How Much Risk Can We Take? The Misuse of Risk Assessment in Corrections

James Austin, Ph.D.
The JFA Institute

The Basics of Risk Assessment
A Closer Look at the LSI-R
Linking Risk to Punishment and Treatment?

AFTER DECADES OF intellectual neglect, the field of corrections has decided to embrace the world of science and adhere to the dictums of "evidence-based" corrections. The term "evidence based" originates from the field of medicine as far back as the 19th century in Europe and means many things to many people. [1] In medicine it is very important that medical procedures and the use of healing drugs and medicine actually demonstrate their effectiveness through rigorous experimental studies before they are brought to market. In the social sciences, evidence-based research suggests that governmental policies must be shaped by scientific evidence that shows the policy has some cause and effect value. For many good reasons, the field of corrections has never had to pass such a high standard. But after American corrections has set world records in the numbers of persons incarcerated and placed on probation and parole, some criminal justice professionals believe the field needs to get serious about its $60 billion a year industry and produce a better product.

Plagued by recidivism rates that have remained stubbornly stagnant for 30 years (or more) and by a general feeling among most politicians that about the only thing that corrections can do is inflict widespread punishment, criminal justice practitioners have seen the more benign goals of treatment and rehabilitation take a back seat to the more politically appealing ideologies of deterrence, incapacitation, and retribution. It's a given that no politician can successfully run on a platform demanding more and better treatment for the two million plus prisoners held in our nation's jails and prisons.

But the times are a changing. Led by a small number of Canadian and American criminologists, there is now a considerable effort to get rehabilitation and treatment back on the map. Their argument is advertised not as ideological but as empirical. The major premise is that treatment does work if it is done right. Therefore, the primary reason treatment is ineffective is because it is more often done wrong.

One major reason that treatment is not done right is that offenders are not properly assessed for risk by most correctional agencies. Without the proper diagnosis, it is not possible to assign prisoners to the proper treatment. Indeed, prior research has shown that assigning low-risk people to treatment they really don't need actually increases recidivism. A recent evaluation of Ohio's community corrections act clearly shows that many correctional programs are not targeting the proper offender, which in turn diminishes the capacity to reduce recidivism rates. [2]

The widespread absence of risk assessment in corrections has historically hampered correctly targeted treatment. It was not until the 1980s that prison systems, due in part to a number of federal lawsuits, finally started using custody classification systems to assign prisoners to the correct prisons. The results have been impressive in most states, with increasing numbers of prisoners now being assigned to minimum security settings. The taxpayers have benefited somewhat because the lower the security, the lower the incarceration costs. Unfortunately, the huge increases in the correctional populations have largely negated whatever savings taxpayers would have realized.

Parole boards, which still govern the date and conditions of release for prisoners in most states, have only recently (and only in a few states) embraced the idea that their decisions would be influenced by some calculation of the prisoner's risk to recidivate. Probation and parole agencies have also begun to implement risk instruments to guide their decisions as to what levels of supervision are most appropriate for their burgeoning caseloads.

But despite these advances, no jurisdiction can point to significant reductions in recidivism rates —and that includes Canada, from which most of the new emphasis on rehabilitation has emanated. [3] Many probation and parole officers seem less interested in risk assessment and case management and more concerned with racking up as many violations of their caseload as they can. I don't recall any prison, parole, or probation department being chastised for having too high a recidivism rate, even though there is considerable evidence that they could have a positive effect on these rates.

The remainder of this article focuses on the state of risk assessment. I concede that in order for rehabilitation to have a meaningful impact on recidivism rates, the proper identification of persons by their risk level is essential. But I now worry that the field is placing too much emphasis on risk assessment with little effort to provide those basic treatment services that are needed.

The Basics of Risk Assessment

Before an agency decides to adopt a risk assessment system, a number of tests need to be completed to ensure it will work. There seems to be a trend in corrections to uncritically accept the latest "innovation" and adopt it without understanding its strengths and limitations. In risk assessment, unless these steps are completed, application of the risk assessment process may prove more harmful than helpful as offenders will be improperly classified.

