

# **Recidivism Risk Prediction and Prevention Assessment in Utah:**

**An Implementation Evaluation of the LSI-R as a Recidivism  
Risk Assessment Tool in Utah**



THE UNIVERSITY OF UTAH

---

*Utah Criminal Justice Center*

COLLEGE OF SOCIAL WORK  
COLLEGE OF SOCIAL & BEHAVIORAL SCIENCES  
UTAH COMMISSION ON CRIMINAL & JUVENILE JUSTICE  
S.J. QUINNEY COLLEGE OF LAW



**Recidivism Risk Prediction and Prevention Assessment in Utah:  
An Implementation Evaluation of the LSI-R as a Recidivism Risk Assessment Tool  
in Utah**

**Kort Prince, Ph.D.  
Robert P. Butters, Ph.D.**

**December 2013**

**Utah Criminal Justice Center, University of Utah**



## Table of Contents

<b>Acknowledgments</b> .....	ii
<b>Executive Summary</b> .....	iii
<b>Recidivism Risk Assessment</b> .....	1
Severity of the Issue.....	1
Risk and Need Assessment Background.....	1
Predicting Recidivism Risk.....	1
Assessing Need to Inform Treatment.....	2
Risk Prediction and Need Instruments.....	3
Third-Generation Instruments.....	4
Fourth-Generation Instruments.....	8
Summary of Instruments.....	10
Risk, Need and Responsivity Assessment in Utah.....	11
LSI-R in Practice.....	11
The Future of Risk, Need, Responsivity Assessment in Utah: The LS/CMI.....	11
<b>Implementation Evaluation of the LSI-R in Utah (within UDC)</b> .....	13
Project Background.....	13
Addressing Data Quality as a Potential Cause of Poor Predictive Validity.....	14
Invalid Assessments (Missing Items).....	14
Discrepant Total Score (Total Scores that are not the Sum of Item-Level Responses).....	15
Prohibited Combinations of Item-Level Responses.....	18
Addressing Administrative Issues as a Potential Cause of Poor Predictive Validity.....	22
Inadequate Training.....	22
Inherent Difficulty of the LSI-R Assessment.....	23
Ambiguities on the LSI-R Response Form.....	24
<b>Discussion</b> .....	26
<b>References</b> .....	27

## **Acknowledgments**

Funding for this project was provided by the state of Utah's Commission on Criminal & Juvenile Justice (CCJJ). Support for project planning was provided by Ben Peterson, PhD, Director of Research & Data at CCJJ. Data for the project were provided by the Utah Department of Corrections (UDC) with special assistance by Julie Christenson, Director of Research and Planning. Please direct concerns, comments or other correspondence to either Rob Butters (UCJC Director; [rob.butters@socwk.utah.edu](mailto:rob.butters@socwk.utah.edu)) or Kort Prince (UCJC Senior Research Analyst; [kort.prince@socwk.utah.edu](mailto:kort.prince@socwk.utah.edu)).

## Executive Summary

This document is divided into two major sections. The first covers recidivism risk and needs assessment, while the second describes an implementation evaluation of the Level of Service Inventory-Revised (LSI-R), which has been used in Utah since 2000 as a risk and needs assessment.

Section one describes the need for accurate, valid risk and needs assessments, and provides a literature review of the most well-known, validated and commonly used instruments as well as a review of tools that are relatively new but show promise. In the literature, the LSI-R remains the dominant risk and needs assessment tool, but other instruments in the third- and fourth-generation of risk and needs assessment tools (see below for generation definitions) have become increasingly more common. The LSI-R has, in some cases, been replaced by its successor, the Level of Service/Case Management Inventory (LS/CMI). Other instruments show similar validity, but also have the added benefit of being easier to use (such as self-administered assessments).

The second major section was originally intended to validate the LSI-R in Utah's population and within demographic subgroups (i.e., race, ethnicity and sex). It was also intended that the research would examine whether Utah's potential adoption of the newer and broader LS/CMI, would improve, reduce or simply maintain the current tool's ability to predict recidivism. While the issue of validation within Utah will remain an important area of future research, preliminary analyses of LSI-R data provided to UCJC revealed problems that precluded a fair and accurate assessment of the validity of either instrument. Problems included both data quality issues (i.e., invalid tests due to too many missing items, total scores that did not match the sum of item-level responses, prohibited combinations of item-level responses) and administrative issues (i.e., inadequate training of administrators, inherent difficulties in using the LSI-R, and ambiguities on the LSI-R response form).

Rather than using inaccurate data in attempting to validate the instrument(s) for use as recidivism risk prediction tool(s) in Utah, the present research necessarily altered focus to examine the extent of the problems. The present research also discusses evidence suggesting that the LSI-R tool, and its inherent difficulties in administration, may have contributed to some of the observed problems. Changes that have already been made and changes that could be adopted to improve the accuracy and fidelity of LSI-R assessments, are discussed.

# **Recidivism Risk Assessment**

## **Severity of the Issue**

The issue of criminal risk and needs assessment addresses the qualitatively different concerns of both predicting and preventing recidivism (through treatment). Both issues are of considerable interest to individuals and society in general, as both individuals and society incur tremendous costs when crimes occur. Crimes have both tangible and intangible costs (Heaton, 2010). The tangible burdens are economic, and are more clearly visible. These include, as examples, adjudication costs, property loss, medical treatment costs and prevention expenses. Intangible costs are more difficult to measure, but can be considerably more crippling, especially at the individual level. The psychological effects of crime, which are often long-lasting, are a primary example of intangible costs, but these also include loss of quality of life and, at the society level, loss of community space in order to prevent crime.

On the lower end of the cost spectrum, motor-vehicle theft, burglary and larceny crimes range from estimated costs of \$2,000 to \$13,000 per crime, while violent crimes range from per event costs of \$67,000 for robbery to \$217,000 for rape and \$8.6-million for homicide (see Heaton, 2010 for a review). Though violent crimes typically have largely intangible costs, they are also the most expensive by either metric. Miller, Cohen and Wiersema (1996) estimated victims bear 77% of the cost associated with violent crimes, while taxpayers bear 14% of these costs and employers bear the remaining costs. At the societal level, taxpayers pay approximately \$1.2-million per murder and \$30,000 per rape.

## **Risk and Need Assessment Background**

In addition to the intangible psychological costs of crime, a great deal of tangible financial resources are allocated to legal proceedings, incarceration and reform efforts of criminals; thus, predicting risk of reoffending and successfully treating offenders with respect to needs are both of paramount interest to the criminal justice community. Despite the importance, both dimensions of criminal assessment met with an inauspicious start.

## **Predicting Recidivism Risk**

The issue of predicting recidivism has nearly a century old history (Andrews & Bonta, 2010), and is generally divided into four generations in criminal justice literature. The first-generation of prediction involved clinical expertise. Typically, a clinical profession would interview an offender in a relatively unstructured format, and might consult patient files and psychological batteries. The primary determination in these first-generation assessments was the clinical professional's "gut feeling". While this approach has considerable face validity, it lacks actual predictive validity (especially when compared to statistical methods). For example, Meehl (1954) summarized 20 studies comparing statistical and clinical methods and determined statistical methods to be more accurate in 19 of the 20 studies. Similarly, research by Steadman and Coccozza (1974) indicated that, by clinical judgment, four non-recidivating psychiatric patients would have been detained for every one that actually committed a future crime. This outcome highlights the issue of criminal justice and ethics when predicting future recidivism.

Though defenders of clinical judgment exist (Holt, 1958), they have paled in frequency to research and literature supporting an actuarial, statistical approach to recidivism prediction (Ægisdóttir, Spengler, Maugherman, Anderson, Cook et al., 2006; Andrews & Bonta, 2010; Grove, Zald, Lebow, Snitz & Nelson, 2000; Hilton, Harris & Rice, 2006). The actuarial approach is often cited as beginning with Burgess' (1928) study of over 3,000 parolees and the 21 factors that differentiated recidivists from non-recidivists. Parolees were assigned a point for every risk factor that was present; the highest risk score



correctly predicted recidivism 76% of the time. By contrast, clinical methods have often been shown to be no better than a coin flip (i.e., random chance) in predicting recidivism, and are sometimes worse (Monahan, 1981). This and subsequent actuarial methods comprise the second-generation risk assessment tools. While they are evidence-based, they are differentiated from third-generation tools (discussed below) by the fact that they are not theory driven, and consist almost entirely of static (historical) risk factors. Accordingly, the second-generation does not give credit to offender reform or treatment programs in their efficacy in reducing future crime. If past behavior were the only relevant predictor of future behavior, 100% of past offenders would recidivate. Because 100% recidivism is empirically false, room for improvement existed in recidivism prediction.

Third-generation assessment tools address the criminogenic needs of offenders in order to help alter future behavior (through treatment) and subsequently reduce risk; that is, they provide a targeted list of factors that might reduce future recidivism at the individual level. Unlike second-generation tools, they assume behavior is malleable and that offenders can be reformed through treatment and environmental alterations (discussed below). Relative to second-generation instruments, third-generation instruments were also differentiated by empirically-based, theory-driven items and factors. Items and scales in this generation of measures were grounded in contemporary understanding of criminality, and naturally flow into the next section, assessing need.

Fourth-generation tools emphasize the connection between risk/need assessment and case management. Case management components are beyond the risk, need, responsivity focus of this report, but examples of fourth-generation tools, and their general use, are discussed in the Risk Prediction and Needs Instruments section below because they often contain sections relevant to risk/need assessment. For example, the Level of Service Case Management Inventory's (LS/CMI) first section is a modification of the third-generation Level of Service Inventory-Revised (LSI-R). As such, section one of the LS/CMI is a useful risk/need assessment tool, and is discussed in the relevant section.

### **Assessing Need to Inform Treatment**

Similar to the issue of assessing risk, the assessment of need has also undergone an evolution after an inauspicious beginning. Andrews and Bonta (1994) delineated two categories of risk factors associated with recidivism: static and dynamic. Static factors are immutable, historic facts, such as age, gender or number of prior convictions. Dynamic factors (also known as criminogenic factors), by contrast, are considered alterable and are therefore subject to intervention. These factors include cognitions and attitudes about crime and anti-social behavior. Static factors have long been considered valid predictors of criminal behavior, but dynamic factors have met with considerably more skepticism. It was not until the Wisconsin classification system was adopted (Baird, Heinz & Bemus, 1979) that criminogenic needs became more commonly assessed; moreover, they have often been historically disregarded as predictive of criminal outcomes.

A refusal to accept the role of dynamic factors in prediction of recidivism has important consequences for the criminal justice system. Because static factors cannot be altered, to the extent that a system of criminal justice assumes these factors are the sole determinants of future behavior, both criminal reform and assessing criminal need become issues of little importance.

The impact of criminal rehabilitation and treatment programs has remained a subject of great debate since Martinson's (1974) famous article that would come to be colloquially known as "Nothing Works." Martinson's meta-analysis of over 20-years of prior rehabilitation efforts to that time concluded that "with few and isolated exceptions, the rehabilitative efforts that have been reported so far have had no appreciable effect on recidivism" (pg. 25). He further concluded that rehabilitative efforts "cannot overcome, or even appreciably reduce, the powerful tendencies of offenders to continue in criminal

behavior” (pg. 49). Conclusions such as these served to foster the belief that the best predictor of future criminal behavior was past criminal behavior, and subsequently debased efforts aimed at criminal reform.

Research conducted in the 1980s and 1990s, however, revealed a more optimistic perspective of criminal rehabilitation through new, more advanced, research methodology techniques (Wilson, Bouffard & Mackenzie, 2005). A key finding of the research conducted in the decades after Martinson’s “Nothing Works” article revealed that many previous programs failed largely because they were based on intuition, not research. A summary of the methods that failed in rehabilitation research can be found in Little, Robinson & Burnette (1992). The research highlighted in this article has documented several well-intentioned components of criminal rehabilitation programs that have, unfortunately, either not impacted recidivism or have actually increased its probability. Missing components in these rehabilitation efforts, as they were typically implemented, were: (1) a therapeutic component that aimed to alter the criminal’s decision-making and thought processes, and (2) interventions that altered the criminal’s environment (including antisocial companions).

Following this resurgence of successful criminal reform programs, the assessment of dynamic criminal needs gained importance. In 1996, Gendreau, Little & Goggin conducted a meta-analysis of 131 studies and 1,141 variables related to recidivism. The study provided a groundbreaking, direct comparison of static versus dynamic predictors of recidivism. While static factors were significant predictors of recidivism averaged across the 131 studies, dynamic factors were statistically more important. The authors concluded: “The time is long past when those risk factors that are dynamic in nature can be cavalierly ignored” (pg. 588).

