A REVALIDATION OF THE LEVEL OF SERVICE INVENTORY–REVISED (LSI-R)

A Thesis
Submitted to the Graduate Faculty
of the
North Dakota State University
of Agriculture and Applied Science

By

Crystal Murani Towers

In Partial Fulfillment of the Requirements
for the Degree of
MASTER OF SCIENCE

Major Program:
Criminal Justice

December 2020

Fargo, North Dakota

North Dakota State University
Graduate School

Title

A REVALIDATION OF THE LEVEL OF SERVICE INVENTORY–REVISED (LSI-R)

By

Crystal Murani Towers

The Supervisory Committee certifies that this ***disquisition*** complies with North Dakota State University's regulations and meets the accepted standards for the degree of

MASTER OF SCIENCE

SUPERVISORY COMMITTEE:

Dr. AJ Myer
Chair

Dr. Kevin Thompson

Dr. Kjersten Nelson

Approved:

12/14/2020 — Date

Dr. Carol Archbold — Department Chair

ABSTRACT

In the United States, the large number of incarcerated individuals presents heavy social and economic burdens. To lessen these strains, many criminal justice agencies utilize risk assessment to determine which individuals are at a higher risk of recidivating and allocate limited intervention resources accordingly. To ensure these interventions are being delivered to those persons most in need, these risk assessment instruments must be tested for predictive validity. The present research seeks to revalidate one such risk assessment tool, the Level of Service Inventory Revised (LSI-R), on an adult offender population of a Midwest state. Additionally, this research expands on previous LSI-R validation studies by assessing the predictive validity of the LSI-R on an understudied population, Native Americans. The analyses utilized in this research include univariate descriptive statistics, bivariate correlation, and Receiver Operator Characteristic/Area Under the Curve analysis. Results are presented. Policy implications and recommendations for future research are discussed.

ACKNOWLEDGEMENTS

Throughout the writing of this thesis I have received a tremendous amount of support and assistance. I would first like to acknowledge the North Dakota Department of Corrections and Rehabilitation for providing me with the opportunity to conduct this research, the data that was utilized for this research, and for the continual professional support I relied on throughout the data collection and cleaning process.

I would like to express my gratitude to my advisor and committee chair, Dr. AJ Myer, who has been an unmatched source of both expertise and morale boosting. Your guidance led me to sharpen my work and brought this thesis to a higher level. I would also like to extend my gratitude to the other members of my committee, Dr. Kevin Thompson and Dr. Kjersten Nelson, for their thoughtful engagement and recommendations.

Additionally, I would like to thank my family for their confidence in me, especially in moments of doubt I had about myself. I would like to thank my parents in particular for their wisdom, perspective, and untiring sympathetic ears. I am so blessed to have you. Finally, I would not have completed this thesis with my good humor intact without the support of my best friend. Kaitlyn, you're the realest one there is.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1: INTRODUCTION

Over the last 40 years, the United States has increased its prison population, resulting in a per-capita incarceration rate of 655 per 100,000 of the national population, the highest in the world (World Prison Brief, 2020). The tremendous social and economic costs associated with mass incarceration have led to discussions on how to strategically divert some criminal justice-involved individuals away from incarceration to other forms of correctional supervision (James, 2018). In selecting who should receive diversion back into the community, there must be consideration for the likely risk a given offender will have for future criminal activity. This notion of predicting risk is where the role of risk assessment tools become salient in corrections. Per Latessa and Lovins (2010), offender assessment serves multiple functions, including the identification of individuals most at risk for recidivating, identification of those individuals requiring intervention and the crime producing needs that should be intervention targets. Additionally, the quantitative nature of offender assessment aids in systematic decision making which also serves to reduce bias against offenders. Finally, the use of assessment tools addresses economic costs concerns arising from mass incarceration by increasing the appropriate placement and resource allocation for offenders, ultimately enhancing outcomes for offenders and thereby increasing safety for the general public (Latessa and Lovins, 2010).

History of risk assessment

Risk assessment has been used in criminal justice for the better part of a century, with origins in the United States circa 1920 (Johnson et al., 2011). Over the years, risk assessment has evolved through several generations, to arrive at the actuarial tools frequently used today.