1. Risk Assessment Instruments Must Be Tested on Your Correctional Population and Separately Normed for Males and Females.

There is a tendency for correctional agencies to simply borrow or buy an instrument that has been developed on another population that may not reflect the attributes of their own offender populations. In research terms this issue has to do with the "external validity" of the instrument and the ability to generalize the findings of a single study of the instrument to other jurisdictions. Generally, if a risk assessment instrument has not been tested on multiple populations under varying conditions, it will not work well on populations it has not been tested on.

Male and female risk assessment is another issue for proposed risk instruments. Men and women are different, behave differently, and respond differently to various forms of treatment and supervision. Yet when it comes to risk assessment we often assume they are the same. Recidivism and career criminal studies consistently show that females are less involved in criminal behavior, are less likely to commit violent crimes and are less likely to recidivate after being placed on probation or parole. Further, since the "criminal population" is largely male, any instrument that is tested on a total correctional population will naturally misclassify females.

2. An Inter-Rater Reliability Test Must Be Conducted

Both an inter-reliability test and a validity test must be completed by independent researchers who have no economic gain in proving the effectiveness of the instrument. Inter-rater reliability has to do with the accuracy and consistency of the instrument being completed by those who will be responsible on a day-to-day basis for completing the form and interpreting the results. Often this work is done by probation and parole officers or parole board hearing examiners. It is a skilled task that not all correctional staff are well suited for.

The inter-rater reliability test would consist of taking a representative sample of offenders (a minimum of 100 cases) who will then be independently scored using the proposed instrument by two staff who have trained in the proposed instrument. Any item on the instrument that does not reach the 80 percent agreement level should be deleted. If the instrument does not demonstrate an agreement level of 90 percent, it should not be implemented.

3. A Validity Test Must Be Conducted

The validity test is designed to see how well the risk factors actually predict recidivism. This test is done by drawing a sample of offenders who were sentenced to probation or released from prison and tracking them for a period of 2 to 3 years. Since most jurisdictions are anxious to have the risk assessment instrument implemented as quickly as possible, the validation sample often consists of persons sentenced or released 2 to 3 years prior to the study being conducted. The research must then be able to perform a variety of bi-variate and multi-variate statistical tests to determine which items should be used, the weights assigned to each item and the proper risk level scale.

4. The Instruments Must Allow For Dynamic and Static Factors That Have Been Well Accepted and Tested In A Number of Jurisdictions

As noted above, the risk instrument should consist of static and dynamic risk items. Table 5 summarizes commonly used risk factors that have been repeatedly validated by a number of validation studies. These are separated into the static and dynamic categories. Of the two, the dynamic factors are generally the more powerful predictors, as they reflect the person's current social and economic environment. If an instrument does not employ dynamic factors, it is likely to not perform accurately.

5. The Instruments Must Be Compatible with the Skill Level of Your Staff

There are a wide variety of risk assessment instruments available to jurisdictions to use. However, they require very different skill levels. The more traditional risk assessment forms generally consist of not more than 10-12 items and are based on factual items that can be gleaned from court and case files and require minimal interpretation by staff trained in their use. Age at first arrest, current age, and number of prior probation violations within the past five years come under this category. For these instruments staff need little academic training to conduct an accurate assessment.

The more complicated risk assessment items require a well-structured interview and a review of all relevant case file data. These instruments often have 40-60 items with several sub-scales reflecting varying domain risk levels. With such instruments it is more difficult to achieve the minimal levels of reliability and validity, unless the staff is highly skilled in the application of psychometric assessment forms. Without such skilled staff, the use of these instruments is not recommended.

6. The Risk Assessment Must Have "Face Validity" And Transparency with Staff, Prisoners, Probationers, Parolees and Policy Makers.