## **Risk Prediction and Need Instruments**

Findings from the historical review above highlight the need for assessment tools that address both risk and need. Accordingly, only instruments that extend beyond second-generation actuarial risk assessment tools are discussed below. The Level of Service Inventory-Revised (LSI-R) is the currently adopted risk tool in the state of Utah; however, a review and utility comparison of other instruments is provided, as consideration of other instruments is warranted in order to ensure Utah is using the best current practices.

Some of the most cited competing instruments (with moderate to strong effect sizes in meta-analytic reviews) in the third<sup>1</sup>- and fourth-generations include the Wisconsin Risk Needs assessment (commonly referred to as the Wisconsin; Baird, Heinz & Bemus, 1979), the Historical, Clinical, and Risk Management Violence Risk Assessment Scheme (HCR-20; Webster, Douglas, Eaves & Hart, 1997), the Self-Appraisal Questionnaire (SAQ; Loza, 1996), the Psychopathy Checklist-Revised (PCL-R; Hare, 2003), the Inventory of Offender Risk, Needs and Strengths (IORNS; Miller, 2006), the Level of Service Inventory-Revised (LSI-R; Andrews & Bonta, 1995), the Correctional Assessment and Intervention System (CAIS; Justice System Assessment & Training, 2010), the Correctional Offender Management Profiling for Alternative Sanctions (COMPAS; Brennan & Oliver, 2000), and the Level of Service/Case Management Inventory (LS/CMI; Andrews, Bonta & Wormith, 2004)<sup>2</sup>. While some of these instruments

---

<sup>1</sup> It is important to note that some literature refers to certain instruments within this group as second-generation while others refer to them as third. The debate centers on the issue that some of these instruments contain mostly static, non-malleable items. Because the focus of this report is assessment tools in general, however, and providing the best recommendations for adoption of an assessment tool in Utah, these tools are reviewed regardless of the theoretical debate of the actual generation in which they belong (unless otherwise noted) as long as they are purported to assess dynamic factors.

<sup>2</sup> Other measures are often used in recidivism research, and to validate and compare the predictive ability of third-generation tools. The Violence Risk Appraisal Guide (VRAG; Harris, Rice, & Quinsey, 1993), for example, is a second-generation (risk only), 12-item actuarial tool consisting of only static items. Its authors indicated no

(e.g., the PCL-R) are oriented toward clinical diagnosis or specific types of criminal recidivism (e.g., violent), they are often utilized as general measures of recidivism as well (Campbell, French & Gendreau, 2009; Gendreau, Little & Goggin, 1996; Loza & Loza-Fanous, 2001), and are therefore reviewed.

### **Third-Generation Instruments**

#### *The Wisconsin*

Of the aforementioned recidivism risk assessment tools, the Wisconsin is the oldest, and has also received the most criticism, particularly with respect to its predictive validity evidence in assessing criminogenic need<sup>3</sup> (Andrews & Bonta, 2010). The Wisconsin scores offenders on two dimensions. The first is conceptualized as a risk dimension, and categorizes offenders for level of supervision. Risk factors include frequent address changes, self-report of substance abuse, percentage of time unemployed in the last year and other factors. The second dimension of the Wisconsin measures needs, scoring offenders on academic/vocational skills, financial issues, companions, mental health, emotional stability, etc. The instrument contains 23 items in all, 11 risk items and 12 needs items.

Some reviews of the Wisconsin have produced favorable outcomes. Gendreau, Little & Goggin's (1996) meta-analytic review found that the Wisconsin was a significant predictor of general recidivism, and was statistically equivalent to the LSI-R and PCL-R (both reviewed later in this report). More recent work has been less favorable, however. Henderson and Miller (2011) reviewed several validation studies using the Wisconsin as a predictive tool. Their review indicated that the instrument was commonly no better than chance (i.e., a coin flip) when predicting recidivism or probation revocation and that it was often worse (Connelly, 2003; Harris, 1994). Moreover, research evidence from their review suggests far too many individuals were classified at the highest risk level, largely due to an extreme weight assigned to assault related crimes. The decision to weight assault crimes more heavily was not statistically derived, but, rather, was derived from concern over liability if assaults reoccurred. Over classifying offenders as high in recidivism risk is disconcerting for several reasons: (1) continued incarceration of objectively low reoffending individuals creates a considerable financial burden, (2) misclassification of risk creates a subsequent misallocation of supervision resources, and (3) over classifying offenders as high risk, and thereby detaining them for greater lengths of time, devalues criminal reform. Attempts to reweight the items from the Wisconsin have not improved its predictive validity in general (Eisenberg, Bryl & Fabelo, 2009).

#### *Historical, Clinical and Risk Management Violence Risk Assessment Scheme (HCR-20)*

The HCR-20 is a hybrid between second and third-generation instruments, but focuses more weight on static, historical factors. As the name suggests, it measures three factors (historical, clinical and risk) in the prediction of violent recidivism using 20-items. The HCR-20 is often cited as a second-generation instrument (Douglas, Ogloff, Nicholls & Grant, 1999), but its clinical and risk management components assess dynamic aspects of violent recidivism risk and prevention need.

Research evidence has suggested the HCR-20 is a moderate to strong predictor of violent behavior (Belfrage & Douglas, 2002; Dernevik, Grann & Johansson, 2002), but, despite its promise in this domain,

---

dynamic variables predicted violent recidivism in their validation work. Because it is an older tool, it fails to consider advances in dynamic predictors of recidivism risk, and because other third-generation tools reveal better prediction of recidivism, the VRAG is not discussed in further detail. As of this report, items from the scale can be reviewed at: <http://www.fotres.ch/index.cfm?&content=9010&spr=en>

<sup>3</sup> The Wisconsin was progressive for the time in which it was developed in that it was the first widely adopted instrument that incorporated dynamic factors.

it is not a measure of general recidivism (it is reviewed here only because of its frequent mention in recidivism research in general), and was not developed by the authors for such a purpose. For example, the historical items ask about previous violence and age at first violent incident, but include no items assessing any other types of crimes. It was also validated, and primarily designed for use with, criminal populations suffering from moderate to severe mental illness, including personality disorders such as antisocial personality (e.g., Douglas & Webster, 1999; Hart, 1998). Its utility in general offender populations has not been well vetted (Manchak, Skeem & Douglas, 2007). Details of its psychometric properties are not reviewed in this report because, despite its usefulness in predicting violent recidivism, the HCR-20 is not a plausible tool to replace the LSI-R in Utah as a general recidivism prediction or needs assessment tool.

### *The Self-Appraisal Questionnaire (SAQ)*

The SAQ (Loza, 1996) is a self-report based risk and need measure. Consisting of 72 items and eight subscales (if the anger subscale is included), the SAQ is touted by its creators as more economical than alternatives. It can be administered in a group setting, requires no clinical judgments, consists of all true/false items and takes approximately 15 to 20 minutes to complete. Over half of the items are designed to assess dynamic factors. Subscales include: (1) criminal tendencies (antisocial attitudes and behaviors), (2) anti-social personality problems (it provides a Diagnostic and Statistical Manual comparable diagnosis of anti-social personality disorder), (3) conduct problems (measures childhood behavior problems), (4) criminal history, (5) alcohol and drug abuse, (6) anti-social associates, and (7) a validity scale (which measures truthfulness to responses within the other scales). An eighth subscale (the anger subscale) is not included in the overall score or in prediction of recidivism because it has not been shown to be a strong predictor. Without this subscale, the instrument's other subscales provide 67 items. The SAQ has been shown to have strong discriminant and predictive validity at 2, 5 and 9-year follow-up (Loza, MacTavish & Loza-Fanous, 2007).

A review of literature in the area of third-generation risk/need instruments suggests the SAQ is a viable challenger to the LSI-R (discussed below). Loza and Loza-Fanous (2001) interviewed a small sample of Canadian offenders prior to release and administered five recidivism risk instruments: the SAQ, the LSI-R, the PCL-R, the VRAG, and the General Statistical Information on Recidivism (GSIR; Nuffield, 1982; not discussed in detail in this document because it is not a third-generation risk and need tool). Offenders were tracked for re-offense (both violent and non-violent) for up to two years. Results were ambiguous due to the low sample size, but indicated the SAQ was a significant predictor and was at least as effective at predicting both types of recidivism when compared to all other measures. Though not statistically different from other measures, it produced the highest correlation coefficient with respect to general recidivism, and was tied with the GSIR, both producing the highest correlations in prediction of violent recidivism. The LSI-R performed slightly less well among these statistically equivalent predictors.

Loza, MacTavish and Loza-Fanous (2007) completed a nine year follow-up study of 657 Canadian sentenced male offenders. All subscales, as well as the total score, were revealed to be significant predictors of recidivism. Sensitivity was 59% for non-violent recidivism and 70% for violent. Specificity was 74% for non-violent and 62% for violent recidivism. Researchers provided results for Relative Improvement Over Chance (RIOC), an atypical but conceptually similar statistic to the more typical Area Under the Curve (AUC). These results indicated improvement over chance based prediction of recidivism of 38% for non-violent recidivism and 49% for violent recidivism.

### *The Psychopathy Checklist – Revised (PCL-R)*

The PCL-R (Hare, 1991, 2003) is a 20-item, clinical rating scale (requiring a clinically trained professional for administration) designed to assess psychopathy. It contains two factors and four facets

(factors 1a and 1b, and factors 2a and 2b). Facet one assesses interpersonal issues; facet two assesses affective issues; facet three assesses lifestyle issues, and facet four assesses anti-social problems. Interestingly, the PCL-R is rarely discussed as belonging to any specific generation of risk assessment tool (though Andrews, Bonta and Wormith (2006) consider it a second-generation tool), owing to the fact that (unlike the other measures) it was not specifically designed as a risk prediction tool, but rather as a diagnostic tool for psychopathy. Accordingly, its focus is on the measurement of personality traits that correlate with general recidivism, including psychopathy. The definition of psychopathy is not the same in Hare's PCL-R as the often interchangeable term antisocial personality disorder (APD) from the DSM-IV-TR. Also, the PCL-R is more general than a simple measure of APD; it also measure narcissistic and histrionic personality disorder. Some research suggests that the PCL-R definition is more useful in clinical and correctional settings because it includes factors that are considered amenable to treatment; however, others suggest that, because it measures relatively stable personality traits, it is not a measure of dynamic factors in the traditional sense (as there is debate as to whether the axis-2 traits it measures are amenable to treatment; Wormith, Olver, Stevenson & Girard, 2007).

As a clinical diagnostic tool, it is more general than a recidivism risk and need prevention tool, but shows promising validity as a predictive measure nonetheless. Gendreau, Little & Goggin's (1996) meta-analytic review found that the PCL-R was a significant predictor of general recidivism, and was statistically equivalent to the LSI-R and Wisconsin. The study revealed an average correlation of .28 between the PCL-R and general recidivism. In the same year, Salekin, Rodgers and Sewell conducted a meta-analysis of 18 studies using the PCL and/or PCL-R and reported a significant average correlation of .27 between the PCL-R and recidivism, and also reported moderate to strong effect sizes.

Unsurprisingly, because the PCL-R is not a direct measure of recidivism risk and prevention need, other studies have shown it to be less predictive than alternatives. The review cited above by Loza and Loza-Fanous (2001) indicated the PCL-R was non-significantly outperformed by the SAQ, and a meta-analytic study by Gendreau, Goggin and Smith (2002) showed the LSI-R outperformed the PCL-R ( $r=.39$  compared to  $r=.25$ ) when the outcome was general recidivism. The same study showed the LSI-R and PCL-R were equivalent on a measure of violent recidivism. Because the PCL-R is not a tool designed specifically for recidivism prediction, reviews have largely concluded that it is better as a clinical diagnostic tool or occasionally as a tool for prediction of violent recidivism (Andrews & Bonta, 2010).

#### *Inventory of Offender Risk, Needs and Strengths (IORNS)*

The IORNS is a 130-item true/false self-report questionnaire assessing static and dynamic risk factors across nine factors/scales (Miller, 2006). The Static Risk Index (SRI) is considered both an index and a scale and measures static criminal history factors. The Dynamic Need Index (DNI) contains 79-items assessing seven dimensions (scales) of dynamic needs: Criminal Orientation, Psychopathy, Intra/Interpersonal Problems, Alcohol and Drugs Problems, Aggression, and Negative Social Influences. It also contains a protective or strength index (PSI) that assesses personal and environmental assets and resources (the last two scales of the instrument). This component makes the IORNS unique in the third generation of assessment tools. Only the fourth-generation LS/CMI (discussed below) possesses a similar asset or strength measurement component; however, and perhaps to the detriment of the LS/CMI, that component is not part of the scoring. For the IORNS, the overall score is created by combining the SRI and DNI and subtracting the PSI; the IORNS, therefore, is unique in considering protective factors in relation to risk factors. Because it is a self-report measure, and is subject to bias, it also contains two validity scales measuring whether the respondent was attempting to make a favorable impression or was inconsistent.

