once they have reached their minimum date. Once an inmate becomes eligible for parole, the decision to release that inmate is guided by an assessment of the inmate's risk to public safety. Assessments can provide highly accurate predictions of how individuals with similar characteristics might behave in the future. These assessments are predicated on the assumption that individual criminal behavior can be predicted by assigning offenders to groups that have explicit re-offending probabilities.

The Pennsylvania Board of Probation and Parole (PBPP) selected the Level of Service Inventory-Revised (LSI-R) instrument as its risk classification tool because it introduces dynamic and more current factors into the risk assessment process, beyond the conventional use of static criminal history and demographic factors. The LSI was developed in the late 1970s in Canada through a collaboration of probation officers, correctional managers, practitioners and researchers. The LSI-R is comprised of 54 static and dynamic items across ten sub-scales (O'Keefe and Wensus, 2001). While the LSI-R has been researched extensively in other jurisdictions, its reliability and validity specifically for Pennsylvania's offender population had not yet been tested. Of particular interest is Pennsylvania's decision to use the LSI-R as a component of its parole decision making guidelines; heretofore, the LSI-R has been used to identify the appropriate level of supervision for probationers and parolees already residing in the community. In this study, the LSI-R's relevance and usefulness as a decision making tool as applied to an incarcerated population is a key line of inquiry.

The Pennsylvania Commission on Crime and Delinquency (PCCD) contracted The Institute on Crime, Justice and Corrections (ICJC) at The George Washington University to conduct a reliability and validation study using the LSI-R scores and recidivism data. The following report summarizes the ICJC's findings.

### **Summary of Research Design and Methodology**

This project consists of two segments: an assessment of the inter-rater reliability in scoring the LSI-R, and the validation of the LSI-R's statistical association with recidivism. The reliability assessment was conducted by selecting a sample of 120 prisoners who were scored on the LSI-R on two separate occasions by two independent PBPP institutional staff. The results of the initial reliability test showed that most of the LSI-R scoring items did not meet a sufficient level of reliability. Consequently, a second reliability test was made in September 2002 on another sample of 156 prisoners to determine if the reliability rates could be improved.

The validation assessment entailed examining recidivists (for the purposes of this study, arrests, detentions, absconders, and returns to prison are considered recidivists) of approximately 1,000 prisoners who were released from nine LSI-R test facilities in 2001. For each of these prisoners an LSI-R form was completed. The follow-up period was for 12 months, which allowed the researchers to determine which items were associated with recidivism within that period.

## **Reliability Findings**

As noted earlier there were there were two reliability assessments made. The results of the first test, in October 2000, can be summarized as follows:

- 18 of 54 LSI-R items (or 33 percent) had reliability scores at or above a minimum 80 percent threshold for inter-rater consistency.
- The items that measure the prisoner's criminal history and other factual -based items had the highest level of agreement.
- There was substantial disagreement between the two interviewers regarding the risk level (high, medium, and low). In total, the interviewers agree on the assessed risk level in 71 percent of the cases.
- For approximately 60 percent of the cases, the two total scores differ by more than three points. These patterns indicate that the divergent risk levels are not the result of differences in one or two items, but rather several items creating significantly different total risk scores.

The second reliability test completed in September 2002 repeated the reliability test with 150 cases. Additional training was provided to staff and the length of time between the first and second assessment was shortened considerably. The major findings of this second test were as follows:

- The scoring reliability had increased. More directly, the level of agreement in the overall risk level had increased from 71 to 88 percent.
- 34 out of 54 items (or 63 percent) now met the minimum threshold of 80 percent or above.
- As in the first test, the more factual criminal history questions had the highest reliability scores.

## **Validation Findings**

The 1,006 prisoner sample used for the validation study were not a representative sample of all prison releases. The pilot involved prisoners who were housed at nine DOC facilities who had an LSI-R completed on them, and were released on parole were included. The major findings associated with this phase of the study are as follows:

- Approximately half of the parolees recidivated within one year of their release date.
- The major reasons for recidivism were technical parole violations, followed by new convictions and absconding. The high percentage of recidivism for technical violations is consistent with statistical findings in many other states.
- Only eight of the 54 LSI-R items were found to be associated with recidivism. These items tend to be “static” measures that reflect prior criminal history and prior drug use. These same eight items also tend to have higher reliability scores.
- While the LSI-R does classify prisoners according to their recidivism rates, the lack of reliability among many of the LSI-R items creates a great deal of “noise” in the instrument and diminishes its level of validity.
- Another version of the LSI-R was considered. Specifically, the LSI-SV or Screening Version consists of eight items of which are several of the same items found to be associated with recidivism. With proper training and analysis, the LSI-SV can be substituted for the LSI-R.
- Using a combination of the eight most reliable and valid items from the LSI-R, plus several other demographic items resulted in the best predictive results.

## **Conclusions and Recommendations**

- The LSI-R instrument effectively separated risk on PA cases into 3 categories, even though these separations (levels of risk) are due to a limited number of questions on the LSI-R.
- The LSI-R as tested in PA institutional settings had problematic reliability. These results do not warrant its use by the PBPP as a method for assessing risk at the time of a parole interview. Instead a more succinct instrument such as the LSI-SV would be more effective for the Board’s risk assessment.
- There are a limited number of LSI-R items for which substantial reliability has been achieved and which have a statistical relationship to recidivism. These (and other) items can be used for assessing risk at the time of a parole interview using a more condensed scoring instrument.

- The LSI-R is best suited for use by the PA DOC upon admission to prison to identify service and program needs and by the PBPP once parole has been granted to identify the level of community supervision and services required.
- The LSI-SV version can be used as one component of the PBPP decision-making process assuming staff are properly trained and tested in the use of the instrument.
- Should the PBPP decide to use the LSI-SV for risk assessment, it should carefully monitor its use over a 12 month period to re-validate its predictive attributes.