First generation assessments

Bonta (1996) described first generation assessments as relying primarily on the professional judgement or gut feelings to arrive at a subjective assessment regarding the offender's risk. This method comprises the collection of offender history paired with an unstructured interview between the offender and a correctional or clinical practitioner (Bonta, 1996). After reviewing the offender's documentation and interview responses, the professional delivers recommendations for the offender. The primary weakness found in first generation assessments is the largely discretionary quantitatively elusive nature of the subjective assessment process. Because the interviews are unstructured, with questions changing across interviewers, recommendations for a given offender could also vary (Bonta, 1996; Johnson et al., 2011). Additionally, the accuracy of subjective assessment predictions in forecasting future offending has little empirical support (Bonta, 1996).

Second generation assessments

In addressing the quantitative issues of the first generation, second generation risk assessment began the utilization of actuarial methods over professional judgement (Johnson et al., 2011). In this approach, variables that have been established to increase the likelihood of reoffending are given numerical scores, allowing for an aggregate score to be given to offenders based on standardized items. In second generation assessments, each item to be scored is given the same weight, rather than allowing for predictors to be weighted by impact. Second generation assessments are limited by their atheoretical nature, resulting in the neglect of theoretically demonstrated factors related to criminal conduct (Andrews and Bonta, 2003). Additionally, scales in second generation assessments are mainly static, meaning that the items, while useful for differentiating between low and high-risk offenders, do little in the way of providing direction for treatment (Bonta, 1996).

Third generation assessments

To remedy the missing future-orientation piece of second-generation assessments, third generation assessments are constructed under the notion of measuring change (Bonta, 1996). Third generation tools further distinguish themselves from second generation assessments by systematically and objectively measuring offenders needs (Andrews and Bonta, 2003). In this generation of assessment, there are dual goals of measuring risk and determining offender treatment needs to manage risk. This is accomplished through the inclusion of the static risk factors of second-generation assessments, now joined by dynamic risk factors, which are more sensitive to change (and thus more malleable to treatment) given an offender's circumstances (Johnson et al., 2011).

Fourth generation assessments

Fourth generation assessments are informed by a series of information including static and dynamic risk factors, responsivity and protective factors, and case plans (Latessa, 2015). Fourth generation risk assessment instruments allow for the measurement of criminogenic risk factors, where criminogenic is understood as factors which may contribute to criminal behavior (James, 2018). Assessment instruments measuring criminogenic risk factors allow for individuals to be assigned to a risk category, generally on a scale of "low," "medium," or "high" risk.

How risk assessments work

As can be gathered, a major purpose of offender risk assessment is classification. As defined by Andrews and Bonta (2003) classification is the process of identification of like offenders into subgroups and assignment of certain interventions based on subgroup. Thus, the utility of risk assessment classification rests in the ability of correctional systems to manage the treatment and services for large numbers of offenders in the most efficient manner possible

(Andrews and Bonta, 2003; Gottfredson and Moriarty, 2006). This efficient service delivery and management is achieved through correctional systems allocating interventions based on the needs and risks of offenders, which are determined through classification. Risk assessments of the third and fourth generations arrive at this classification through the adherence to a system of three principles: risk, need, and responsivity.

Risk principle

The risk principle states that higher levels of service should be reserved for higher risk cases, while lower risk cases should receive less intensive to minimal services (Andrews and Bonta, 2003; Lowenkamp, Holsinger, Brusman-Lovins, & Latessa, 2004). In essence, referencing an individual's risk level for reoffending dictates which, if any, correctional interventions should be provided for that offender (Andrews and Bonta, 2003). Programs that assign interventions based on the results of risk assessment instruments have significantly greater effects in reducing recidivism compared to programs that do not adhere to the risk principle (Latessa, 2004). Therefore, the risk principle dictates that it is necessary to measure risk through an objective risk assessment instrument that has established validity and reliability. Decisions for offenders should be based on that individual's risk level. Thus, high risk clients should receive the highest intensive services as they have more risk categories that can be targeted for treatment (Latessa, 2004). In contrast, because low risk offenders should receive minimal treatment and services, as correctional interventions can inadvertently increase these individuals' risk through associating with high risk offenders and/or disrupting the prosocial factors in their lives that keep them low risk (Latessa, 2004).