The IORNS is a relative newcomer to the RNR assessment literature, but it possesses several attributes that make it an appealing assessment option. Its self-report nature makes it easy to administer, and it

requires only a third-grade reading ability. Scoring can be performed without the need for expert training in psychology or related fields. However, little validation work has been performed using the instrument. The creators' own validation research revealed significant prediction of recidivism defined as rule violations, but correlations could not be found relating the subscales or the overall score to general misdemeanor or greater recidivism at specific follow-up periods. Correlational analysis comparing the overall scores of the IORNS to the well-documented LSI-R revealed a significant correlation (0.45), but the measure correlated more strongly with two measures of psychopathology, leading one to question whether it is an accurate measure of recidivism risk in the general offender population.

### *The Level of Service Inventory – Revised (LSI-R)*

The LSI-R is a third-generation, theory driven risk and need assessment tool (Andrews & Bonta, 1995) for the criminal offender population. It contains 54-items and 10 subscales: Criminal History, Education and Employment, Financial, Family/Marital, Accommodations, Leisure and Recreation, Companions, Alcohol and Drugs, Emotional and Personal, and Procriminal Attitude Orientation. Items and subscales were derived from a literature review of the factors theoretically related to criminal conduct. While the LSI-R does not require a clinical professional for scoring, training in use of the instrument is necessary, and the instrument does require an interview with offenders. Scores from the 10 subscales of the LSI-R create an overall score (summation method) and a risk classification into one of five categories (although the LSI-R recommends these cutpoints, jurisdictions can alter them for their population). The cutpoints reflect risk of one-year recidivism.

Three decades of research support the reliability and validity of the LSI-R (Andrews, 1982; Andrews, Bonta & Wormith, 2006). Andrews, Bonta and Wormith (2006) found average correlations between the LSI-R and general recidivism to be .36, while Gendreau, Goggin and Smith (2002) indicated a correlation of .39. With respect to violent recidivism, the same studies found respective correlations of .25 and .28. Comparisons with other instruments have indicated the LSI-R generally predicts recidivism as well or better than other instruments in prediction of both general and violent recidivism, including the Wisconsin and the PCL-R (Gendreau, Little & Goggin, 1996). The only review that showed promise for another measure relative to the LSI-R indicated that the SAQ measure was a non-significantly better predictor of recidivism (Loza & Loza-Fanous, 2001).

Similar to the SAQ, the majority of items on the LSI-R are dynamic in nature, and both are well-regarded in the research for assessing criminogenic need; however, the LSI-R has been the subject of considerably more research in the area of dynamic validity. Raynor, Kynch, Roberts & Merrington (2000), for example, found that, among low-risk offenders who were evaluated more poorly on the LSI-R at reassessment, recidivism rates were higher; conversely, initially high-risk offenders who scored more favorably at reassessment were less likely to recidivate. Other studies have also indicated that the dynamic items do, in fact, change at reassessment and are malleable as posited (Belfrage & Douglas, 2002).

The LSI-R stands out from most instruments for recidivism prediction in terms of both predictive utility and the sheer volume of supporting research. While other instruments show promise (e.g., the SAQ), without jurisdiction specific evidence questioning its validity, it is difficult to argue in favor of replacing the LSI-R as an assessment tool due to the considerable range of validation studies in which it has been involved. It has been shown to predict violent and general recidivism (Andrews, Bonta and Wormith, 2006), probationer recidivism (Andrews & Bonta, 2003), parolee recidivism (Kroner & Mills, 2001), recidivism after long-term incarceration (Manchak, Skeem & Douglas, 2007), and has shown promise with both male and female offenders (Vose, Lowenkamp, Smith & Cullen, 2009).

## Fourth-Generation Instruments

Third-generation risk and needs instruments were developed for purposes of allocating resources for supervision and intervention, and thereby address both risk and need. While these tools consider need and responsivity factors above second-generation tools, fourth-generation instruments emphasize the connection between assessment and case management, and identify areas within which intervention strategies need to be modified to maximize reduction in risk. They also address special responsivity factors that help maximize the benefits of treatment. Tools in this generation address the problem created by the fact that, although Risk, Need, Responsivity (RNR) inventories are administered, they are not always utilized to assist in offender reform. Examples of instruments in this class include the Correctional Offender Management Profile for Alternative Sanctions (COMPAS; Brennan & Oliver, 2000), the Correctional Assessment and Intervention System (CAIS; National Council on Crime and Delinquency, 2004), and the Level of Service/Case Management Inventory (LS/CMI<sup>4</sup>; Andrews, Bonta & Wormith, 2004).

Though important tools for correctional and case management professionals, the application of the fourth-generation tools within the state of Utah extends beyond the risk and needs focus of this report. However, tools within this generation contain sections or components relevant to RNR assessment. For example, the LS/CMI's Section I (presented below) is a tool similar to the LSI-R. For purposes of this report, only the sections/components of these fourth-generation tools relevant to RNR assessment are discussed. If future research seeks to evaluate case management needs, the adoption and evaluation of these fourth-generation RNR and case management tools should be considered.

### *Correctional Assessment and Intervention System (CAIS)*

The CAIS is a fourth-generation instrument that adds case management capabilities to the Wisconsin instrument reviewed earlier. Because of concerns regarding the usefulness of the Wisconsin as a predictive tool (discussed above), the CAIS is not discussed in further detail<sup>5</sup>.

### *Correctional Offender Management and Profiling Alternative Sanctions (COMPAS)*

The COMPAS was designed to assist decision-making in terms of offenders' placement, supervision and case management (Northpointe Institute for Public Management, n.d). As a fourth-generation risk assessment, the COMPAS is intended to function as a dynamic part of case planning and implementation. It is comprised of multiple subscales: criminal involvement, history of noncompliance, history of violence, current violence, criminal associates, substance abuse, financial problems, vocational or educational, criminal attitudes, family criminality, social environment, leisure, residential instability, criminal personality and social isolation. The needs assessment component classifies offenders as having low, medium or high need for services across the range of subscales.

---

<sup>4</sup> Because the LS/CMI includes the same risk items as the LSI-R, it can be considered equally valid as a predictor of general and violent recidivism. Relative to the LSI-R, it has already been established that the LS/CMI adds specific additional risk and need factors as well as special responsivity considerations. Relative to the LS/RNR, the LS/CMI provides a case management specific section requiring correctional staff to prioritize the criminogenic needs of the offender. They must also assist the offender by creating tangible targets for change and identifying the means by which the offender would reach these goals. The case management professional also keeps a record of offender progress toward concrete goals, noting improvement, deterioration or lack of change.

<sup>5</sup> Additional information on the instrument is available at: <http://www.nccdglobal.org/assessment/correctional-assessment-and-intervention-system-cais>

The COMPAS was designed to assess risk of violence recidivism, general recidivism, non-compliance and failure to appear. The COMPAS also includes modules for tracking decision-making, treatment, offender outcomes and organizational variables. Brennan, Dieterich and Ehret (2009) found that COMPAS had stronger predictive validity for person and felony offenses than for the category of any new offense, was equally accurate for both men and women and performed equally well for African American and white males. It was shown to have particularly high predictive utility with females committing person offenses, and White males committing either person or felony offenses (AUCs > .75). However, the AUC value for the total sample for commission of any re-offense (a more general measure of recidivism) was .68, below the threshold for instruments generally considered of acceptable predictive utility, and below the values reported in meta-analytic studies of the LSI-R and LS/CMI (discussed more below; see also Andrews, Bonta & Wormith, 2004)<sup>6</sup>. Within some demographic groups, or when considering specific crimes, the COMPAS occasionally reveals better predictive utility than other tools. Adoption of the COMPAS may, therefore, be of greater utility with specific subgroups of the criminal population.

### *The Level of Service/Case Management Inventory (LS/CMI)*

As a result of work beginning in 1994 and culminating in 2004, the LSI-R was expanded as an assessment tool referred to as the Level of Service/Case Management Inventory (LS/CMI; Andrews, Bonta & Wormith, 2004; 2008). This instrument, like the LSI-R, considers offender needs and the dynamic aspects of risk. While the LSI-R has one section consisting of 54-items, the LS/CMI has 11 sections. Items from the LSI-R were incorporated into section one of the LS/CMI; hence, for comparison purposes, and based on the goal of RNR assessment, only section one of the LS/CMI is discussed in detail here.

Section one of the LS/CMI is largely a restructuring of the LSI-R, but also reduces the number of items. For example, though typically explained in the literature in terms of its 10 subscales (listed above), the LSI-R does include items measuring what are known as the “big four” and the “central eight” risk and needs factors.<sup>7</sup> The LS/CMI section one reorganizes the items from the original LSI-R to correspond to the “central eight”, including restructuring financial, emotional/personal and accommodation factor items, removing other individual items that were part of the original 10 subscales, and adding an antisocial pattern factor that encompasses items from the financial and accommodation subscales of the LSI-R. To the LSI-R capabilities, the LS/CMI has also added a “strengths” notation section, allowing the user to remark on certain protective factors.

Because the LS/CMI section one is largely a reorganization and refinement of the LSI-R, validation evidence from the latter instrument largely applies to the former as well. However, some evidence suggests that the restructuring and reduction in items has actually improved the predictive ability of the instrument (Andrews, Bonta & Wormith, 2004). With respect to the “central eight”, the items from the LSI-R are simply reorganized into the new subscales of the LS/CMI section one.

They are not directly relevant to this research, but for purposes of comprehensiveness, it is worth noting that, in addition to including the “big four” and “central eight” risk and needs factors addressed by the LSI-R (and the LS/CMI section one), the additional sections of the LS/CMI add specialized risk and need related factors including sexual assault, weapon use, homelessness and victimization experiences, and responsivity factors such as cultural, gender and ethnic issues. It also contains sections on prison

---

<sup>6</sup> The interpretation of AUC values is described later in the results section of this document.

<sup>7</sup> The big four are the most predictive factors and are a subset of the larger “central eight” factors. Together, they represent the most powerful, consistently supported predictors of criminal behavior. They include: (1) history of anti-social behavior, (2) anti-social personality patterns, (3) anti-social cognition, (4) anti-social associates (these represent the “big four”), (5) family and/or marital issues, (6) school and/or work issues, (7) leisure and/or recreation issues, and (8) substance abuse (5 to 8 are also called the “moderate four”).



experience, client social, health and mental issues, and sections for case management planning and progress reporting.

The LS/CMI is also regarded as a gendered (i.e., gender-informed) tool. The LSI-R is considered a gender-neutral tool; that is, outcomes (including risk) are assumed to be gender invariant. While some research has supported the validity of the LSI-R with female offenders (Smith, Cullen & Latessa, 2009), other research has questioned its gender invariance and applicability to women (Holtfreter & Cupp, 2007). Regardless of the stance one takes on the issue of gender and LSI-R validity, the issue of its jurisdictional gender invariance is empirical. Cutpoints can be set differently if jurisdictional data indicate the tool is gendered. The LS/CMI, however, ostensibly eliminates this concern. According to the website for Multi-Health Systems (MHS; the host and authoring site for all Level of Service products<sup>8</sup>), the LS/CMI provides gender-based norms and does not require gender specific add-ins. Howell (n.d.) reports that the tool has been normed with over 20,000 women and that it provides “prediction of re-offending as effectively with women as with men” (pg. 4).

### **Summary of Instruments**

A thorough review of the literature on recidivism risk and prevention needs assessment supports the use of the LSI-R and the comparable LS/CMI section one as valid, empirically supported tools. While other instruments for the prediction of recidivism exist and also show promise, most are unviable as RNR tools because of an overdependence on or overrepresentation of static factors. Instruments in this group offer little information about how to intervene to alter future behavior, and they are also better-suited to the prediction of violent (rather than general) recidivism (e.g., the PCL-R or HCR-20). Other reviewed tools designed with dynamic factors and general recidivism in mind, show poorer predictive ability compared to the LSI-R or LS/CMI section one (e.g., the Wisconsin), though the empirical predictive ability of the Level of Service tools in Utah is examined later in this report. Some tools (e.g., the COMPAS) showed promise as instruments equally predictive when compared to the Level of Service instruments, but adopting an equally predictive tool that would require additional training but is similar in cost makes little fiscal sense.