## **I. INTRODUCTION**

### **A. Background**

The Commonwealth of Pennsylvania uses sentencing guidelines to help determine whether an offender will be incarcerated (disposition) and for how long (duration). Sentencing guidelines are used to determine a minimum and maximum sentence for offenders committed to county jails or state correctional institutions. State sentenced prisoners are eligible for parole consideration once they have reached their minimum date. Once an inmate becomes eligible for parole, the decision to release that inmate is guided by an assessment of the inmate's risk to public safety. Assessments can provide highly accurate predictions of how individuals with similar characteristics might behave in the future. These assessments are predicated on the assumption that individual criminal behavior can be predicted by assigning offenders to groups that have explicit re-offending probabilities.

Parole-eligible prisoners are reviewed by the Pennsylvania Board of Probation and Parole (PBPP), which decides whether the inmate is a suitable risk for release and what conditions of parole supervision should be imposed if release is granted. As part of the state's indeterminate sentencing structure, the PBPP has discretionary release powers, and uses parole guidelines to assist in the exercise of its discretion. These guidelines were first developed in the 1980s and have been revised on several occasions. Over time, the PBPP has undertaken steps to develop and implement a new set of guidelines that would better reflect its current policies regarding parole.

In 1997, the PBPP was selected to participate in a multi-state project on Structured Release funded by the National Institute of Corrections (NIC). The NIC project lasted two years and resulted in the creation of new guidelines which included a process for determining the inmate's risk to public safety. This risk assessment process relies upon the Level of Service Inventory-Revised (LSI-R). The PBPP endorsed using the LSI-R instrument as its risk classification tool because it introduced dynamic and more current factors into the process of risk assessment, beyond the conventional use of static criminal history and demographic factors. Further, the PBPP sought to have a standard risk classification instrument that could be used to assess risk for parole release as well as to determine the appropriate level of supervision after release.

The LSI-R consists of 54 items that are sorted into the following ten substantive areas believed to be related to future criminal behavior:

1. Criminal History (10 items)
2. Education and Employment (10 items)
3. Financial (2 items)
4. Family and Marital (4 items)
5. Accommodations (3 items)
6. Leisure and Recreation (2 items)
7. Companions (5 items)

8. Alcohol and Drugs (9 items)
9. Emotional and Personal (5 items)
10. Attitude and Orientation (4 items)

Through an interview process, offenders are rated on items requiring either a “yes/no” response, or the use of a structured scale ranging in value from 0 to 3. Based on these responses, the interviewer scores the offender on each item, totals the item scores, and determines the offender’s overall risk level using the PBPP scale.

In October 1999, the PBPP undertook an initial study to test the administration of the LSI-R, and to set cut-off points for the scale based on locally-relevant policies. The key finding of this effort was that the locally determined cut-off points created manageable workload levels. However, these cut-off points needed to be validated using recidivism data in order to assess the predictive validity of the calculated risk levels. Adjustments to the cut-off points of the scale may be prudent based on the outcome of this research.

## **B. Project Objectives**

This project had two major components: an assessment of the inter-rater reliability in scoring the LSI-R, and the validation of the LSI-R’s ability to assess prisoners according to their level of risk to recidivate.

Reliability refers to the level of consistency in scoring offenders on the LSI-R’s various factors that assess the inmate’s suitability for release to parole. There are two types of reliability, inter- and intra-rater reliability. The former involves whether two persons computing the LSI-R score on the same individual reach the same rating. Intra-rater reliability refers to whether a single rater scoring the LSI-R for an inmate will reach the same rating on repeated applications. This analysis assesses only the level of inter-rater reliability. The specific research questions for the reliability study are:

1. *To what extent is the scoring of the LSI-R reliable?*
  - a. *Are LSI-R risk scores computed in a consistent manner across PBPP staff?*
  - b. *Which LSI-R risk items appear to have the greatest inter-rater reliability?*
  - c. *Which should be deleted or redefined to ensure consistency in assessing an inmate’s risk for release?*

A validation study sought to determine the LSI-R’s statistical relationship to recidivism among prisoners released to parole. This analysis examined the recidivism rates across the major LSI-R risk levels, as well as the individual LSI-R scoring items. The specific research questions to be addressed are stated below:

2. *To what extent is the LSI-R a valid predictor of recidivism for prisoners*



*released to parole?*

- a. Which LSI-R risk factors are statistically associated with recidivism?*
- b. To what extent is the overall LSI-R risk levels associated with recidivism?*
- c. What adjustments can be made to the LSI-I or other versions of the LSI-R that might improve the risk assessment process?*

## **II. RELIABILITY ASSESSMENT OF LSI-R**

### **A. Methodology**

The reliability assessment was conducted by selecting a sample of approximately 120 prisoners who were screened using the LSI-R on two separate occasions by two independent PBPP institutional staff. The initial assessments were conducted in September and October of 2000. The same 120 prisoners were then re-assessed approximately two months later by a different PBPP staff person also trained in the use of the LSI-R. The sample was stratified by facility location (approximately 15 per facility) to ensure it included a wide array of PBPP staff participating in the exercise.

Once each case was scored, PBPP staff entered the data into a Microsoft Excel spreadsheet application. Of the total sample of 120 cases, only two cases were deleted due to the absence of a completed re-test survey, resulting in a final sample size of 118 prisoners. Once all cases were entered, the PBPP forwarded a copy of the data file to ICJC staff, who then converted the file into an SPSS database for statistical analysis. Using these data, descriptive analyses (e.g., frequencies and cross-tabulations) were conducted to produce tables that reflect the level of agreement on each item, the total score, and the assessed risk level. For the purposes of this reliability assessment, a “percent agreement” statistic was computed to reflect the extent to which the two interviewers were consistent in their assessment of each inmate on each of the items.

### **B. Results**

Table 1 summarizes the background attributes of the 120 prisoners included in the sample.<sup>1</sup> The sample is predominantly male (89%) and non-white (60%), with a median age of 36 years. A large number of prisoners (36%) are age 40 or older. Sixty-five percent of the cases are being considered for parole for the first time, 27 percent are being reviewed after a previous denial by the PBPP, and eight percent are being reviewed after a previous

---

<sup>1</sup>Demographic and offense characteristics of the population from which the sample was drawn were not available; therefore, a comparative analysis to ensure the representativeness of the sample could not be conducted.