Finally, the instrument and the entire risk assessment process needs to be credible with all of the parties that are being directly impacted by it. Staff assigned to the risk assessment process must believe that the instrument actually works and will help inform the decision process for sentencing, release, and supervision decisions. The decision makers (judges, parole boards, and

correctional administrators) must also have confidence in the risk assessment process and demonstrate through their decisions that they are using it. In particular, statistics should show that offenders assessed as low risk should have lower rates of being sentenced to prison, have shorter sentences, have high rates of being paroled and receive lower levels of supervision. High-risk offenders should show just the opposite trends.

The people who are being assessed for risk must also believe that the process is credible and will be used by decision makers. The process should also be transparent and not some mysterious process where the offender is unaware of what factors are being used and how each is scored. This is especially helpful for risk instruments that employ dynamic risk factors—items that can change based on the offender's social and economic situation (employment, residency, and family relations). By understanding these dynamic risk factors, the offender can take actions or seek support that will actually reduce the risk to public safety.

A Closer Look at the LSI-R

As the interest in risk assessment has grown, so too has the private industry engaged in developing and distributing these systems. Currently there are two major privately held risk assessment systems available to corrections. The most widely advertised system is the Level of Service Inventory—Revised (or LSI-R), which was first developed in Canada and has now been adopted by a number of U.S. correctional agencies. [4] LSI-R is owned and distributed by Multi Health Systems, Incorporated, which distributes a wide array of psychologically based assessment tools. [5] The other is the Correctional Offender Management Profiling for Alternative Sanctions (or Compas), owned and distributed by the Northpointe Institute for Public Management, Inc., which also offers a privately held prison and jail classification system. [6]

Few independent validation studies of these two systems appear in the literature. By "independent" I mean studies done by researchers who have no financial interest in the two companies. Because the LSI-R has been around longer and is more widespread than the Compas, there have been a few recent studies in Washington, Pennsylvania, and now Vermont. As will be shown below, these studies show that many of the individual factors used in the LSI-R scale are not predictive of re-offending behavior. [7]

Why is this so? The principal problem with the LSI-R is that it is difficult to achieve a sufficient level of inter-rater reliability on many of its items. 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)

The LSI-R scorer is expected to make either a dichotomous "yes" or "no" to 37 items and a likert scale rating of satisfactory, relatively satisfactory, relatively unsatisfactory or very unsatisfactory for the other 17 items. For example, one question in the family/marital domain requiring a level of satisfactory response is "dissatisfaction with marital or equivalent situation." The scorer is instructed to base this assessment on a review of the case file data and an interview with the subject. On the accommodation domain, one question requiring a yes/no response is "three or more address changes last year." Such questions and the associated response raise important questions about whether correctional staff (most of whom have little if any

training in psychometric testing) can correctly use this assessment.

Researchers in Washington State conducted one of the first independent validation studies of the LSI-R as it was being applied to released state prisoners. [8] The authors found that the LSI-R criminal history factors were strong predictors of recidivism and produced most of the predictive power for the instrument. Put differently, many of the numerous other LSI-R items do little to enhance the LSI-R predictive attributes. These findings led the researchers to recommend that some of the LSI-R items be combined with other non-LSI-R factors, like current age and gender, to provide for a better risk instrument.

A recent study of the LSI-R as used by the Pennsylvania Parole Board and the Department of Corrections is instructive on difficulties associated with the LSI-R scoring process. [9] In particular, it provides the results of an inter-rater reliability study—a study that should be done for any risk assessment system. The Pennsylvania Parole Board was using the LSI-R scores to determine the suitability for release from prison. However, there had been no attempt to validate the LSI-R on Pennsylvania prisoners, which no doubt are somewhat different from the Canadian prisoners on whom the LSI-R had been developed. Further, the concept of using LSI-R for parole release considerations suggests a serious misapplication of the LSI-R, since many of the items have to do with the prisoner's life prior to incarceration. For example, how does one assess whether a prisoner who has been incarcerated for several years has "some criminal acquaintances" or few anti-criminal friends"? Given that many months or years may have passed since the offender was living in the community, the problems of accurate recall and the relevance of the questions for prisoners are rather obvious.