Because no tools seemed to reliably outperform the Level of Service instruments, the decision to adopt a new tool would likely consider three factors: the new tool should be (1) equally predictive and valid, (2) have lower cost in training and adoption, and (3) have a lower overall cost per survey and for administration. Two reviewed tools, the IORNS and the SAQ, show some promise across these three factors. The IORNS is a relatively new tool and does not have enough validation research currently to justify its adoption, but it is a relatively low cost and easily adopted alternative. The SAQ has a relatively stronger research base than the IORNS, shows promise as a general and violent recidivism prediction tool, and contains mostly dynamic items. Like the IORNS, it is a relatively low cost, self-administered instrument. While research evidence reviewed above suggests that the SAQ is largely comparable to the LSI-R in terms of predictive validity, the relative lack of empirical research compared to the LSI-R detracts from serious consideration of it as a replacement for the LSI-R at this point. To the extent that the cost of other assessment instruments becomes a more pressing consideration, both the SAQ and IORNS are worthy of consideration as assessment tools in the future and once more validation research has been conducted with the tools. Alternatively, the state of Utah may want to consider evaluating these tools in the Utah criminal population and as alternatives to the LSI-R and LS/CMI.

---

<sup>8</sup> See <http://www.mhs.com/product.aspx?gr=saf&prod=lsirn&id=overview> for a review.

## **Risk, Need and Responsivity Assessment in Utah**

In Utah, the LSI-R has been the assessment tool of choice for approximately the last decade. It is used as part of the presentence investigation (PSI). Scores from the LSI-R are provided to the courts and risk factors of concern are identified. Scores are also used to determine supervision classification levels for probation and parole. The LSI has been used as a screening tool for Utah's Drug Offender Reform Act (DORA), and partially determines eligibility for treatment funds. At the prison level, the LSI is used to determine the type and extent of prison services and programming.

### **LSI-R in Practice**

The LSI-R is available in a paper-and-pencil format using a QuikScore form and through a software application. The software application is the focus of this section describing the scoring of the LSI-R, but the same scoring methods apply to the paper-and-pencil version. Assessment users are presented with a graphic user interface and links which connect to the 10 subscales of the LSI-R; the 10-subscales represent separate electronic pages with items composing them listed on the same page. Responses are predominately in the yes/no format where a yes ("1") indicates a problem exists and a no ("0") indicates no problem on the item. However, some items within subscales are ordinal categories. For example, in the education/employment subscale/domain, the question "currently employed?" is answered as yes or no. However, an item in the financial section ("financial problems?") is answered on a 0, 1, 2 or 3 scale where 0 is completely unsatisfactory and 3 is satisfactory with no need for improvement. Ultimately, all scores are recoded so that a "1" is an unfavorable response and a "0" is a favorable response. For items coded 0, 1, 2, or 3, scores of 0 or 1 are recoded into a 1 while scores of 2 or 3 are recoded into a 0. The total number of summed 1s creates a person's assessment score, and can range from 0 to 54.

Domain scores are calculated for each subscale as the user enters responses to the individual items. Some items allow for free text entry of comments, but these comments are not used for calculation of domain/subscale or total scores. Additionally, some items allow you to further detail a response by selecting options from a dropdown menu. For example, under the alcohol/drug problem domain/subscale, the item "drug problem, currently?" allows the user to specify the type of drug, with the option to select multiple drugs (if applicable).

After answering all questions, the assessor is presented with an assessment score page which provides a total score for each subscale and an overall score. Error checks prevent invalid or illogical entries. For example, if the question assessing whether the individual completed grade 10 is marked "no," the question assessing whether the individual completed grade 12 must also be "no." Also, the system will not allow an assessment to be posted if more than five questions are unanswered. Professional overrides can be made, but a reason must also be entered. Finally, offenders can be grouped into classification categories that place them at greater risk of recidivism, and also help indicate the level of supervision of service they need.

Published values for the five categories of offenders are: Low 0-13, Low/Moderate 14-23, Moderate 24-33, Moderate/High 34-40, High 41 and above. Utah, by comparison, uses a four category system: Low 0-13, Moderate 14-23, High 24-40, Intensive 41-54. The main difference is that (with the exception of the value of 23 from the published low/moderate category) Utah collapses across the published moderate and moderate/high categories.

### **The Future of Risk, Need, Responsivity Assessment in Utah: The LS/CMI**

The Utah Department of Corrections is considering a transition from the LSI-R to the newer LS/CMI section one<sup>9</sup>. The details of the LS/CMI section one were outlined in the previous section. In summary, it reorganizes items from the LSI-R (deleting and moving some items) to better fit the domains important to RNR assessment as dictated by the “Central Eight” domains. The domains measured by the LSI-R and LS/CMI (as well as the number of items in each domain score) are outlined in Table 1. The first seven domains overlap, and measure similar (in some cases identical) constructs.

Table 1: Summary of Domains and Number of Items from the LSI-R Relative to the LS/CMI Section I.

<b>LSI-R Domains</b>	<b>LS/CMI Section I Domains</b>
Criminal History (10)	Criminal History (8)
Education/Employment (10)	Education/Employment (9)
Family/Marital (4)	Family/Marital (4)
Leisure/Recreation (2)	Leisure/Recreation (2)
Companions (5)	Companions (4)
Alcohol/Drug Problem (9)	Alcohol/Drug Problem (8)
Procriminal Attitudes/Orientation (4)	Procriminal Attitudes/Orientation (4)
Accommodations* (3)	Antisocial Pattern (4)
Financial* (2)	
Emotional/Personal (5)	

\*Indicates the domain or an item from the domain was moved into the Antisocial Pattern subscale of the LS/CMI Section I

<sup>9</sup> Utah already uses proprietary software for case management; hence, the entire LS/CMI is not the subject of this report. Only section one of the LS/CMI, which measures RNR factors, is discussed. From this point forward, references to the LS/CMI refer only to section one, but the term LS/CMI is used for parsimony.

## **Implementation Evaluation of the LSI-R in Utah (within UDC)**

Despite the use of the LSI-R for nearly a decade in Utah, and the potential adoption of the LS/CMI as an updated assessment tool, little is known about the tool's predictive validity within the state. This research project, therefore, began with an original goal of examining and validating the LSI-R and LS/CMI (which is a shorter assessment and can be calculated completely using the items from the LSI-R) as recidivism assessment tools in the Utah population. While the issue of validation within Utah will remain an important areas of future research, preliminary analyses of LSI-R data provided to UCJC revealed problems in the software system the UDC uses to enter item-level scores for the LSI-R and in calculation of total scores for risk classification. As described in more detail below, these problems made a test of the predictive validity of the LSI-R in Utah impossible at the present time.

Rather than using inaccurate data in attempting to validate the instrument(s) for use as recidivism risk prediction tool(s) in Utah, the present research necessarily altered focus to examine the extent of the problems in score calculation and risk classification that resulted from software-level issues. The present research also discusses evidence suggesting that the LSI-R tool, and its inherent difficulties in administration, may have contributed to item-level, total-score and risk calculation discrepancies going unnoticed for an extended period of time.

### **Project Background**

Originally, a data request was submitted to the Utah Department of Corrections soliciting demographic variables, offense histories and LSI-R item scores for all individuals beginning probation or released to parole between 1/1/2008 and 12/31/2010 (regardless of when the original crime for which they were convicted was committed). A two year follow-up was planned to examine recidivism occurring in the two-year period following probation start or community release.

When data analyses first began with a goal of testing the predictive validity of the LSI-R in Utah, UCJC calculated domain and total scores for both the LSI-R and LS/CMI using a UDC provided file containing the item-level responses of probationers and parolees meeting the aforementioned eligibility criteria. Analyses (using Area Under the Curve (AUC) procedures) revealed relatively poor recidivism prediction for the overall and domain scores of the LSI-R and LSI/CMI<sup>10</sup>. Because the AUC values were relatively low when compared to prior research, the current project sought to better understand the nature of the poor predictive values.

As a result of the unexpected findings, three hypotheses were postulated to explain why the LSI-R and LS/CMI were not as predictive in Utah as other jurisdictions. These included: 1) the tools are simply not as predictive, and are not applicable, in the Utah offender population, 2) a data quality problem exists (e.g., data are not entered or recorded as intended by the LSI-R administrators), or 3) administrative problems exist (e.g., a lack of sufficient trainings exists or the LSI-R is simply too difficult to complete with fidelity).

In order to investigate whether hypothesis one was most accurate, the latter two would first need to be demonstrated to be false. If data quality or administrative issues were found to exist, hypothesis one could not be examined to the extent that data do not accurately represent the intent of the LSI-R tool or, alternatively, an administrator/user. The issues of data quality and administrative difficulties were

---

<sup>10</sup> Although the LS/CMI showed greater predictive validity, it is not clear whether this is an artifact of the problems in the data or a potential true improvement that might remain when new, accurate data become available. The specific AUC values are not provided because they are not meaningful given the data quality issues.

examined separately, but it is important to note that one does not preclude the other. They are not mutually exclusive, and both could be potential causes of poor predictive validity that are unrelated to the true predictive validity of the LSI-R if it were administered and recorded as intended.

Because the original intent of this research project was data driven, concerns surrounding data quality were given relatively greater attention. The issue of administrative concerns was examined only superficially because the Institutional Review Board (IRB) proposal for the project did not cover interviews with LSI-R administrators. Therefore, discussion of the potential for administrative concerns is limited to UCJC staff interacting with the tool the state of Utah uses to complete LSI-R assessments, unsolicited information volunteered by UDC staff who assisted in demonstrating the tool to UCJC staff, and prior research addressing the LSI-R in Utah (findings relating to use of the LSI-R in Utah are derived from the a report (discussed in more detail later) by the Office of the Legislative Auditor General for the State of Utah (2013)).

Because the issues surrounding data quality and administrative concerns applied to all LSI-R assessments, the dataset was expanded from the original, time-restricted dataset to one that included all LSI-R assessments from its inception as an assessment in Utah (2000) to September 2013. The database included 97,641 assessments. The average number of assessments per person was 3.87 but ranged from one to 24 with a median of four assessments. Twenty-two percent (21.9%) of individuals had more than five assessments, and 1.6% had more than 10. The majority of the assessments were conducted within the parole population (59.9% relative to 40.1% probation). Parole or probation status was not static, however, and the same individual could be a probationer at one time and a parolee at another time.

## **Addressing Data Quality as a Potential Cause of Poor Predictive Validity**

In order to examine whether data quality factors contributed to low AUC values, three primary analyses were conducted. The first analysis examined the number of invalid tests with total scores still computed. According to the owners of the LSI tools, Multi-Health Systems (MHS), and the LSI-R manual (Andrews & Bonta, 1995), an assessment becomes invalid if it has more than five missing items (this is not the standard protocol in Utah). A second analysis compared total scores calculated by the UDC to total scores calculated by UCJC using SPSS syntax adopted from the owners of the LSI tools, MHS. A third analysis examined the data for item-level response combinations that are not allowed per instructions of the LSI-R manual. For example, if a person is marked as having two or more prior convictions (LSI-R item two) or three or more prior convictions (item three) he or she must also be marked as having “any” prior convictions on item one. The issue of invalid item-level response combinations is relevant both as a data quality issue and an administrative issue; hence, it is discussed here and in the next section regarding administrative concerns.

### **Invalid Assessments (Missing Items)**

Of the 97,641 total assessments over the 13-year period, 13.7% (13,329) should have been considered invalid by MHS standards because too many items were missing to calculate a meaningful LSI-R score. The mean number of missing items was 1.87 per assessment, but 6.1% had over 10 missing items. The policy recommended by the LSI-R developers when more than five items are missing is to locate the relevant information before finalizing an assessment (Andrews & Bonta, 1995; Andrews, Bonta, & Wormith, 2004). Because LSI-R total scores (as well as corresponding risk levels) were calculated with many items missing, total scores (and consequentially risk categories) are inevitably lower on average than an indeterminable true score. Utah may want to consider adopting the MHS standards and requiring at least a minimum number of items be entered before submission of the assessment can occur.

**Discrepant Total Score (Total Scores that are not the Sum of Item-Level Responses)**

A preliminary analysis comparing UDC-provided total scores to UCJC-calculated total scores revealed the UDC scores did not perfectly match UCJC-computed total scores. Furthermore, when the individual item scores from UDC data were summed, those values were different from the UDC (or UCJC) total score (recall the total score is the sum of all items with a recoded value of “1” from the assessment). Table 2 shows the means and standard deviations of the LSI-R scores as calculated by UDC and UCJC separated by valid and invalid assessment status (i.e., five or less missing items and more than five missing items, respectively).