Table 1 Demographic and Offender Characteristics of LSI-R Reliability Sample					
Characteristic	N=120	100%	Characteristic	N=120	100%
<b>Race</b>			<b>Gender</b>		
White	48	40.0	Male	107	89.2
Black	63	52.5	Female	13	10.8
Hispanic	9	7.5	<b>Present Offense</b>		
<b>Current Age</b>			<i>Violent</i>	59	49.2
Under 21	2	1.7	Murder/Manslaughter	9	7.5
21-29	31	25.8	Aggravated Assault	19	15.8
30-39	44	36.7	Rape/Sexual Assault	11	9.2
40+	43	35.8	Robbery	15	12.5
Average Age	37 years		Kidnapping	3	2.5
Median Age	36 years		Simple Assault	2	1.7
<b>Parole Status</b>			<i>Property</i>	32	26.7
First Review	78	65.0	Theft/Other Property	17	14.2
Previous Denial	32	26.7	Burglary	15	12.5
Previous Parole Failure	10	8.3	<i>Weapons</i>	3	2.5
			<i>Drugs</i>	19	15.8
			<i>Other</i>	7	5.8

failure on parole that returned them to prison. About half of the sample (49%) was convicted of a violent crime, with most of these crimes being robbery or aggravated assault. Among the non-violent crimes, the most prevalent offenses are drugs, theft and burglary.

Table 2 summarizes the overall risk level determined by each of the two interviewers, taking into account the scores on each of the 54 LSI-R scoring items. It is important to note that the vast majority of prisoners on both the first and second assessment were scored as high risk, although there is a certain amount of disagreement between the two interviewers. Specifically, the first interviews identified 73 percent of the prisoners as “high risk,” while the second interviews of the same prisoners reduced that figure to 65 percent. Similarly, the first interviewers scored 23 percent of the sample as “medium risk,” compared to 28 percent for the second wave. There were very few “low risk” cases in the sample, and little difference in the proportion of cases scored as low risk. With regard to risk level, it is also important to

note that the scale cut-off points were determined by the PBPP and not the developers of the LSI-R.

Table 3 presents a more direct comparison of the level and direction of disagreement between the two interviewers in terms of the *risk level*. The shaded, diagonal boxes in the table represent those cases on which the two interviewers agreed on the prisoners' risk level. In total, the assessors agreed in 71 percent of the total sample. However, the interviewers differed by only one risk level (e.g., one scored the inmate as "low" and the other scored the inmate as "medium") rather than two risk levels (e.g., "low" versus "high"). In some cases, an inmate's total score differed between the two raters, but translated into the same level of risk. In these instances, the assessed risk level was considered to be consistent.

To better understand the source of disagreement between the two scored risk levels, an item-by-item analysis was conducted and is shown in Table 4. The level of reliability across items varies significantly, ranging from 53 percent to 96 percent. Items that proved to be highly reliable included criminal history items based on factual information readily available in the inmate's record, such as the number of prior convictions, or whether the inmate had been incarcerated previously. Items with low levels of reliability included items that are rarely documented and those that are subjective in nature, such as whether the inmate could make better use of his or her time.

These item-by-item results were obtained using a relatively liberal definition of agreement. Those items with a range of 0 to 3 points were scored in a truncated manner (0-1 points or 2-3 points). In other words, if the first rater scores '0' and the second rater scores '1' for the same item, both of which are considered a "yes" response, the item was considered to be in agreement. Despite this more liberal reliability test, there continues to be a high level of disagreement or inconsistency for certain items. As will be discussed later in detail, this low level of reliability creates difficulty in the validation of the individual items. For some of the items, there is a 50 percent likelihood that the inmate would score differently depending on who interviews the inmate.

Given the high level of disparity across the two interviewers, we undertook several analyses in an effort to locate the source of the discrepancies. First, we examined whether some of the differences in the overall risk level were due to addition errors in computing the total score. The sum of the scores on all items was computer generated and the resultant risk level was compared to the original risk level assigned by the interviewers. The risk level is properly identified for all cases on the first set of interviews and only three errors were made on the second interview. Thus, the level of disagreement between the first rater and the second rater cannot be explained by errors in calculation.

<b>Table 2</b>				
<b>Comparison of PBPP Risk Levels Across Interviewers of the Reliability Sample</b>				
<b>LSI-R Risk Level</b>	<b>1<sup>st</sup> Interview</b>		<b>2<sup>nd</sup> Interview</b>	
	<b>N</b>	<b>100%</b>	<b>N</b>	<b>100%</b>
Low (0 through 15)	5	4	7	6
Medium (16 through 22)	28	23	33	28
High (23 and above)	87	73	78	65
Missing	0	0	2	2
<b>Total</b>	120	100	120	100

<b>Table 3</b>				
<b>Cross-tabulation of the First and Second LSI-R Interviews of the Reliability Sample</b>				
<b>1<sup>st</sup> Interview</b>	<b>2<sup>nd</sup> Interview</b>			
	<b>Low</b>	<b>Medium</b>	<b>High</b>	<b>Total*</b>
Low (0 through 15)	3	2	0	5
Medium (16 through 22)	3	14	11	28
High (23 and above)	1	17	67	85
<b>Total</b>	7	33	78	118
% Disagreement: 29%; % Disagreement One Level: 28%; % Disagreement Two Levels: 1%				
*Note: Two cases were not scored a second time and were excluded from this analysis. Source: Pennsylvania Board of Probation and Parole				