But even with these issues, one must also determine if the assessors are able to produce reliable scoring results. To this end, several reliability tests were conducted. The basic test is relatively straightforward and easy to do. A sample of 120 cases was selected for the test. Within two weeks, two staff were required to independently score the sampled cases and determine the appropriate score for each case. The results are shown in [Table 1](). The table contains only the 16 items that reached the 80 percent level of inter-rater agreement. The other 38 items had scores in the 60-70 percent range. If we use the more generous criteria of risk level, the level of disparity is reduced but remains at an unacceptable range, with a 29 percent disagreement on the risk level. It is also noteworthy that the items that have an acceptable reliability score are the more factual ones that are found in the more traditional risk assessment instruments.

With such a level of "noise" in the scoring process, it is not surprising that only a few of the LSI-R items were found to be associated with recidivism. A recidivism study of 1,006 prisoners who were scored on the LSI-R and had been released for at least one year was conducted. 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 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?

[Table 2]()

A regression analysis was done to see which of these 11 items had an independent effect on recidivism. This resulted in the following eight items being used: 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.

As shown in Table 3, only a small number of the 54 LSI-R scoring items are useful and most of them are not contributing to the risk assessment process. We also found that 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 that the condensed instrument does a better job of selecting those prisoners representing the most significant danger to public safety.

In Table 4, 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 re-offending.

In Vermont a similar set of findings were noted. The study was similar to the Pennsylvania one where the Parole Board was desirous of adopting a risk instrument to guide their decision-making process. Although a formal reliability was not completed, a validation study was made on the LSI-R and other factors believed to associate with risk to public safety. Two measures of recidivism rates were tested (return to prison and a new conviction), and became the basis for determining which items should be included in the final risk assessment instrument. The first step was to conduct a bi-variate statistical analysis to determine which items had a simple association with the three measures of recidivism.

The original validation study was based on 2,533 sentenced prisoners who were released in 2002. Of this number only 644 had completed LSI-R scoring results. As was found in the Washington state and Pennsylvania studies, only a relatively small number (13) of the LSI-R 54 factors are consistent and strong predictors of recidivism (items 4, 8, 9, 11, 13, 14, 16, 17, 31, 34, 39, 40 and 50). And another set of variables that are not part of the LSI-R was found to be associated with recidivism rates. These included current age, marital status, education level, measures of institutional conduct, and completion of certain programs while incarcerated (see Table 5).

Linking Risk to Punishment and Treatment?

The above studies show that risk assessment is doable but that it need not be complicated or expensive. Before adopting a particular system, an agency needs to rigorously assess what model it can afford and administer in a professional and accurate manner. If the wrong decisions are made in terms of what model to buy you may end up with little if any enhancements to your ability to assess risk.

I want to close on another matter that seems to be receiving little attention; namely, the requirement to administer or provide the proper "intervention" that is consistent with risk. The major assumption in evidence-based policy is that prisoners, probationers and parolees are to be " serviced" and punished relative to their risk. But reaching this standard can fail for two reasons. First, the assessed risk level becomes moot if there are no high quality programs or interventions to assign the "client" to once the assessment has been completed. For example, in Vermont only 14 percent of the released prisoner sample had completed an educational, substance abuse, or sex treatment program while incarcerated, even though 31 percent of the sample were assessed as high risk.

On the other end of the spectrum, we need to recognize that a very large proportion of the prison, probation, and parole populations is low risk; these offenders are being punished and even treated beyond their threat to public safety. It's like a hospital that decides to provide intensive care for patients who have a cold—the treatment is not only unnecessary but expensive.