Interestingly, invalid tests calculated by UDC have the highest mean score, and UCJC calculations for invalid tests have the lowest mean score. Scores for valid assessments are similar across the two agencies. This finding may suggest that missing items were being included in the total calculated by the software used by UDC. However, if missing items were being included in the total score calculation, they were not included consistently, or in a systematic fashion, because the total scores calculated by UDC also did not equal the sum of the missing items plus the individual items marked as “1” after recoding.

Table 2: Mean LSI-R Score by Invalid/Valid Status by UDC and UCJC Calculations

Source	Invalid			Valid		
	Mean	Std. Dev.	N	Mean	Std. Dev.	N
UDC Provided Total Score	23.7	8.62	13,329	22.5	9.36	84,312
UCJC Calculated Total Score	18.0	7.27		22.3	9.37	

Simply because UCJC total scores were calculated by summing the raw scores for items recoded as “1” did not necessarily dictate that UCJC totals were correct. Problems might also exist at the individual item-level that precluded

correct summation. In order to examine the accuracy of the UCJC calculated total scores, UCJC staff was provided with a tutorial of the software system UDC uses to administer assessments and create corresponding total scores. In a meeting with UDC staff, UCJC staff provided UDC staff with a sample of offender numbers for individuals whose total scores did not match the sum of item scores and for whom UDC and UCJC total scores did not match.

In order to examine the issue, UDC staff created a copy of an assessment for which the discrepancy was present and opened the copy in a test server (eliminating the possibility that changes could accidentally be recorded and saved). When this assessment was submitted under the newest version of UDC software (timeframe of adoption of the new system is explained below), the new UDC total score did not match the original UDC total score for the assessment, but did match the UCJC total score that was the sum of item-level responses. Inspection of a sample of discrepant assessments revealed the same pattern of results, providing support for the conclusion that the original UDC total scores were somehow not the sum of recoded item scores or a combination of the item scores and missing items.

For a small number of these discrepant tests, the total calculated by the new UDC system also did not match the UCJC total, but the cause of this discrepancy was easy to identify. New validation checks in the UDC system will not allow the user to submit a test with response combinations that are not allowed; the older software did not include these validation checks, and assessments could be submitted with logical errors. In one specific case, for example, the new system calculated a total score one point higher than UCJC’s total score. The additional point was the result of the assessment administrator noting that a person had two or more prior convictions, but failing to note that he or she had “any” prior convictions. In the older system, this administrative error was submitted as a “0” for the “any” prior convictions item, which ultimately reduced the sum of the item-level responses by one. The new UDC system caught the

error, flagged the impossible response combination for review, and then fixed both values to “1” because two or more prior convictions necessitates that the individual also has “any” prior convictions.

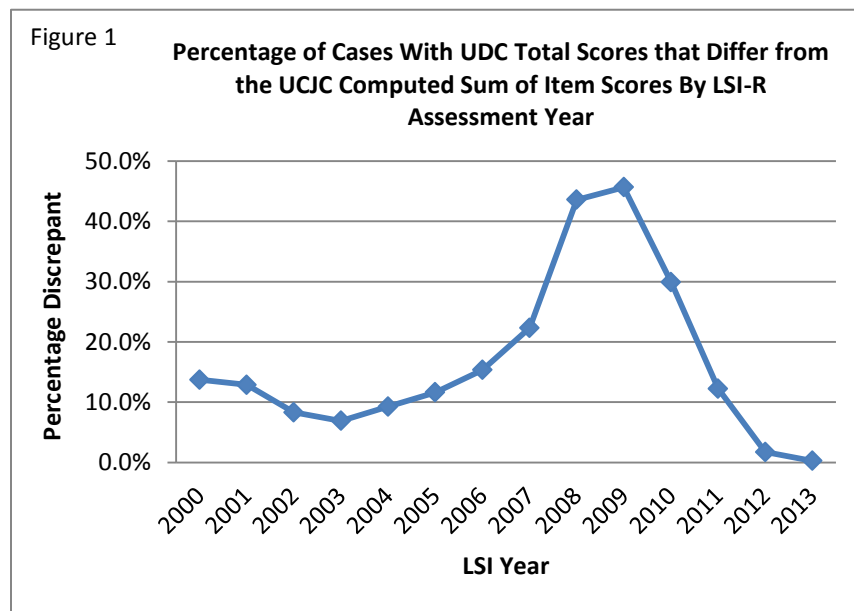
Because the new UDC system seemed to fix errors in total score calculation and impossible item-level responses, the logical next step was to identify whether the discrepancies were specific to a period of time during which software was updated or replaced. Table 3 and Figure 1 demonstrate that the problem in total score calculation has been pervasive since adoption of the LSI-R as an assessment tool; however, the rate at which the discrepancies occur begins to drop from 2011 to 2013. The degree of discrepancy fluctuates from year to year, reaching a zenith in 2009 with 45.6% of all assessments conducted in that year not matching the sum of item-level responses. Though 2013

Table 3: Discrepant Total Score Cases by Year: UDC Provided Relative to UCJC Calculated (Sum of Item-Level Scores)

LSI Year	N Discrepant	Total N	% Discrepant
2000	7	51	13.7
2001	151	1,172	12.9
2002	142	1,710	8.3
2003	128	1,856	6.9
2004	453	4,894	9.3
2005	609	5,233	11.6
2006	940	6,132	15.3
2007	1,529	6,852	22.3
2008	5,269	12,091	43.6
2009	7,194	15,763	45.6
2010	5,134	17,167	29.9
2011	1,425	11,639	12.2
2012	142	8,247	1.7
2013	13	4,818	0.3
Total	23,136	97,625	23.7

actually shows a discrepancy percentage of 0.3%, the discrepant assessments are limited to the first quarter of 2013. After March of 2013, there were no UDC total scores that differed from the sum of item-level responses. To the extent that the individual item-level responses reflect the administrator’s true intent on scoring an LSI-R, deviations from the sum of item-level responses can be regarded as incorrect

total scores. Across the period from 2000-2013, 23.7% of all LSI-R assessments reflected inaccurate total scores.



Discussions with UDC staff failed to elucidate a cause of the vacillating percentage of discrepant scores. Software improvements were a continual process, and new validation software was adopted in 2008, but, as shown in the figure and table, the discrepancies were greatest during the 2008-2009 period. Because UDC was not aware of the discrepancy, it is not clear how the newest software

iteration fixed the problem. It could be that new software was developed starting with new program code, which serendipitously fixed the problem. Regardless of how the newest version was developed, the UDC system now (and since March 2013) appears to be functioning correctly in generating a total score, and

also includes validation checks that fix disallowed item-level response combinations (discussed more below).

Because of the problems in total score calculations, mis-categorization of individuals into inaccurate risk levels is also a concern, particularly because assigned risk level dictates the nature and intensity of services and supervision one receives.

Table 4 shows the number of individuals at each Utah risk classification level according to the original UDC calculations, and then provides the percentage of individuals who should have been classified at a lower risk level or a higher risk level, and the percentage who were classified correctly despite the total score discrepancy<sup>11</sup>. The table also provides (in parentheses) the LSI-R scores that fall within a classification level as defined by the state of Utah. Columns in the last row of the table (“Overall”) represent the overall number and percentage of individuals misclassified when taking into account that misclassification at the lower risk levels accounted for a greater number of people (despite representing a smaller percentage relative to higher risk levels). Rows within this column also account for the fact that a misclassification into a lower risk category cannot occur within the low risk group, and a misclassification into a higher risk category cannot occur within the intensive risk group.

As seen in Table 4, misclassification most often placed a person into a higher risk category than was appropriate based on the true LSI-R total score. The columns within the “True Classification” row indicate where the individual should have been placed relative to where he or she was actually placed (i.e., relative to the “UDC Classification” column). Thus, the 8.6% value in the true classification column “Lower Level” indicates that 8.6% of people who were classified as moderate risk by the UDC classification should have been in a lower classification category. Conversely, 0.2% of those classified as moderate risk by UDC classification should have been in a higher classification. The greatest discrepancy in terms of percentage misclassified occurred within the intensive risk category, wherein 22.3% of individuals classified as intensive should have been classified in a lower risk category. Although this was the category with the greatest percentage of misclassified cases, it impacted the smallest number of people (i.e., 22.3% of 2,086). Overall, 8.5% of all LSI-R assessments between 2000 and 2013 resulted in a classification of an offender into a higher risk than was appropriate (i.e., true classification should have been a “Lower Level”); only 0.1% of all cases were classified into a lower risk level than was appropriate (i.e., true classification should have been a “Higher Level”).

Table 4: Percentage of Misclassifications by UDC Original Classification

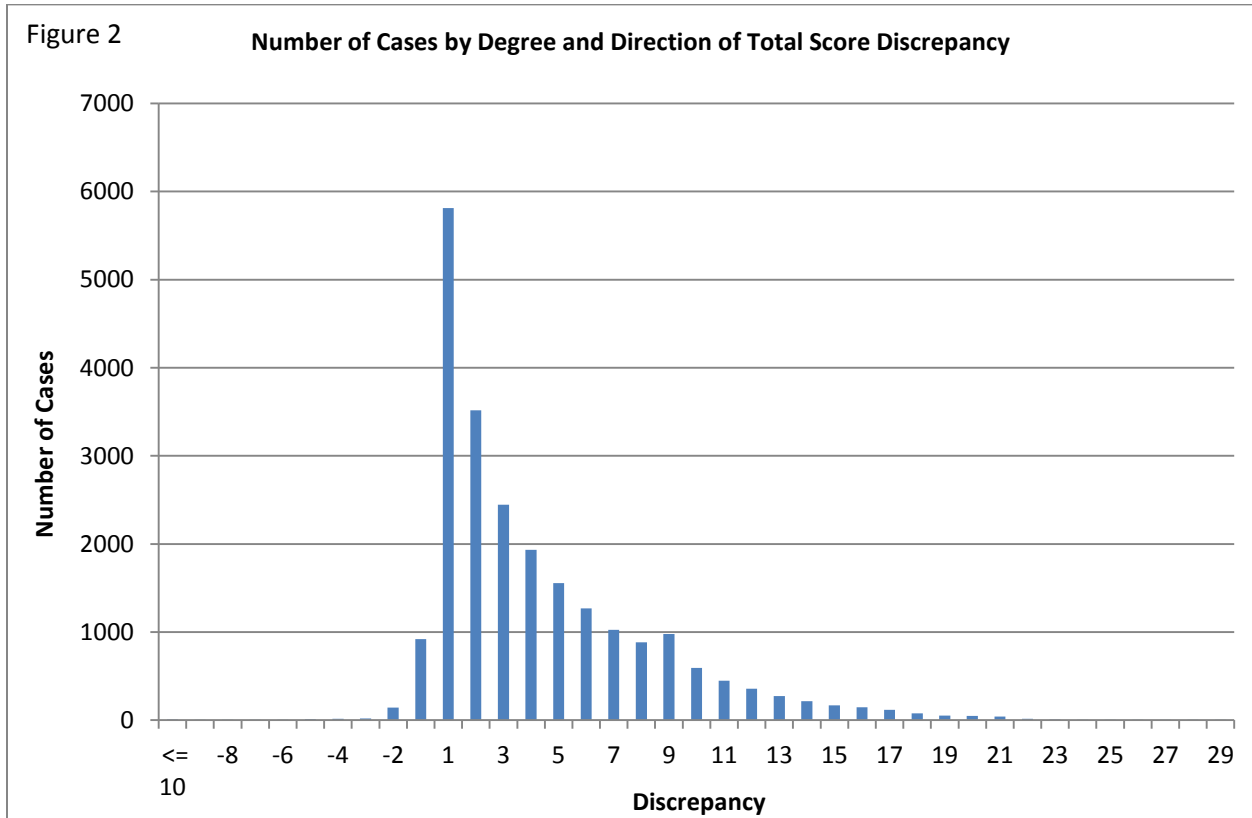
UDC Classification	True Classification (Sum of Item-Level Responses)			
	N	Lower Level	Same Level	Higher Level
Low Risk (0-13)	18,102	NA	99.8%	0.2%
Moderate Risk (14-23)	34,536	8.6%	91.2%	0.2%
High Risk (24-40)	42,917	11.3%	88.6%	0.1%
Intensive (41-53)	2,086	22.3%	77.7%	NA
Overall	97,641	8.5%	91.4%	0.1%

Figure 2 shows a visual interpretation of the degree and direction of the total score discrepancy. For purposes of scaling, cases where a discrepancy was zero (i.e., the UDC reported total score matched the sum of item-level responses) were removed from the figure; 74,500 cases had total scores where the UDC score and the sum of item-level responses matched. The axis labeled “discrepancy” shows the difference

<sup>11</sup> Note that denoting classifications as “correct” is not entirely accurate. As discussed later in the report, administrative errors that led to prohibited item-level response combinations also affect the accuracy of the total score. Here, in the absence of validation checks correcting for impossible item-level response combinations, the sum of the item-level responses is considered the best approximation of the “correct” or “true” LSI-R score.



between the UDC-reported score and the calculated sum of item-level responses. Therefore, a value of one on the x-axis (which is most common) indicates that the UDC score was one point higher than the true score. As seen in Figure 2, the data are skewed toward a higher score discrepancy, and relatively few cases exist in which the UDC score was lower than the true score (this pattern supports data presented in Table 4).