**Table 4**  
**LSI-R Agreement Rates for the 54 Scoring Items of the Reliability Sample**

<b>Variable</b>	<b>% Agreement</b>	<b>Variable</b>	<b>% Agreement</b>
<b>I. Criminal History</b>		<b>V. Accommodations</b>	
1. Any prior convictions?	96	27. Unsatisfactory	64
2. Two or more prior convictions?	93	28. Three or more address changes last year?	82
3. Three or more convictions?	93	29. High crime neighborhood?	67
4. Three or more present offenses?	81	<b>VI. Leisure/Recreation</b>	
5. Arrested under age 16?	78	30. Absence of recent participation in an org. activity?	53
6. Ever incarcerated upon conviction?	95	31. Could make better use of time	53
7. Escape history from a correctional facility?	81	<b>VII. Companions</b>	
8. Ever punished for institutional misconduct?	87	32. A social isolate?	68
9. Charge made for probation/parole suspended during prior community supervision?	91	33. Some criminal acquaintances?	77
10. Official record of assault/violence?	86	34. Some criminal friends?	59
<b>II. Education/Employment</b>		35. Few anti-criminal acquaintances?	61
11. Currently employed?	86	36. Few anti-criminal friends?	59
12. Frequently unemployed?	72	<b>VII. Alcohol/Drug Problem</b>	
13. Never employed for a full year?	72	37. Alcohol problem, ever?	76
14. Ever fired?	78	38. Drug problem, ever?	88
15. Less than regular grade 10?	85	39. Alcohol problem, currently?	72
16. Less than regular grade 12?	88	40. Drug problem, currently?	68
17. Suspended or expelled at least once?	76	41. Law violations?	79
18. Participation/performance	78	42. Marital/family?	68
19. Peer interactions	76	43. School/work?	62
20. Authority interactions	75	44. Medical?	78
<b>III. Financial</b>		45. Other indicators?	63
21. Problems?	60	<b>IX. Emotional/Personal</b>	
22. Reliance upon social assistance?	69	46. Moderate interference?	84
<b>IV. Family/Marital</b>		47. Severe interference, active psychosis?	93
23. Dissatisfaction with marital or equivalent situation?	64	48. Mental health treatment, past?	87
24. Non-rewarding, parental	62	49. Mental health treatment, present?	89
25. Non-rewarding, other relative	65	50. Psychological assessment indicated?	66
26. Criminal family/spouse?	68	<b>X. Attitudes/Orientation</b>	
		51. Supportive of crime	68
		52. Unfavorable toward convention	78
		53. Poor, toward sentence?	72
		54. Poor, toward supervision?	62

Further, there are no significant differences in the mean, minimum or maximum scores when comparing the original and re-computed score. Most of the scores were calculated correctly for the first and second interview (91% and 86%, respectively). For the first interview, four percent have positive errors (original was higher than re-computed score) and five percent have negative errors (original was less than re-computed score). On the second interview, four percent have positive errors and eight percent have negative errors. These findings confirm that the differences in ratings across interviewers cannot be attributed to calculation errors.

Finally, we assessed the distribution of the total scores for both sets of interviews to see if large numbers of cases are clustered along the cut-off points of the three risk levels. If this is the case, disagreement on individual items that changed the total score by only a few points would have also changed the assessed risk level for a significant number of cases. Very few cases are clustered around the low-medium cut-points, although between 20 and 25 percent of cases are clustered around the medium-high cut points.

These cases (n=52) were examined more closely to ascertain whether small differences in total score between the first and second interviewer were responsible for the divergent risk categorizations. For example, if a single case has a total score of 22 points by the first rater and a total score of 23 points by the second rater, the risk level would not have matched (the first rater would score the inmate as medium risk and the second rater would score the inmate as high risk) even though the actual magnitude of the difference in scores is quite small. Only 40 percent of the cases fall into this category. Conversely, for 60 percent of these cases, the risk level categorizations do not agree because the total score by the first and second rater diverge significantly.

Based upon these reliability results, the PBPP undertook a second test after institutional parole staff had received additional training in using the LSI-R. A sample of 156 cases was selected by the PBPP and scored twice by independent scorers in September 2002. The same analysis completed for the first reliability sample was again performed on this sample with the results shown in Tables 5 through 7. Here one can see that the percentage of agreement has improved over previous results. In contrast to the first reliability study in which only 18 of 54 items (33%) had an agreement rate of 80 percent, a minimally acceptable performance standard at the PBPP, in the second study inter-rater agreement improved to 34 of 54 items (63%) at the 80 percent threshold. Likewise, in the first study, only 6 of 54 items (11%) had an agreement rate at an optimal level of 90 percent or higher in contrast to 19 of 54 items (35%) in the second study that met the 90 percent agreement threshold. The rate of agreement in the overall risk level was increased from 71 percent to 88 percent. These results indicate that it is possible to increase the level of reliability of the LSI-R with additional and more intensive staff training. However, the more subjective items (i.e. financial, family/marital), continue to be unreliable.

<b>Table 5</b>				
<b>Comparison of PBPP Risk Levels Across Interviewers of the Reliability Sample</b>				
<b>LSI-R Risk Level</b>	<b>1<sup>st</sup> Interview</b>		<b>2<sup>nd</sup> Interview</b>	
	<b>N</b>	<b>100%</b>	<b>N</b>	<b>100%</b>
Low (0 through 15)	7	5	7	5
Medium (16 through 22)	36	23	31	20
High (23 and above)	113	72	118	75
<b>Total</b>	156	100	156	100

<b>Table 6</b>				
<b>Cross-tabulation of the First and Second LSI-R Interviews of the Reliability Sample</b>				
<b>1<sup>st</sup> Interview</b>	<b>2<sup>nd</sup> Interview</b>			
	<b>Low</b>	<b>Medium</b>	<b>High</b>	<b>Total*</b>
Low (0 through 15)	4	2	1	7
Medium (16 through 22)	3	25	8	36
High (23 and above)	0	4	109	113
<b>Total</b>	7	31	118	156

% Disagreement: 12%; % Disagreement One Level: 11%; % Disagreement Two Levels: 1%