It would be helpful for those in the risk assessment business to start advocating a more reasonable level of intervention that matches the risk they have so carefully calibrated.

back to top

Endnotes

Table 1: LSI-R Reliability Scores at the 80% Level

Variable	% Agreement
1. Any prior convictions?	96%
2. Two or more prior convictions?	93%
3. Three or more convictions?	93%
4. Three or more present offenses?	81%
6. Ever incarcerated upon conviction?	95%
7. Escape history from a correctional facility?	81%
8. Ever punished for institutional misconduct?	87%
9. Charge/probation/parole suspended during prior community supervision?	91%
10. Official record of assault/violence?	86%
11. Currently employed?	86%
12. Less than regular grade 10?	85%
13. Less than regular grade 12?	88%
14. Three or more address changes last year?	82%
15. Drug problem, ever?	88%
16. Moderate interference?	84%
17. Severe interference, active psychosis?	93%
18. Mental health treatment, past?	87%
19. Mental health treatment, present?	89%

	2nd Interview			
1st Interview	**Low**	**Medium**	**High**	**Total***
Low (0 through 15)	3	2	0	5
medium (16 through 22)	3	14	11	28
High (23 and above)	1	17	67	85
Total	7	33	78	118

Table 2: Cross-tabulation of the First and Second LSI-R Interviews of the Reliability Sample

*Note: Two cases were not scored a second time and were excluded from this analysis.

Point Distribution		Total		Recidivated
		N	%	%
TOTAL		948	100.0%	53.4%
Low	0	17	1.8%	17.6%
	1	17	1.8%	29.4%
	2	58	6.1%	22.4%
	3	54	5.7%	33.3%
	Total Low Risk	146	15.4%	33.6%
Moderate	4	119	12.6%	52.9%
	5	115	12.1%	47.8%
	6	133	14.0%	51.9%
	7	160	16.9%	56.3%
	8	87	9.2%	58.6%
	Total Moderate Risk	614	64.8%	53.4%
High	9	115	12.1%	70.4%
	10	23	2.4%	73.9%
	11	43	4.5%	60.5%
	12	1	0.1%	0.0%
	13	6	0.6%	83.3%
	Total High Risk	188	19.8%	68.6%

Table 3: Score using Select LSI-R Items by Failure Variable
Pennsylvania Board of Probation and Parole

Table 4: Score using Select LSI-R and Demographic Items by Failure

Pennsylvania Board of Probation and Parole

Point Distribution		Total		Recidivated
		N	%	%
TOTAL		848	100.0%	53.2%
Lowest	1	1	0.1%	0.0%
	2	7	0.8%	0.0%
	3	23	2.7%	21.7%
	Total Lowewt Risk	31	3.7%	16.1%
Low	4	20	2.4%	45.0%
	5	43	5.1%	37.2%
	6	64	7.5%	37.5%
	Total Low Risk	127	15.0%	38.6%
Moderate	7	89	10.5%	51.7%
	8	91	10.7%	46.2%
	9	115	13.6%	52.2%
	10	104	12.3%	57.7%
	11	94	11.1%	56.4%
	Total Moderate Risk	493	58.1%	52.9%
High	12	92	10.8%	65.0%
	13	41	4.8%	78.0%
	14	29	3.4%	89.7%
	15	19	2.2%	52.6%
	16	12	1.4%	66.7%
	17	4	0.5%	75.0%
	Total High Risk	197	23.2%	69.0%

Table 5: Vermont Parole Board Final Risk Assessment Simulation Score by Item and Overall Risk Level

Static Item	N-644	%	% Reconvicted
1. Age at First Arrest*			
16 years or older	458	71%	45%
Under 16 years	186	29%	60%
2. Prior Charges/Suspensions under Supervision*			
No	169	26%	36%
Yes	475	74%	54%
3. Crime Seriousness			
Level 1,2,4,5 and 10	334	52%	37%
Level 3,6,7,8,9 and 11	310	48%	61%
4. Drug/Alcohol Abuse*			
None	388	60%	42%
Behavior problems/Positive test	256	40%	60%
5. Prior Convictions*			
None	387	60%	45%
One	109	17%	49%
Two or more	148	23%	60%
6. Criminal Acquaintances at Admission*			
No	43	7%	26%
Yes	601	93%	51%
7. Employed at least 12 months prior to Admission*			
Yes	221	34%	41%
No	423	66%	53%
Dynamic Item	**N-644**	**%**	**% Reconvicted**
8. Current Age			
50 and above	46	7%	20%
40-49	133	21%	33%
24-39	325	50%	55%
Under 23	140	22%	61%