### Prohibited Combinations of Item-Level Responses

The LSI-R manual and scoring guide both dictate several items with if-then logic; that is, if one statement is true (or false), then another statement must also be true (or false). The following are a list of item-level dependencies based on if-then logic:

#### Convictions

- (1) If a person has two or more prior convictions (item two), he or she must be marked as having “any” prior convictions (item one).
- (2) If a person has three or more prior convictions (item three), he or she must be marked as having two or more prior convictions (item two) and “any” prior convictions (item one).

#### Employment Problems

- (1) Current unemployment (item 11) dictates that work specific participation/performance (item 18), peer interactions (item 19) and authority interactions (item 20) must also be marked as problems.

### Grades Completed

- (1) If a person has less than a grade 10 education (item 15), he or she must also be marked as having less than a grade 12 education or equivalent (item 16).

### Social Isolation, Criminal and Prosocial Friends/Acquaintances

- (1) If a person is noted as a social isolate not by his or her choice, but due instead to his or her lack of capacity to form social bonds (such as due to mental or physical disabilities or social skills deficits; item 32), he or she cannot be marked as having criminal friends (item 34).
- (2) If a person is noted as a social isolate not by his or her choice, but due instead to his or her lack of capacity to form social bonds (such as due to mental or physical disabilities or social skills deficits; item 32), he or she must be marked as having an absence of prosocial friends (item 36).
- (3) If a person is marked as having some criminal friends (item 34), he or she must also be marked as having some criminal acquaintances (item 33).

### Alcohol and Drug Problems

- (1) If a person is marked as having a current alcohol problem (item 39), he or she must also be marked as having ever had an alcohol problem (item 37).
- (2) If a person is marked as having a current drug problem (item 40), he or she must also be marked as having ever had a drug problem (item 38).
- (3) If a person does not have a current alcohol problem (item 39) or a current drug problem (item 40) he or she cannot have any problems related to current alcohol or drug use including law violations, marital/family problems, school/work problems, medical problems, or other related problems (items 41-45)

### Active Psychosis

- (1) If a person is marked as having severe interference from emotional and cognitive problems (item 47), he or she must also be marked as having moderate interference (item 46) and psychological indicators of emotional/personal problems (item 50).

The current research project investigated the prevalence of violations of the if-then logic in the UDC-provided dataset using syntax to create flags for violations. Table 5 show the frequency of violations numbered according to the headings provided above for the if-then logic, and includes the item numbers for reference to the LSI-R or its scoring guide. For example, values in the table for the row labeled “Convictions (1)” refer to violations of the first issue described under conviction if-then logic above. A flag for violations was created using specific syntax within the Statistical Product and Service Solutions (SPSS; formerly Statistical Package for the Social Sciences) platform. The following is an example of the form of the syntax:

```
COMPUTE CurrAlcFlag=0.  
IF (q39=1 AND q37 NE 1) CurrAlcFlag=1.  
EXECUTE.
```

For purposes of this report, it is not necessary to understand the syntax beyond the issue of understanding the meaning of the final discrepancy flag it calculates. The first line, a compute statement, tells SPSS to compute a variable called “CurrAlcFlag.” It sets the value for this variable to “0” for every person. The next line creates the flag for a prohibited combination of item-level responses. Specifically, it says that if

question 39 (alcohol problem, currently) equals a value of “1” (which indicates yes, a current problem exists), and question 37 (alcohol problem, ever) does not equal “1” (which indicates yes, a problem existed at some point), then reset the value of “CurrAlcFlag” from “0” to “1”. Note that this syntax flags any cases where a person has a current alcohol problem, but was not also noted as having ever had one.

Also important to note is the use of the term “NE” in the syntax. This line tells SPSS to create the flag if question 39 equals “1” and question 37 does not equal “1.” The values of any LSI-R item (when scored) can only be expressed as “0” (not a problem) or “1” (problem); however, by explicitly stating that the flag be created if question 37 does not equal “1,” the syntax also counts missing values. In other words, if question 37 is “0” or missing, a flag will be created indicating a prohibited item-level response combination. The use of the “NE” term is important because the logic of question 37 is dictated by question 39; if question 39 was answered, and the answer was a value of “1,” then question 37 should never be missing a value because its value was already established by question 39. As of March 2013, the UDC software system will not allow a user to enter any value other than a value indicating an alcohol use problem for question 37 if question 39 indicates a current alcohol problem.

As seen in Table 5, violations of the LSI-R rules regarding if-then logic were pervasive and also varied considerably, by issue, in the frequency of their occurrence. The row for “Convictions (1)” shows that 1.1% of LSI-R assessments had an offender marked as having two or more convictions, but not marked as having “any” convictions. Similarly, 0.9% of assessments (see “Convictions (2)”) had an offender marked as having three or more convictions but not marked as having two or more, “any” prior convictions, or both.

By a wide margin, the most common discrepancy with respect to prohibited response combinations can be found for the items comprising “Employment (1).” The prohibited response combination of marking item 11 as problematic without marking at least one of items 18-20 as well occurs 41.6% of the time. If current unemployment (item 11) is marked as a problem, if-then logic dictates that work specific participation/performance (item 18), peer interactions (item 19) and authority interactions (item 20) must all be marked as problems. To some extent, this may seem counterintuitive to the LSI-R administrator: if a person does not have a job, administrators might not interpret the person as having participation or interaction problems at work. The intent of marking these items as “1” despite the lack of actual employment is to weight unemployment more heavily, such that, if one is unemployed, it is a relatively greater factor in determining the overall score. By nature of marking items 18-20 as problems if the person is also unemployed, unemployment equates to four additional points on the LSI-R<sup>12</sup>. Though this is the psychometric logic behind scoring items 18-20 as problems even when a person does not have a job, the reasoning is not explained in the manual, and is perhaps often misunderstood.

The least common discrepancy, occurring in only 0.3% of cases, involved the items relating to level of education completed. Rarely, administrators marked an individual as having completed grade 12 (item 16), but not grade 10 (item 15).

In 4.4% of cases, a person noted as being a social isolate (item 32) was inappropriately marked as having criminal friends (item 34; see “Social Isolation and Friends (1)”). Somewhat less frequently (3.5% of cases), a person was noted as a social isolate but was not marked as having an absence of prosocial friends (item 36; see “Social Isolation and Friends (2)”). The least common discrepancy among socialization items occurred for items 33 and 34. In 1.1% of cases, a person marked as having some criminal friends (item 34) was not also marked as having some criminal acquaintances (item 33; see “Social Isolation and Friends (3)”).

---

<sup>12</sup> A person is not considered unemployed if he or she is working part-time, is a seasonal worker, is in a work skills training program, is a full-time student, a homemaker or a pensioner, is retired, employed in the institution in which he or she is incarcerated, or is serving less than two years and will verifiably return to the same job after release.

Alcohol and drug related items revealed relatively few prohibited item-level responses. In 1.1% of cases, a person was marked as having an alcohol problem currently (item 39), but not as having ever had one (item 37; see “Alcohol and Drugs (1)”). In 0.3% of cases, a person was marked as having a drug problem currently (item 40), but not as having ever had one (item 38; see “Alcohol and Drugs (2)”). The most common issue regarding prohibited item-level response combinations on alcohol- and drug-related items involved the combination of items 39-45. As mentioned, the LSI-R manual dictates that, if a person does not have a current alcohol problem (item 39) or a current drug problem (item 40), he or she cannot have any problems related to current alcohol or drug use including law violations, marital/family problems, school/work problems, medical problems, or other related problems (items 41-45). In 3.3% of cases neither a current alcohol nor drug problem was indicated, but a separate problem relating to an alcohol or drug problem was indicated (see “Alcohol and Drugs (3)”).

In 3.8% of cases, a person was indicated as having a severe emotional or cognitive problem that interfered with his or her quality of life and response to life stressors, but was not also marked as having moderate interference from the same causal factors. Excluding prohibited item-level responses related to employment (i.e., Employment (1), which revealed by far the highest

Table 5: Frequency of Violations of If-Then LSI-R Logic

If-Then Issue	Items	N (total N=99,641)	%
Convictions (1)	1 and 2	1,069	1.1
Convictions (2)	1, 2 and 3	921	0.9
Employment (1)	11, 18, 19 and 20	40,619	41.6
Grades (1)	15 and 16	277	0.3
Social Isolation and Friends (1)	32 and 34	4,269	4.4
Social Isolation and Friends (2)	32 and 36	3,391	3.5
Social Isolation and Friends (3)	33 and 34	1,066	1.1
Alcohol and Drugs (1)	37 and 39	669	1.0
Alcohol and Drugs (2)	38 and 40	316	0.3
Alcohol and Drugs (3)	39-45	3,247	3.3
Active Psychosis (1)	46, 47 and 50	3,679	3.8

percentage of prohibited item-level responses), 15.1% of all cases (not shown in table; N=14,764) had at least one combination of responses that violated the if-then logic of the LSI-R.

Mistakes in if-then logic are both a data quality issue and an administrative issue; however, because they can be stopped at the data-quality level through software-level data validation checks, they have been discussed in the data quality section. Also, it is important to note that, as of this report, data quality checks are present in the UDC system that will not allow submission of an LSI-R assessment if any of the if-then logic is violated. The user will receive an error message upon attempted submission, and will be prompted to fix items violating the if-then logic. Total scores are also now calculated correctly, and the sum of item-level responses equals the calculated total score in all cases. However, as of this report, a user could still submit an assessment with more than five missing items.

The extent of the problems displayed in Table 5 speaks to the issue of why a validation study could not be conducted on the LSI-R as a recidivism prediction tool in Utah using data collected from software in place prior to March of 2013. While total scores could have been fixed by using a new total score that was simply the sum of all item-level responses, and invalid assessments could have been deleted from analyses, discrepancies in if-then logic could not be fully repaired in hindsight because some items cannot be fixed without knowing the administrator’s intent.

To illustrate this point, consider items 39-45. Item 39 assesses existence of a current alcohol problem, while item 40 assesses existence of a current drug problem. Items 41-45 assess life problems that are caused by an alcohol problem, a drug problem, or both (including substance-related law violations, marital/family problems, school/work problems, medical problems or other related problems). If neither items 39 or 40 were marked as indicating a current alcohol or drug problem, but one of items 41-45

indicated a life problem relating to a current substance use problem, it would not be clear whether the administrator: (1) mistakenly did not indicate a current drug or alcohol problem despite noting a life problem related to current alcohol or drug use; (2) meant to mark a current alcohol, drug problem or both; or (3) completed the life problem items about the offender's life in general rather than specifically related to a current substance abuse problem (and a substance abuse problem did not, in fact, exist).

As a more straightforward example, consider items 15 and 16. Item 15 assesses whether the person completed grade 10, while item 16 assesses whether he or she completed grade 12. Failure to mark a person as having completed grade 10 when he or she did complete grade 12 could indicate the administrator accidentally mismarked item 10, intending to indicate the offender passed both grades 10 and 12, or it could indicate that the administrator mistakenly marked the offender as completing grade 12 when, in actuality, he or she completed neither grade 10 or 12.

Because there is no way of knowing the administrator's intent when completing these types of items, there is no way to correct the prohibited response combinations; hence, there is no way of providing an accurate assessment of the LSI-R's validity. Once the data validation checks adopted in March of 2013 have been in place for a year or more, the issue of the LSI-R's predictive validity in Utah can be more appropriately addressed.

## **Addressing Administrative Issues as a Potential Cause of Poor Predictive Validity**

The current research also examined administrative-level factors as potential causes of lower-than-expected predictive validity for the LSI-R in Utah: (1) inadequate training, (2) inherent difficulty of the LSI-R assessment, and (3) ambiguities on the LSI-R response form.