**Table 7**  
**LSI-R Agreement Rates for the 54 Scoring Items of the Reliability Sample**

<b>Variable</b>	<b>% Agreement</b>	<b>Variable</b>	<b>% Agreement</b>
<b>I. Criminal History</b>		<b>V. Accommodations</b>	
1. Any prior convictions?	94	27. Unsatisfactory	<b>64</b>
2. Two or more prior convictions?	94	28. Three or more address changes last year?	<b>89</b>
3. Three or more convictions?	95	29. High crime neighborhood?	<b>83</b>
4. Three or more present offenses?	<b>89</b>	<b>VI. Leisure/Recreation</b>	
5. Arrested under age 16?	<b>89</b>	30. Absence of recent participation in an org. activity?	<b>76</b>
6. Ever incarcerated upon conviction?	99	31. Could make better use of time	<b>58</b>
7. Escape history from a correctional facility?	92	<b>VII. Companions</b>	
8. Ever punished for institutional misconduct?	97	32. A social isolate?	<b>88</b>
9. Charge made for probation/parole suspended during prior community supervision?	90	33. Some criminal acquaintances?	91
10. Official record of assault/violence?	92	34. Some criminal friends?	<b>78</b>
<b>II. Education/Employment</b>		35. Few anti-criminal acquaintances?	<b>77</b>
11. Currently employed?	98	36. Few anti-criminal friends?	<b>79</b>
12. Frequently unemployed?	<b>83</b>	<b>VII. Alcohol/Drug Problem</b>	
13. Never employed for a full year?	<b>86</b>	37. Alcohol problem, ever?	<b>86</b>
14. Ever fired?	<b>87</b>	38. Drug problem, ever?	93
15. Less than regular grade 10?	92	39. Alcohol problem, currently?	<b>59</b>
16. Less than regular grade 12?	97	40. Drug problem, currently?	<b>58</b>
17. Suspended or expelled at least once?	<b>88</b>	41. Law violations?	<b>88</b>
18. Participation/performance	99	42. Marital/family?	<b>83</b>
19. Peer interactions	99	43. School/work?	<b>78</b>
20. Authority interactions	99	44. Medical?	92
<b>III. Financial</b>		45. Other indicators?	<b>74</b>
21. Problems?	<b>54</b>	<b>IX. Emotional/Personal</b>	
22. Reliance upon social assistance?	<b>78</b>	46. Moderate interference?	<b>85</b>
<b>IV. Family/Marital</b>		47. Severe interference, active psychosis?	<b>100</b>
23. Dissatisfaction with marital or equivalent situation?	<b>59</b>	48. Mental health treatment, past?	<b>74</b>
24. Non-rewarding, parental	<b>62</b>	49. Mental health treatment, present?	<b>75</b>
25. Non-rewarding, other relative	<b>69</b>	50. Psychological assessment indicated?	<b>81</b>
26. Criminal family/spouse?	<b>87</b>	<b>X. Attitudes/Orientation</b>	
		51. Supportive of crime	<b>63</b>
		52. Unfavorable toward convention	<b>72</b>
		53. Poor, toward sentence?	92
		54. Poor, toward supervision?	<b>88</b>



## C. Conclusions and Recommendations

The above analysis clearly shows that, while scoring reliability improved as a result of recent training efforts, significant improvements are still needed to ensure a suitable level of reliability. These improvements should strive to bring clarity to each item and to reduce subjectivity. It is critically important that PBPP staff learn to administer the LSI-R in a fashion that results in consistent scoring across staff members. Without this reliability, it will be extremely difficult to monitor the system's on-going effectiveness in properly assessing the risk to public safety posed by Pennsylvania's parole-eligible prisoners.

However, the process by which these improvements can be made must account for existing limitations in the flexibility of the assessment system itself. Because the LSI-R is a privately-owned screening system it cannot be altered or modified without the permission of the vendor. Furthermore, the PBPP already has invested considerable resources in staff training and the automation of this system. For these reasons, the following recommendations are focused on providing additional staff training in the use of the LSI-R. We also urge the PBPP to forward this report to the LSI-R vendor to seek their guidance for enhancing the performance of the LSI-R for the Pennsylvania inmate population.

**1. Continue to focus on reliability.** No matter how the LSI-R is to be used, continued efforts are required to improve the reliability of the scoring of the instrument. Conclusions about the validity and impact of the system cannot be made without confidence in the reliability of the instruments. We recommend that the PBPP strive to achieve at minimum 80 percent agreement rates, but optimally 90 percent agreement rates for each of the 54 items and the risk classification level.

**2. Develop a training strategy.** Clearly, the recent training effort helped to improve understanding of individual items and the accuracy of scoring among staff. In addition to clarifying the intent of each item, the training effort should also focus on directing staff to appropriate sources of information and acceptable means of verifying information. Given the training results demonstrated by PBPP with this and other instruments, a small group or "train-the-trainer," format appears to work well as an initial step to developing a shared understanding of the meaning of individual items.

Consultations among staff to discuss the reasons for divergent ratings also appear to be a useful tool. This training will need to extend to all staff members who are responsible for scoring the LSI-R, and should be on-going to ensure that new staff is provided with intensive instruction. Trainings should include practice sessions for scoring actual cases, and the levels of agreement across staff should be continuously monitored until minimal 80 percent, or the optimal 90 percent threshold is reached consistently. The complexity of this endeavor should not be underestimated.

### **III. VALIDATION OF THE LSI-R**

The purpose of the validation study is to measure how well the LSI-R instrument predicts success or failure among PA prisoners under state parole supervision. A valid risk assessment instrument will identify distinct groups of offenders with different likelihoods of re-offending. In other words, the recidivism rate of the group of offenders identified as “low risk” will be significantly lower than the recidivism rate of those identified as “medium risk”, which will be significantly lower than “high risk.”

It should be noted that the results of this analysis are limited by the weakness revealed by the inter-rater reliability study. Until the scoring of the LSI-R (or an alternative risk assessment instrument) reaches an acceptable level of consistency across items and risk level, the association between risk level and recidivism should be considered preliminary.