9. Most Severe Disciplinary report				
	None in the past 12 months	450	70%	44%
	major A or major b	194	30%	59%
10. Completed education/Substance Abuse program				
	Yes	87	14%	45%
	No	557	86%	50%
11. Current Custody Level				
	Minimum	0	0%	N/A
	Medium	0	0%	N/A
	Else	0	0%	N/A
12. Current marital Status				
	Married, Divorced	204	32%	43%
	Single	440	68%	52%
Scored Risk Level				
	Low	151	23%	26%
	Moderate	293	45%	49%
	High	200	31%	67%

*Denotes LSI-r factors

representatives in the sample for some combinations of gender, race/ethnicity, court type, and age.

[6] This includes those with negative scores.

[7] We include the square of age as a factor in the logistic regression because age has a curvilinear relationship with rearrest. Little (n.d.) used a similar analytic approach in her evaluation of the SDRRC.

How Much Risk Can We Take? The Misuse of Risk Assessment in Corrections

[1] See http://bmj.bmjjournals.com/cgi/content/full/312/7023/71. Also see http://www.ahrq.gov/clinic/epc/ for a listing of the growing number of evidence-based practice medical and mental health centers in the U.S. and Canada.

[2] Lowenkamp, Christopher T. and Edward J. Latessa. 2005. "Evaluation of Ohio's CCA Funded Programs. Final Report." Cincinnati, OH: University of Cincinnati, Center for Criminal Justice.

[3] Austin, James. 2006. "What Should We Expect From Parole?" American Probation and Parole.

[4] For a recent summary of the validity of the LSIR and its history see Girard, Lina and J. Stephen Wormith, 2004. "The Predictive Validity of the Level of Service Inventory- Ontario Revision on General and Violent Recidivism among Various Offender Groups" *Criminal Justice and Behavior*, Vol. 31, No.2:150-181 and Violent Recidivism among Various Offender Groups" *Criminal Justice and Behavior*, Vol. 31, No.2:150-181.

[5] For more information on MHS, Inc see their website at http://www.mhs.com/index.htm.

[6] For more information about Northpointe see their website at http://www.northpointeinc.com/contact.htm.

[7] Washington State Institute for Public Policy. *Washington's Offender Accountability Act: An Analysis of the Department of Corrections' Risk Assessment.* December 2003. Olympia, Washington; James Austin, Dana Coleman, Kelly Dedel-Johnson, and Johnette Payton. 2003. *Reliability and Validity of the LSI-R Risk Assessment Instrument* . Washington, DC: The Institute on Crime, Justice and Corrections at the George Washington University; and James Austin, 2006. *Vermont Parole Board Risk Based Guidelines, Technical Assistance Report #2.* Washington, DC: National Institute of Corrections.

[8] Washington State Public Policy Institute. 2003. p.4

[9] Austin, James, Dana Coleman, and Kelly Johnson. 2002. "Reliability and Validity of the LSI-R for the Pennsylvania Board of Parole and Probation." Washington, DC: The Institute on Crime, Justice and Corrections, The George Washington University.

From Theory to Practice: The Lifecycle Document for the Results-Based Management Framework for the Federal Probation and Pretrial Services System

[1] The terms "outcome-based" and "results-based" are used interchangeably.

[2] The term "evidence-based practice" implies that 1) there is a definable outcome(s); 2) it is measurable; and 3) it is defined according to practical realities (e.g. recidivism).