### **Inadequate Training**

Issues addressed above, under data quality concerns, demonstrate that concerns also exist regarding how the LSI-R is administered and the adequacy of training administrators of the assessments receive. For example, if administrators were fully aware of the if-then logic of many of the items, errors revealing prohibited item-level response combinations could be reduced substantially. The lack of understanding of these response dependencies is exemplified by the item assessing current unemployment. As noted above, if a person is unemployed, three subsequent LSI-R items must also be noted as problems. However, the three subsequent items (items 18-20) were not marked as problems when unemployment existed in 41.6% of LSI-R assessments. This outcome suggests a lack of familiarity with the LSI-R manual, and perhaps a lack of training.

Another issue suggesting a lack of sufficient training and familiarity with the LSI-R's directions involved the frequency with which certain items were left blank or skipped. While all items were skipped to some degree (range 0.5% missing/skipped to 8.9% missing/skipped), certain items were skipped at a relatively alarming rate (i.e., nearly 1 in 11 assessments skipped the items; see Table 6); moreover, these frequently-skipped items clustered into specific domains of the LSI-R. For example, recall that item 11 assesses current unemployment, and that items 18, 19 and 20 (assessing work related participation/performance, work peer interactions, and work authority interactions, respectively) must be marked as problems if the individual was unemployed. These four related items were four of the five most commonly skipped items in the LSI-R data. The frequency with which these items were skipped may suggest a lack of familiarity with the LSI-R's if-then logic, and the concomitant weighting of unemployment as an important risk factor. The aforementioned finding that these items often revealed prohibited response combinations augments the possibility that administrators simply do not know how to respond or do not understand the logic of the directions with respect to these items. Particularly with offenders who are, or have recently

been, incarcerated, administrators may find these items difficult to complete. However, the LSI-R manual does offer guidelines regarding how to score these items for incarcerated individuals.

Table 6: Ten Most Commonly Skipped Items on the LSI-R

Item	Skip Frequency (%)
40. Drug problem, currently?	8.9
18. Work participation/performance	8.6
20. Work authority interactions	8.4
11. Currently unemployed?	8.2
19. Work peer interactions	8.0
41. Law violations?	7.4
31. Could make better use of free time	6.6
21. Financial problems?	6.3
27. Unsatisfactory accommodations	6.1
23. Dissatisfaction with marital or equivalent situation?	5.7

In order to administer the LSI-R, users must (1) have completed graduate level course work in tests and measurement, (2) have completed training conducted by an approved MHS trainer, or (3) be directly supervised by someone who meets criteria one and two. Unfortunately, the present study could not interview LSI-R administrators in order to determine whether they had received proper training or supervision because the original intent of the project, and hence the approved IRB proposal,

covered retrospective data analysis only. It did not cover interviews or collection of new data regarding training and knowledge of the LSI-R by its administrators. However, a report issued by the Office of the Legislative Auditor General for the State of Utah (2013) addresses the issue of training.

Interviews conducted with Adult Probation and Parole (AP&P) staff as part of the audit indicated that “some staff members stated that agents have not been trained on the proper administration of the LSI-R, instead learning on the job from other employees and self-study” (p. 21). The audit references an independent report (also reviewed as part of the present research) by Lowenkamp, Latessa and Holsinger (2004) providing evidence that the predictive validity of the LSI-R declines substantially when staff are inadequately trained in its administration. Recall that the LSI-R’s predictive validity in Utah was initially examined as part of the current research project, but the outcomes from those analyses were jettisoned because both data quality and administrative-level errors were found. In light of the frequency with which prohibited item-level response options were found, it is not surprising that the jettisoned analyses revealed poor predictive values for the LSI-R in Utah. Until LSI-R administrators are all trained to the standards set forth by MHS, any validation of the LSI-R in Utah will be misrepresentative of the actual predictive validity of the tool.

Even with greater training, findings from the AP&P audit suggest that an additional threat to the LSI-R’s accuracy derives from the fact that some staff do not believe in the validity of the tool. Interviews conducted as part of the audit revealed that both staff and their supervisors do not have confidence in the LSI-R’s validity. Other research conducted with parole agents in Utah also corroborates the concern over agents’ perceptions of the validity of the LSI-R (Peterson, Hickert & Dorsey, 2008). The extent to which this lack of confidence influences the thoroughness with which administrator’s complete the LSI-R is unknown; however, it seems plausible that if they doubt its validity (and, as the audit found and other research indicates, sometimes do not use it to help focus case management goals on important risk factors), administrators may be less careful when completing the assessment. Speculatively, they may, for example, merely skip difficult or confusing items rather than seeking clarification.

### **Inherent Difficulty of the LSI-R Assessment**

The LSI-R is so widely utilized as a risk and needs assessment tool that it is largely unquestioned whether it is actually difficult to administer in practice. Though it has already been acknowledged that a lack of training decreases the predictive ability of the tool, the issue of whether training alone is sufficient to

make a person a proficient user of the tool remains largely unexamined. It could be, for example, that some people (based on individual proclivities) are simply not well-suited to the task of using a complicated psychometric instrument such as the LSI-R.

Austin (2006) raises concerns regarding difficulties administering the LSI-R, and warns of the subjective nature of many of the LSI-R items. Regardless of research on the instrument, however, it is easy to see (upon reviewing the actual instrument) how different administrators could have difficulty reliably providing the same responses to an individual's LSI-R assessment. For example, one can see how it would be difficult for an administrator, interviewing an individual incarcerated for many years, to decide whether the individual has only "some criminal friends" (where a response of "relatively satisfactory situation" does not count as a point on the instrument (and hence greater risk), and a response of "relatively unsatisfactory situation" does).

Similarly, absence of participation in an organized activity (yes or no) is considered a risk factor leading to an additional point, but the definition of organized activity is not immediately clear. MHS suggests church qualifies as an organized activity, but only if one participates in church activities beyond mere attendance. The definition of a "formal organization" also seems ambiguous. A group of friends playing a sport on a regular basis, for example, may not be considered "formal" despite the fact that it promotes activity that is perhaps antithetical to criminal activity and is a structured, patterned behavior. Furthermore, the response (yes or no) to whether the individual has ever had an alcohol problem or a drug problem states only: "the assessment of an 'alcohol' (drug) problem depends upon the interviewer's assessment and not the offender's evaluation" (Andrews, Bonta & Wormith, 2004, p. 18). The criteria for making such an assessment are not further specified.

Some evidence of administrative difficulties/inconsistencies was found in the UDC data. Prohibited item-level response combinations and frequently skipping items may be partially driven by a lack of understanding of the LSI-R manual, but, as the examples above illustrate, even with an adequate knowledge of the manual's scoring guide, some items are simply ambiguous and subject to a degree of speculation and interpretation.

### **Ambiguities on the LSI-R Response Form**

LSI-R administrators complete the assessment on a software platform developed by UDC. A member of UCJC staff was able to review the system with a member of the supervisory staff from UDC, and was able to address specific anomalies/concerns found in the LSI-R data during preliminary analyses. What follows are merely observations about the platform. Comments are limited to impressions of these two individuals, and should be considered starting points for discussion of how the overall user experience might be improved for more accurate responses.

It is the opinion of these staff that the LSI-R's software format may create ambiguities in how to respond to certain items. For example, recall that some items on the LSI-R are scored on a scale of 0, 1, 2, or 3. A score of "0" or "1" on the four-point response scale indicates an unsatisfactory situation, while a "2" or "3" indicates an acceptable one. Next to each of the items with the four-point response scale are two sets of numbers: a series of "0"s and "1"s (which indicate how the items will ultimately be scored by the software), and a "0," "1," "2," and "3" (representing the actual response scale the administrator should use to score the item). Each of these respective sets of numbers is in a separate column. The actual four-point response scale is in column two (next to the selectable radio button), and the first column contains the "0" and "1" values included in the final scoring. The administrator need only mark the answer provided in the response scale of column two; they do not need to know how it is scored by the system. The first column simply indicates the score the person ultimately receives that counts toward his or her total. A score of "0" or "1" in column two means a user would mark a response corresponding to a "1" in

the first column, while a score of “2” or “3” corresponds to a “0” in the first column. Because the software converts the administrator’s score into the actual point score, it is unnecessary for the user to see the first column, and could ultimately be confusing (see Figure 3).

Additionally, section headings are not always clear in the UDC web-based system. Unlike the paper version, where questions in a domain are clearly marked by a large header, the web-based version has a small header, and once a user scrolls down on the page, the heading is no longer visible. Consequently, it is not always clear that subsequent questions relate to the previous header/domain. Under the Financial section, for example, a question simply asks: “Problems in the last year?” If a user lost track of where he or she was, it may not be clear that the question refers to financial problems. A similar problem occurs with questions assessing outcomes relating to or caused by substance use. A question asking about “law violations” is asking specifically about violations relating to substance use, but, again, this may not be clear if a person lost track of the section in which he or she was responding. It might remedy the problem if each question repeated the header or topic as part of the question.

Figure 3

Section: ALCOHOL/DRUG PROBLEMS	
<b>37 ALCOHOL PROBLEM, EVER</b>	
1.0	<input checked="" type="radio"/> YES
	<input type="text" value="under age"/> 0.111
0.0	<input type="radio"/> NO
	<input type="text"/> 0.0
<b>38 DRUG PROBLEM, EVER</b>	
1.0	<input checked="" type="radio"/> YES
	<input type="text"/> 0.111
0.0	<input type="radio"/> NO
	<input type="text"/> 0.0
<b>39 ALCOHOL PROBLEM, CURRENTLY - 1 YEAR REGARDLESS OF SETTING</b>	
0.0	<input type="radio"/> 3
	<input type="text"/> 0.0
0.0	<input checked="" type="radio"/> 2
	<input type="text"/> 0.0
1.0	<input type="radio"/> 1
	<input type="text"/> 0.111
1.0	<input type="radio"/> 0
	<input type="text"/> 0.111
<b>40 DRUG PROBLEM, CURRENTLY/SPECIFY DRUG - 1 YEAR REGARDLESS OF SETTING</b>	
0.0	<input type="radio"/> 3
	<input type="text"/> 0.0
0.0	<input checked="" type="radio"/> 2
	<input type="text"/> 0.0
1.0	<input type="radio"/> 1
	<input type="text"/> 0.111
1.0	<input type="radio"/> 0
	<input type="text"/> 0.111



## Discussion

A review of the literature on the current state of science in Risk, Need and Responsivity assessment indicated that the Level of Service tools (the LSI-R and LS/CMI) remain at the forefront of recidivism prediction instruments. While many tools in the same generation as the LSI-R (third) or LS/CMI (fourth) reveal promise, none have as much research demonstrating they are efficacious predictors of general recidivism.

The state of Utah has utilized the LSI-R since 2000, and, despite the vast research support for the LSI-R, little is known about the predictive utility of the instrument in the Utah population. This research project, therefore, began with an original goal of examining and validating the LSI-R and LS/CMI (which is a shorter assessment and can be calculated completely using the items from the LSI-R) as recidivism assessment tools in the Utah population. However, preliminary analyses of LSI-R data provided to UCJC revealed problems that precluded testing the validity of the LSI-R in Utah. Problems included both data quality issues (i.e., invalid tests due to too many missing items, total scores that did not match the sum of item-level responses and prohibited combinations of item-level responses) and administrative issues (inadequate training of administrators, inherent difficulties in using the LSI-R and ambiguities on the LSI-R response form).

Problems in these areas made a test of the predictive validity of the LSI-R in Utah impossible at the present time. While the issue of validation within Utah will remain an important area of future research, a great deal of work remains to be done to ensure such an evaluation is a fair assessment of the LSI-R's validity. Some work has already been accomplished: UDC's newest LSI-R software iteration (adopted in March of 2013) now correctly calculates total scores, and will not allow prohibited item-level response combinations. The decision to continue to use assessments that MHS considers invalid due to too many missing items continues to threaten the predictive validity of the tool.