#### **A. Methodology**

##### ***1. Sampling Method***

Ideally, a true validation sample would consist of all prisoners who were released from prison during a given time frame. This was not possible for a number of reasons. First, the LSI-R requires an interview with the prisoner. Since the LSI-R was being pilot tested at selected facilities, it was not possible to locate a cohort of LSI-R scored prisoners that represented all prisoners being released. Second, since we needed to complete this study in a timely manner, prisoners who were scored on the LSI-R at these facilities but who were not released on parole or who later discharged from prison without the benefit of parole were also excluded from the sample. Thus the final validation sample consisted of 1,006 male prisoners who had an LSI-R completed on them and who were released to parole from nine test facilities. Put differently, the validation results measures how well the LSI-R instrument predicts the incidence of recidivism for prisoners released to parole.

##### ***2. Data Collection and Analysis***

Data extracted from PBPP files provide a profile of the sample’s demographic and offense characteristics. In addition, LSI-R data were available for all cases, which included responses to each item on the instrument, the total score and resultant assessed risk level. The offender’s success on parole was tracked, and recidivism data for those who failed during the 12-month follow-up period was collected. These outcome data for cases that returned to a correctional facility were extracted from PBPP files, which included the date of the recidivism event and the nature of the event. For the purposes of this study, arrests, detentions, absconders, and returns to prison are considered recidivists.

---

<sup>2</sup>Demographic and offense characteristics for the population from which the sample was drawn were not available; therefore, a comparative analysis to ensure the representativeness of the sample could not be conducted.

The PBPP forwarded the data file to the ICJC, where it was converted into an SPSS database for statistical analysis. Descriptive (e.g., frequencies and cross-tabulations) analyses were conducted in order to produce tables that present demographic and legal characteristics, the distribution of scores across items, and predictive attributes of the LSI-R items and risk levels.

## B. Results

The outcome variable of interest is recidivism. Descriptive statistics are used to develop a profile of the total sample, and the group of prisoners who recidivated. The distribution of scores across all items is also presented, with comparisons between the successful and unsuccessful parolees.

Table 8 presents the distribution of offenders across various parole outcomes, both successful and unsuccessful. Of the 53 percent who recidivated, over 60 percent were recommitted for technical violations. Less than 20 percent for committed a new crime or absconding. Approximately 32 percent of the sample continues to report regularly, while 12 percent have had their sentences expire.

<b>Table 8</b>		
<b>Summary of One Year Recidivism Measures</b>		
<b>Pennsylvania Board of Probation and Parole</b>		
<b>Outcome Measure</b>	<b>N</b>	<b>%</b>
<b>Total</b>	<b>1,006</b>	<b>100%</b>
<b>Successful</b>	<b>468</b>	<b>47%</b>
Reporting Regularly	322	32%
Sentence Expired	117	12%
Residing in CCC	23	2%
Other	6	1%
<b>Unsuccessful (recidivism)</b>	<b>538</b>	<b>53%</b>
Recommitted for Technical Violation	328	33%
Recommitted for New Crime	125	12%
Absconded	74	7%
Detainer (County, Federal, State)	7	1%
Other	4	0%

The first task was to perform an item by item test of 54 LSI-R scoring items to see which ones were associated with recidivism. This analysis showed that only the following 11 LSI-R items had a statistical association with recidivism:

1. Any prior convictions?
2. Two or more prior convictions?
3. Three or more prior convictions?
4. Arrested under age 16?
5. Escape History?
6. Probation/parole suspension during prior community supervision?
7. Three or more address changes the past year?
8. Current drug problem?
9. Drug problem related to law violations?
10. Drug problem related to school or work problems?
11. Mental health problems in the past?

We then assessed the extent to which the computed LSI-R score was associated with recidivism (see Table 9). The results shown here show that the LSI-R point total is associated with recidivism. Forty three percent of offenders who scored as “low risk” recidivated, while 51 percent of “medium risk” offenders and 58 percent of “high risk” offenders recidivated. The same relationship appears whether one uses the technical violation or new conviction criteria. These results show that only a small number of the LSI-R scoring items are useful and that most of them are not contributing to the risk assessment process.

Table 10 presents select items from the LSI-R instrument and their relationship to the failure variable. These items were selected by examining the differential rates of recidivism within each item. Items in which offenders scoring in one direction exhibit distinctly different rates of recidivism from the group of offenders scoring in the other direction are considered to be useful for their predictive validity. These items are: any prior convictions, two or more prior convictions, arrested under age 16, prior probation/parole suspension, three or more address changes within the last year, current drug problem, problem affecting school/work, and mental health treatment in the past.

Table 11 presents the distribution across risk levels for the condensed instrument, and the respective rates of recidivism. Compared to the risk groups created by the full LSI-R, the condensed instrument creates risk categories with greater distinctiveness in terms of recidivism. Not only do these items have better predictive ability, but also they reduce the “high risk” category. According to this instrument, only 188 prisoners would be classified as “high risk,” compared to 522 using the full LSI-R instrument. More importantly, the high risk group created by the condensed instrument has a 69 percent recidivism rate, compared to the 58 percent recidivism rate of the LSI-R high risk group, indicating the condensed instrument does a better job of selecting those prisoners representing the most significant danger to public safety.