Need principle

The risk classification of offenders is determined through the assessment of risk factors, which are the personal characteristics and circumstances that are predictive of future offending (Andrews et al., 1990). Risk factors are understood to fall into two categories: static and dynamic. Static risk factors are those attributes related to crime which cannot be influenced by treatment, e.g. age at first arrest or gender (James, 2018). Dynamic risk factors refer to those crime-producing or “criminogenic” factors that can be changed through intervention (Latessa and Lovins, 2010). Andrews and Bonta (2003) identified eight major risk factors that should be addressed for services: antisocial/procriminal attitudes, values, and beliefs; procriminal peers and isolation from prosocial peers; temperament and personality factors like impulsivity, risk-taking, and pleasure seeking; history of antisocial behavior; family history of criminality or lack of care; low levels of educational or vocational achievement; lack of prosocial leisure activities; drug and alcohol abuse. The need principle posits that services should target criminogenic needs of offenders, those dynamic risk factors that are highly correlated with criminal behavior. Of the previously listed eight major risk factors several are criminogenic in nature such as antisocial attitudes, antisocial peers, personality factors, and substance abuse. Programs built on the need principle target those dynamic risk factors yield more effective correctional intervention programs when compared to those programs developed to address noncriminogenic needs (Dowden and Andrews, 1999).

Responsivity principle

The responsivity principle broadly states that in addition to considering risk and need factors, effective treatment should also be delivered in a way that is consistent with the ability and learning style of the offender receiving treatment (Latessa and Lovins, 2010; James, 2018). Responsivity can also be split into categories of general and specific. The general responsivity principle posits that for effecting change in the target areas previously stated, cognitive-behavioral and social learning therapies are the most effective form of intervention (Bonta and Andrews, 2007; James, 2010; Landenberger and Lipsey, 2005; Lipsey, Landenberger, and Wilson, 2007; Lipsey, 2009). The specific responsivity principle states that the personal attributes of the individual offender should be assessed as well. This is because areas like mental and emotional issues, cognitive functioning, and motivation to change all vary by person and can impact the individual's engagement in treatment (Latessa and Lovins, 2010). When a program incorporates the principles of risk, need, and responsivity, it is more likely to serve as an effective intervention for reducing recidivism when compared to programs that do not utilize these principles (Dowden and Andrews, 1999; Latessa and Lovins, 2010).

Since research has demonstrated that risk can be measured and categorized, there has been a proliferation in efforts to create tools for a multitude of offender types and subtypes (Andrews et al., 2006). Many new assessments have been developed with specific crime producing issues in mind, built in the school of the previously described third and fourth generations: COMPAS (Blomberg et al., 2010; Brennan, Dieterich, and Ehret (2009), ORAS (Latessa et al., 2009; Lovins et al., 2018), PCRA (Johnson et al., 2011), and the tool of present interest, the LSI-R (Flores et al., 2007; Chenane et al., 2015).

Purpose of present study

The present research seeks to conduct a revalidation of the Level of Service Inventory – Revised (LSI-R) against an actual adult offender population of a Midwest state. The LSI-R was designed to be a third generation general actuarial tool, meaning it should have predictive validity across a broad swath of the general offending population, from males and females to Whites and non-Whites (Andrews and Bonta, 2010; Andrews et al., 2006; Holtfreter and Cupp, 2007; Luong and Wormith, 2011; Manchak, Skeem, Douglas, and Siranosian, 2009; Whiteacre, 2006). The LSI-R is a widely used risk assessment tool in the United States and Canada and has been validated nationally on multiple samples (Flores et al., 2006; Folsom and Atkinson, 2007; Schlager and Simourd, 2007; Fass et al., 2008; Austin, 2011; Chenane et al., 2015; Ellison et al., 2016; Olver, Stockdale, and Wormith, 2014; Ostermann and Salerno, 2016; Vose et al., 2009; 2013). Given the lack of consensus on the predictive validity for subpopulations (Vose et al., 2009; Chenane et al., 2015; Holsinger, Lowenkamp, and Latessa, 2006; Ostermann and Salerno, 2016; Wilson and Gutierrez, 2014), this study will attempt to test the predictive validity of the LSI-R on both a general sample, and understudied subpopulation—Native Americans. As demonstrated by several studies (Schlager and Simourd, 2007; Chenane et al., 2015; Ellison et al., 2016), the subcomponents of the LSI-R also vary in their predictive validity. The research questions are thus: What is the predictive validity of the LSI-R on a Midwest adult population? What is the predictive validity of the LSI-R on a substantial Native American population?