Moving forward, improvements can be made that will facilitate a fair and accurate evaluation of the LSI-R's predictive validity (and its overall usefulness as a risk and needs classification tool) by: 1) ensuring all users are adequately trained and 2) improving the ease of interpreting the software's user interface. Difficulties inherent in the LSI-R will remain an important issue for discussion, but much of this difficulty can be attenuated by ensuring minimal standards of training on the instrument. Increased training may have the additional benefit of helping users better understand the instrument, creating greater faith in its validity and overall usefulness

## References

- Ægisdóttir, S., White, M.J., Spengler, P.M., Maugherman, A.S., Anderson, L.A., Cook, R.S., et al. (2006). The meta-analysis of clinical judgment project: Fifty-six years of accumulated research on clinical versus statistical prediction. *The Counseling Psychologist, 34*, 341-382.
- Andrews, D. A., & Bonta, J. (1994). *The psychology of criminal conduct*. Cincinnati, OH: Anderson
- Andrews, D. A., & Bonta, J. (1995). *The level of service inventory—revised*. Toronto, Ontario, Canada: Multi-Health Systems.
- Andrews, D. A., & Bonta, J. L. (2003). *The level of service inventory revised: U.S. norms manual supplement*. Toronto, Ontario, Canada: Multi-Health Systems.
- Andrews, D. A., & Bonta, J. L. (2010). *The psychology of criminal conduct* (5th ed.). New Providence, NJ: LexisNexis.
- Andrews, D. A., Bonta, J. L., & Wormith, J. S. (2004). *Level of service/case management inventory (LS/CMI): An offender assessment system, user's manual*. Toronto, Ontario, Canada: Multi-Health Systems.
- Andrews, D. A., Bonta, J., & Wormith, S. J. (2006). The recent past and near future of risk/need assessment. *Crime and Delinquency, 52*, 7–27.
- Austin, J. (2006). How much risk can we take? The misuse of risk assessment in corrections. *Federal Probation, 70*(2), 58-63.
- Austin, J., Coleman, D., Peyton, J., & Johnson, K. D. (2003). *Reliability and validity study of the LSI-R risk assessment instrument*. Washington, D.C.: Institute on Crime, Justice, and Corrections at The George Washington University.
- Baird, S., Heinz, R. C., & Bemus, B. J. (1979). *The Wisconsin case classification/staff deployment project: Two-year follow-up report*. Madison: Wisconsin Division of Corrections.
- Belfrage, H., & Douglas, K. S. (2002). Treatment effects on forensic psychiatric patients measured with the HCR-20 violence risk assessment scheme. *International Journal of Forensic Mental Health, 1*, 25-36.
- Bonta, J., Harman, W. G., Hann, R. G., & Cormier, R. B. (1996). The prediction of recidivism among federally sentenced offenders: A re-validation of the SIR scale. *Canadian Journal of Criminology, 38*, 61-79.
- Brennan, T., Dieterich, W., & Ehret, B. (2009). Evaluating the predictive validity of the COMPAS risk and needs assessment system. *Criminal Justice and Behavior, 36*(1), 21 – 40.
- Brennan, T., & Oliver, W. L. (2000). *Evaluation of reliability and validity of COMPAS scales: National aggregate sample*. Traverse City, MI: Northpointe Institute for Public Management.
- Burgess, E.W. (1928). Factors determining success or failure on parole. In A. A. Bruce, A. J. Harno, E. W. Burgess, and J. Landesco (Eds.) *The Working of the Indeterminate Sentence Law and the Parole System in Illinois*. Springfield, Illinois, State Board of Parole: 205-249.

- Campbell, M., French, S., & Gendreau, P. (2009). The prediction of violence in adult offenders: A meta-analytic comparison of instruments and methods of assessment. *Criminal Justice and Behavior*, 36, 567-590.
- Connelly, M. (2003). Critical examination of actuarial offender-based prediction assessments: Guidance for the next generation of assessments. *Dissertation Abstracts International*, 64(12), 4635.
- Dernevik, M., Grann, M., & Johansson, S. (2002). Violent behaviour in forensic psychiatric patients: Risk assessment and different risk management levels using the HCR-20. *Psychology, Crime, and Law*, 8, 83-111.
- Douglas, K., Ogloff, J., Nicholls, T., & Grant, I. (1999). Assessing risk for violence among psychiatric patients: The HCR-20 Violence Risk Assessment Scheme and the Psychopathy Checklist: Screening Version. *Journal of Consulting and Clinical Psychology*, 67, 917-930.
- Douglas, K. S., & Webster, C. D. (1999). Predicting violence in mentally and personality disordered individuals. In R. Roesch, S.D. Hart, & J. R. P. Ogloff (Eds.), *Psychology and law: The state of the discipline* (pp. 175-239). New York: Plenum.
- Eisenberg, M., Bryl, J., & Fabelo, T. (2009). Validation of the Wisconsin department of corrections risk assessment instrument. *Council of State Governments Justice Center*.
- 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, 397-426.
- Gendreau, P., Little, T., & Goggin, C. (1996). A meta-analysis 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, 19-30.
- Hare, R. D. (1991). *Manual for the revised Psychopathy Checklist*. Toronto, Ontario, Canada: Multi-Health Systems.
- Hare, R. D. (2003). *The Hare Psychopathy Checklist—Revised manual* (2nd ed.). Toronto, Ontario, Canada: Multi-Health Systems.
- Hart, S. D. (1998). The role of psychopathy in assessing risk for violence: Conceptual and methodological issues. *Legal and Criminological Psychology*, 3, 121-137.
- Harris, G. T., Rice, M. E., & Quinsey, V. L. (1993). Violent recidivism of mentally disordered offenders: The development of a statistical prediction instrument. *Criminal Justice and Behavior*, 20, 315-335.
- Harris, P. (1994). Client management classification and prediction of probation outcome. *Crime & Delinquency*, 40, 154-174.
- Heaton, P. (2010). *Hidden in plain sight: What cost-of-crime research can tell us about investing in police*. Santa Monica, CA: RAND.
- Henderson, H. & Miller, H. (2011). The (twice) failure of the Wisconsin risk needs assessment instrument. *Criminal Justice Policy Review*, 24(2), 199-221.

- Hilton, N. Z., Harris, G. T., & Rice, M. E. (2006). Sixty-six years of research on the clinical versus actuarial prediction of violence. *Counseling Psychologist*, 34 (3), 400-409.
- Howell, T. (n.d.). *LSI-R, LS/RNR and LS/CMI Documentation*. Toronto, Ontario, Canada: Multi-Health Systems. Retrieved April 19, 2013 from [https://docs.google.com/viewer?a=v&q=cache:6KpaF\\_v63ksJ:www.scstatehouse.gov/archives/citizen\\_sinterestpage/SentencingReformCommission/Miscellaneous/exhibittoDanford100809presentationsLSIdocumentation.doc+&hl=en&gl=us&pid=bl&srcid=ADGEESjGMjyHxyBUwJV1xncplUzX1OS0A1q3Aq9qr2CWqhEJnmNtuScyBiDOWojghOMTvNgGCxc4NfBTnDBytAGvd6M9Pk1YvYo\\_Gx20b44j2l62Re1u\\_GwdhLfH2oXPJaPTk8FJMj&sig=AHIEtbTJwMDdLbqMOTn9rY6gAQQhwgD-yg](https://docs.google.com/viewer?a=v&q=cache:6KpaF_v63ksJ:www.scstatehouse.gov/archives/citizen_sinterestpage/SentencingReformCommission/Miscellaneous/exhibittoDanford100809presentationsLSIdocumentation.doc+&hl=en&gl=us&pid=bl&srcid=ADGEESjGMjyHxyBUwJV1xncplUzX1OS0A1q3Aq9qr2CWqhEJnmNtuScyBiDOWojghOMTvNgGCxc4NfBTnDBytAGvd6M9Pk1YvYo_Gx20b44j2l62Re1u_GwdhLfH2oXPJaPTk8FJMj&sig=AHIEtbTJwMDdLbqMOTn9rY6gAQQhwgD-yg)
- Justice System Assessment & Training (2010). *Wisconsin Risk & Needs tool (WRN)*. Retrieved from: <http://www.jsat.com/Toolkit/Adult/adf6e846-f4dc-4b1e-b7b1-2ff28551ce85>
- Kroner, D., & Mills, J. (2001). The accuracy of five risk appraisal instruments in predicting institutional misconduct and new convictions. *Criminal Justice and Behavior*, 28, 471–489.
- Lowenkamp, C. T., Latessa, E. J., & Holsinger, A. M. (2004). Empirical evidence on the importance of training and experience in using the Level of Service Inventory-Revised. *Topics in Community Corrections*, 49-53.
- Loza, W. (1996) *Self-Appraisal Questionnaire (SAQ): A tool for assessing violent and nonviolent recidivism*. Unpublished manuscript.
- Loza, W., & Loza-Fanous, A. (2001). The effectiveness of the Self-Appraisal Questionnaire (SAQ) in predicting offenders' post-release outcome: A comparison study. *Criminal Justice and Behavior*, 28, 105-121.
- Loza W., MacTavish A., Loza-Fanous A. (2007). A nine-year follow-up study on the predictive validity of the Self-Appraisal Questionnaire for predicting violent and nonviolent recidivism. *Journal of Interpersonal Violence*, 22, 1144-1155.
- Loza, W., Neo, L. H., Shahinfar, A., & Loza-Fanous, A. (2005). Cross-validation of the Self-Appraisal Questionnaire: A tool for assessing violent and non-violent recidivism with female offenders. *International Journal of Offender Therapy and Comparative Criminology*, 49, 547-560.
- Little, G. L., Robinson, K. D., & Burnette, K. D. (1992) Cognitive-behavioral treatment for offenders: the successful approach of Moral Reconciliation Therapy. *The IARCA Journal on Community Corrections*, 5-8.
- Manchak, S., Skeem, J., & Douglas, K. (2007). Utility of the Revised Level of Service Inventory (LSI-R) in predicting recidivism after long-term incarceration. *Law and Human Behavior*, 32, 477-488
- Martinson, R. (1974). What works? Questions and answers about prison reform. *The Public Interest*, 35, 22-54.
- Meehl, P. E. (1954). *Clinical vs. statistical prediction: A theoretical analysis and a review of the evidence*. Minneapolis: University of Minnesota Press.
- Miller, H. (2006). A dynamic assessment of offender risk, needs, and strengths in a sample of pre-release general offenders. *Behavioral Sciences & the Law*, 24, 767-782.

- Miller, T. R., Cohen, M. A., & Wiersema, B. (1996). *Victim Costs and Consequences: A New Look*. Department of Justice, Office of Justice Programs, National Institute of Justice. Washington, D.C.: U.S.
- Monahan, J. (1981). *The clinical prediction of violent behavior*. Rockville, MD: National Institute of Mental Health.
- Northpointe Institute for Public Management. (n.d.) Retrieved from:  
<http://www.northpointeinc.com/publications>
- Nuffield, J. (1982). *Parole decision making in Canada: Research towards decision guidelines*. Ottawa, ON: Solicitor General of Canada.
- Office of the Legislative Auditor General State of Utah. (2004). *Performance audit of the division of adult probation and parole: Report to the Utah Legislature Number 2013-08*. Salt Lake City, UT: Author.
- Peterson, B., Hickert, A. O., & Dorsey, N. (2008). *Factors related to parole violations and revocations: Analysis of the Utah parole system and outcomes*. Salt Lake City, UT: Utah Commission on Criminal and Juvenile Justice (CCJJ) and Utah Criminal Justice Center (UCJC).
- Raynor, P., Kynch, J., Roberts, C., & Merrington, S. (2000). *Risk and need assessment in probation services: An evaluation*. Home Office Research Study Number 211. London, England: Home Office.
- Salekin, R. T., Rogers, R., & Sewell, K. W. (1996). A review and meta-analysis of the Psychopathy Checklist and Psychopathy Checklist-Revised. *Clinical Psychology: Science and Practice*, 3, 203-215.
- Smith, P., Cullen, F. T., & Latessa, E. J. (2009). Can 14,737 women be wrong? A meta-analysis of the LSI-R and recidivism. *Criminology and Public Policy*, 8, 183-208.
- Steadman, H. J. & Cocozza, J. J. (1974) *Careers of the criminally insane: Excessive social control of Deviance*. Lexington, MA: Lexington Books.
- Vose B., Lowenkamp C. T., Smith P., Cullen F. T. (2009). Gender and the predictive validity of the LSI-R: A study of parolees and probationers. *Journal of Contemporary Criminal Justice*, 25, 459-471.
- Webster, C. D., Douglas, K. S., Eaves, S. D., & Hart, S. D. (1997). Assessing risk of violence to others. In C. D. Webster & M. A. Jackson (Eds.), *Impulsivity: Theory, assessment, and treatment* (pp. 251-277). New York: Guilford.
- Wilson, D. B., Bouffard, L. A., & MacKenzie, D. L. (2005). A quantitative review of structured, group-oriented, cognitive-behavioral programs for offenders. *Criminal Justice & Behavior*, 32, 172-204.
- Wormith, J. S., Olver, M. E., Stevenson, H. E., & Girard, L. (2007). The long term prediction of offender recidivism using diagnostic, personality, and risk/need approaches to offender assessment. *Psychological Services Review*, 17, 330-342.