**Table 9**  
**Current LSI-R Score by Failure Variable**  
**Pennsylvania Board of Probation and Parole**

Point Distribution		Total		Recidivated	Technical	Convicted
		N	%	%	%	%
Low Risk	0	0	0.0%	N/A	N/A	N/A
	1	1	0.1%	N/A	N/A	N/A
	2	1	0.1%	100.0%	0.0%	0.0%
	3	0	0.0%	N/A	N/A	N/A
	4	0	0.0%	N/A	N/A	N/A
	5	0	0.0%	N/A	N/A	N/A
	6	0	0.0%	N/A	N/A	N/A
	7	0	0.0%	N/A	N/A	N/A
	8	2	0.2%	N/A	N/A	N/A
	9	2	0.2%	N/A	N/A	N/A
	10	2	0.2%	N/A	N/A	N/A
	11	2	0.2%	50.0%	0.0%	50.0%
	12	5	0.5%	60.0%	40.0%	20.0%
	13	5	0.5%	40.0%	40.0%	0.0%
	14	10	1.0%	50.0%	40.0%	0.0%
	15	13	1.3%	38.5%	15.4%	7.7%
	16	20	2.0%	40.0%	20.0%	10.0%
17	23	2.3%	52.2%	39.1%	4.3%	
<b>Total Low Risk</b>		<b>86</b>	<b>8.7%</b>	<b>43.0%</b>	<b>26.7%</b>	<b>7.0%</b>
Medium Risk	18	34	3.5%	50.0%	26.5%	5.9%
	19	41	4.2%	58.5%	34.1%	14.6%
	20	56	5.7%	41.1%	23.2%	8.9%
	21	51	5.2%	52.9%	31.4%	7.8%
	22	62	6.3%	53.2%	29.0%	8.1%
	23	74	7.5%	50.0%	39.2%	6.8%
	24	59	6.0%	62.7%	35.6%	6.8%
<b>Total Medium Risk</b>		<b>377</b>	<b>38.3%</b>	<b>50.9%</b>	<b>31.8%</b>	<b>8.2%</b>
High Risk	25	79	8.0%	50.6%	34.2%	11.4%
	26	63	6.4%	54.0%	34.9%	6.3%
	27	50	5.1%	62.0%	40.0%	10.0%
	28	49	5.0%	51.0%	20.4%	12.2%
	29	47	4.8%	70.2%	40.4%	19.1%
	30	56	5.7%	69.6%	37.5%	19.6%
	31	24	2.4%	54.2%	20.8%	16.7%
	32	37	3.8%	59.5%	40.5%	10.8%
	33	34	3.5%	58.8%	29.4%	14.7%
	34	24	2.4%	54.2%	37.5%	0.0%
	35	15	1.5%	53.3%	26.7%	6.7%
	36	19	1.9%	36.8%	15.8%	15.8%
	37	10	1.0%	40.0%	30.0%	0.0%
	38	4	0.4%	75.0%	25.0%	0.0%
	39	3	0.3%	66.7%	33.3%	0.0%
	40	3	0.3%	100.0%	100.0%	0.0%
	41	0	0.0%	N/A	N/A	N/A
	42	2	0.2%	100.0%	50.0%	0.0%
	43	2	0.2%	100.0%	50.0%	50.0%
	44	0	0.0%	N/A	N/A	N/A
45	0	0.0%	N/A	N/A	N/A	
46	0	0.0%	N/A	N/A	N/A	
47	1	0.1%	100.0%	0.0%	100.0%	
<b>Total High Risk</b>		<b>522</b>	<b>53.0%</b>	<b>57.9%</b>	<b>33.5%</b>	<b>12.1%</b>

**Table 10**  
**Select LSI-R Items by Failure Variable**  
**Pennsylvania Board of Probation and Parole**

Parole Guidelines Item	Total		Recidivated	Technical	Convicted
	N	%	%	%	%
	<b>TOTAL</b>	<b>1,006</b>	<b>100.0%</b>	<b>53.5%</b>	<b>32.6%</b>
<b>Any prior convictions?</b>					
Yes (1)	858	77.6%	54.7%	33.7%	9.9%
No (0)	145	13.1%	46.9%	26.2%	10.3%
Missing	3	0.3%	N/A	N/A	N/A
<b>Two or more prior convictions?</b>					
Yes (1)	733	72.9%	55.5%	33.8%	9.8%
No (0)	270	24.4%	48.1%	29.3%	10.4%
Missing	3	0.3%	N/A	N/A	N/A
<b>Arrested under age 16?</b>					
Yes (2)	446	44.3%	60.1%	35.9%	12.8%
No (0)	543	54.0%	48.3%	30.2%	7.6%
Missing	17	1.7%	N/A	N/A	N/A
<b>Prior Prob/Parole suspension?</b>					
Yes (2)	740	73.6%	57.2%	34.5%	10.4%
No (0)	257	25.5%	42.8%	26.8%	8.6%
Missing	9	0.9%	N/A	N/A	N/A
<b>Three or more address changes last year?</b>					
Yes (2)	120	11.9%	59.2%	36.7%	11.7%
No (0)	868	86.3%	52.5%	31.9%	9.7%
Missing	18	1.8%	N/A	N/A	N/A
<b>Drug problem currently?</b>					
Yes (1)	542	53.9%	55.4%	34.9%	10.0%
No (0)	448	44.5%	50.9%	29.7%	9.6%
Missing	16	1.6%	N/A	N/A	N/A
<b>School/work?</b>					
Yes (2)	429	42.6%	59.9%	36.6%	11.7%
No (0)	570	56.7%	48.6%	29.5%	8.6%
Missing	7	0.7%	N/A	N/A	N/A
<b>Mental health treatment, past?</b>					
Yes (2)	261	25.9%	58.6%	37.9%	7.7%
No (0)	739	73.5%	51.7%	30.6%	10.8%
Missing	6	0.6%	N/A	N/A	N/A

<b>Table 11</b>						
<b>Score using Select LSI-R Items by Failure Variable</b>						
<b>Pennsylvania Board of Probation and Parole</b>						
Point Distribution		Total		Recidivated	Technical	Convicted
		N	%	%	%	%
<b>TOTAL</b>		<b>948</b>	<b>100.0%</b>	<b>53.4%</b>	<b>32.5%</b>	<b>9.7%</b>
Low	0	17	1.8%	17.6%	0.0%	11.8%
	1	17	1.8%	29.4%	5.9%	23.5%
	2	58	6.1%	22.4%	19.0%	10.3%
	3	54	5.7%	33.3%	25.9%	1.9%
	<b>Total Low Risk</b>	<b>146</b>	<b>15.4%</b>	<b>33.6%</b>	<b>17.8%</b>	<b>8.9%</b>
Moderate	4	119	12.6%	52.9%	35.3%	6.7%
	5	115	12.1%	47.8%	28.7%	8.7%
	6	133	14.0%	51.9%	31.6%	9.8%
	7	160	16.9%	56.3%	36.3%	10.0%
	8	87	9.2%	58.6%	27.6%	13.8%
<b>Total Moderate Risk</b>	<b>614</b>	<b>64.8%</b>	<b>53.4%</b>	<b>32.4%</b>	<b>9.6%</b>	
High	9	115	12.1%	70.4%	48.7%	7.8%
	10	23	2.4%	73.9%	47.8%	13.0%
	11	43	4.5%	60.5%	34.9%	11.6%
	12	1	0.1%	0.0%	0.0%	0.0%
	13	6	0.6%	83.3%	16.7%	50.0%
<b>Total High Risk</b>	<b>188</b>	<b>19.8%</b>	<b>68.6%</b>	<b>44.1%</b>	<b>10.6%</b>	