CHAPTER 2: VALIDITY OF THE LSI-R AND OTHER TOOLS IN USE

Description of LSI-R

History

Andrews' (1982) original Level of Supervision Inventory (LSI) was composed of 58 items that surveyed offender characteristics and circumstances relevant for level of supervision decisions. The items were scored as "0-1" and were grouped into 11 different risk-related categories: criminal history (4), education/employment (10), financial (2), family/marital (4), accommodation (3), leisure/recreation (2), companions (5), alcohol/drug problems (9), emotional/personal (5), probation/condition (4), attitudes/orientation (4) (Andrews, 1982). The LSI matched or surpassed the predictive ability of other contemporary measures for each risk category upon evaluation (Andrews, 1982). Andrews' LSI-R (1995) revised the LSI by removing the 4 items measuring probation/condition, resulting in the 10 subcategory, 54 item assessment currently in use.

Overview of LSI-R

The LSI-R is administered to the offender via structured interview by an assessor trained and certified in LSI-R administration. Andrews and Bonta (1995) also recommend the collection of supporting documents from multiple sources e.g. family members, employers, case files, and drug tests. Once the assessment is administered, the 54 items can be scored, and a classification given. The overall LSI-R classification works by tallying a raw score from the 54 items, which is then plotted on a risk scale ranging from 0-54. The higher the raw score, the higher the overall classification level and the greater likelihood of recidivism. The categories range broadly from low to high risk/need. Per Andrews and Bonta (1995), the breakdown of each category and cut-off score can be referenced below: low risk (0-13), low-moderate risk (14-23), moderate risk (24-33), high-moderate risk (34-40), and high risk (41-54)

The individual LSI-R score can also be actuarially assigned a recidivism likelihood based on the actual observed recidivism rates of offenders who were previously assessed using the LSI-R (Flores et al., 2006). Additionally, the subcategory scores of the LSI-R are used to identify what the offender's greatest areas of risk/need are. This is useful for informing assessors and providers on which treatment is most suitable for the offender (Andrews and Bonta, 1995; Flores et al., 2006).

While these are the cut-off scores that have been standardized and validated in previous studies, agencies standardize cut-off scores to adjust to their needs. The ND DOCR uses the following cut-off scores for LSI-R risk categories: low risk (0-19), low-moderate risk (20-23), moderate risk (24-29), high-moderate risk (30-38), and high risk (39-54).

Importance of validation

Importance of revalidation

For the benefits of an assessment tool like the LSI-R to be fully realized, it must have empirical support. This means reviewing the benefits observed in practice with those previously identified in research evaluations (Flores et al., 2006). Lovins et al. (2018) identified several reasons highlighting the importance of risk assessment tool validation. First, if a risk assessment tool is not valid and thus unable to predict recidivism and likelihood to reoffend, the usefulness of the risk assessment is undercut. Failure in this area means a tool will be unable to place individuals into appropriate levels of service by their need, resulting in a misuse of resources, a greater likelihood of recidivism, and an increased risk to public safety (Lovins et al., 2018). Another benefit to the repeated revalidation of risk assessment tools relates to their refinement. The more a given risk assessment tool is studied and revised, the greater its validity will be across populations, allowing for easier adoption by agencies with fewer legal challenges (Latessa and Lovins, 2010; Lovins et al. 2018).

Importance of validation across populations

Clear (1997) notes that the cross-jurisdictional and populational validity of risk assessment tools is of ever-increasing importance as more criminal justice agencies adopt the use of such tools. While many risk assessment tools may claim to be valid against a general criminal justice population, it is necessary to ensure that these tools remain valid when tested against unique populations. Lovins et al. (2018) states the importance of valid risk assessment tools particularly in the context of sentencing considerations, to ensure equitable outcomes across populations.

As previously alluded to, risk assessment tools are growing in the span of their use along with the occasion for use. While the LSI-R is the tool in question of the present study, the state of its validation efforts is be better understood in context and comparison of other offender risk instruments (Andrews and Bonta, 2003). Therefore, the following section will provide an overview of risk assessment instruments currently in use, commenting on each tool’s level of demonstrated validity. The section will end with the LSI-R, which is one of the most widely used assessment instruments, requiring consistent evaluation of its predictive accuracy (Olver, Stockdale, and Wormith, 2014). The final section of this chapter will include a review of existing literature on the predictive validity for Native American populations, demonstrating the paucity of research in this area.