<b>Table 12</b>						
<b>Score using Select LSI-R and Demographic Items by Failure</b>						
<b>Pennsylvania Board of Probation and Parole</b>						
Point Distribution		Total		Recidivated	Technical	Convicted
		N	%	%	%	%
<b>TOTAL</b>		<b>848</b>	<b>100.0%</b>	<b>53.2%</b>	<b>31.8%</b>	<b>10.1%</b>
Lowest	1	1	0.1%	0.0%	0.0%	0.0%
	2	7	0.8%	0.0%	0.0%	0.0%
	3	23	2.7%	21.7%	8.7%	8.7%
<b>Total Lowest Risk</b>	<b>31</b>	<b>3.7%</b>	<b>16.1%</b>	<b>6.5%</b>	<b>6.5%</b>	
Low	4	20	2.4%	45.0%	25.0%	5.0%
	5	43	5.1%	37.2%	18.6%	14.0%
	6	64	7.5%	37.5%	20.3%	6.3%
<b>Total Low Risk</b>	<b>127</b>	<b>15.0%</b>	<b>38.6%</b>	<b>20.5%</b>	<b>8.7%</b>	
Moderate	7	89	10.5%	51.7%	29.2%	13.5%
	8	91	10.7%	46.2%	34.1%	2.2%
	9	115	13.6%	52.2%	34.8%	10.4%
	10	104	12.3%	57.7%	29.8%	11.5%
	11	94	11.1%	56.4%	37.2%	10.6%
<b>Total Moderate Risk</b>	<b>493</b>	<b>58.1%</b>	<b>52.9%</b>	<b>33.1%</b>	<b>9.7%</b>	
High	12	92	10.8%	62.0%	37.0%	13.0%
	13	41	4.8%	78.0%	31.7%	19.5%
	14	29	3.4%	89.7%	65.5%	3.4%
	15	19	2.2%	52.6%	36.8%	0.0%
	16	12	1.4%	66.7%	50.0%	16.7%
17	4	0.5%	75.0%	0.0%	50.0%	
<b>Total High Risk</b>	<b>197</b>	<b>23.2%</b>	<b>69.0%</b>	<b>40.1%</b>	<b>12.7%</b>	

In Table 12, the analysis is taken a step further. Along with the eight LSI-R items in the condensed instrument, we also include these descriptive variables: age at release, marital status, committing offense, and release type. This instrument, combining a small number of reliable LSI-R items with a few demographic items, produced the best risk assessment results. In this analysis, we are able to develop greater specificity within the “low risk” category and to identify groups of prisoners with more distinct rates of reoffending.

There is another version of the LSI-R instrument that closely resembles the condensed version noted above. The LSI-SV (or Screening Version) was first developed by the developers of the LSI-R for the state of Washington for its jail system. The desire was to develop an abbreviated assessment process that could be applied quickly and accurately for the large volume of persons admitted to jail. The LSI-SV consists of eight items – several of which are the same ones found in this study to be associated with recidivism. It is the opinion of the authors of this report, that the LSI-SV could be used as a risk assessment instrument.

## **B. Conclusions and Recommendations**

The LSI-R, in its full version, capably identifies distinct groups of PA offenders with different likelihoods of recidivating. However, the measured lack of reliability in many of the items created “noise” in the instrument, preventing it from being sufficiently precise. Using a condensed instrument that has items with higher rates of reliability, would result in greater specificity in the identification of prisoners with the highest, and lowest, risk to public safety.

Thus, we recommend that the PBPP not use the full LSI-R for parole consideration decisions in assessing risk. Instead a more succinct instrument such as the LSI-SV would be more effective for the Board’s risk assessment. The LSI-R, in its full version, is best suited for institutional case planning upon admission to prison and to determine the level of community supervision required once parole has been granted.

Should the PBPP decide to use the LSI-SV for risk assessment, it should carefully monitor its use over a 12 month period to re-validate its predictive attributes.



## **Addendum to the Study**

Although not part of the original study, the PBPP Research Committee requested that we comment and make recommendations on how the results of this research could impact the PBPP guidelines. Those guidelines consist of the following five factors:

1. Violence Indicator;
2. LSI-R Score;
3. Sex Offender Risk Assessment Score;
4. Institutional Programming Score; and,
5. Institutional Behavior Score.

Under this system, the LSI-R factor is just one of five factors in the guidelines ending up with a recommendation to the Board of either “likely parole” or “unlikely to parole”. For this cohort, 84% of the releases were scored as “likely to parole”. However, the LSI-R score showed that 53% of the parolees were scored as “high risk”. The proposed change to the to the risk assessment instrument would result in only 23 percent of the cases being scored as “high risk” and a much larger proportion being scored as “moderate” and “low” risk.

Assuming the PBPP adopts the LSI-SV as the means for assessing risk, it should be a three level scale that conforms to the existing PBPP risk categories of high, moderate and low with a distribution that will approximate the distribution reported in this report (approximately 20-25 percent “high risk”, 50 percent “moderate risk” and the remaining 25% low risk.

In so doing the number of prisoners assessed as “likely parole” will increase slightly assuming no other changes are made in the other four guidelines items and the guideline scale. Here again, the PBPP should carefully monitor and evaluate how the prisoners are being scored on the guidelines over the next 12 months and make whatever adjustments are warranted.