Validity of risk assessment tools in use

COMPAS

Correctional Offender Management Profiling for Alternative Sanctions (COMPAS) is an entirely web-based, fourth-generation risk assessment tool developed by the Northpointe Institute for Public Management (Brennan, Dieterich, and Ehret, 2009; Blomberg et al., 2010). Used throughout the United States, COMPAS uses information obtained through official records, standardized interviews with clients, and self-report questionnaire information provided by clients to direct pretrial release, probation, community corrections, institutional programming, reentry, and parole decisions (Blomberg et al., 2010). COMPAS is composed of 22 scales: criminal involvement, history of non-compliance, history of violence, current violence, criminal associates/peers, criminal opportunity, leisure and recreation, social isolation, substance abuse, criminal personality, negative social cognitions, criminal thinking observation, criminal thinking self-report, anger/violence, cognitive behavior, family criminality, socialization failure, financial problems, vocational/educational profile, social environment, residential instability, and social adjustment problems (Skeem and Louden, 2007; Blomberg et al., 2010). Results from these scales provide risk level scores ranging from one to ten for three separate outcomes: risk of violence, risk of recidivism, and risk of failure to appear (Blomberg, 2010). While fourth-generation assessment tools are relatively new, some validation studies have been conducted on these emerging tools.

Brennan, Dieterich, and Ehret (2007) reported the findings of several recent COMPAS a validation studies in Ohio, New York, and California. The Ohio study used a sample of 493 participants from an Ohio community corrections center, with an outcome measure of program failure. Results indicated that overall, COMPAS had high predictive validity, with risk of violence, risk of failure to appear, and risk of recidivism all significant predictors of program failure

(Brennan et al., 2007). New York's study on 393 COMPAS assessed offenders looked at two outcome measures: new crime and days until new crime occurred. Similarly to the Ohio study, the risk of recidivism scale had significant predictive validity, with an area under curve (AUC) score of .72 (Brennan et al., 2007). The California validation study used a sample of 20,898 California parolees to determine at the predictive validity of COMPAS against an outcome measure of parole failure. While the average follow-up was 6-months in length, results indicated the risk of recidivism scale and risk of violence scale significantly predicted risk of parole failure, in line with the other cited by Brennan et al. (2007).

Brennan, Dieterich, and Ehret (2009) also examined the reliability and predictive validity of COMPAS for a community-based sample of males and females. A sample of 2,328 probationers entering 18 county-level probation agencies from the years 2001-2004 received the COMPAS assessment. Recidivism was measured through three outcomes: an arrest for any offense, an arrest for a person offense, and an arrest for a felony offense. Brennan et al. (2009) used Cox proportional hazard models to measure the risk of failure, in this case, recidivism. Results indicated that for both males and females, the COMPAS scales generally had strong (>.70) reliability. Excluding current violence, financial problems, and residential instability, all COMPAS scales significantly predicted felony recidivism. COMPAS had similar predictive validity for both males and females, with high AUC values ranging from .71 to .80 (Brennan et al., 2009).

Farabee, Zhang, Roberts, and Yang (2010) also completed a test of the predictive power of COMPAS to measure future recidivism on a sample of California parolees. Recidivism was measured as two outcomes: subsequent arrest for any reason following release, and a subsequent arrest for a violent offense, including homicide, assault, sexual assault, robbery, domestic violence, and kidnapping. During the 24-month follow-up period, only the COMPAS recidivism and violence scales significantly predicted re-arrests, and only the recidivism scale achieved the .70 AUC threshold (Farabee et al., 2010).

ORAS

The Ohio Risk Assessment System (ORAS) was conceived as a set of tools that conformed to the principles of effective intervention which addressed the risk and needs of Ohio offenders at multiple stages in the criminal justice system including pretrial, community supervision, institutional intake, and community reentry levels (Latessa, Lemke, Makarios, Smith, and Lowenkamp, 2009; 2010). Within the ORAS, there are five assessment tools for use at each criminal justice stage: The Pretrial Assessment Tool (PAT), The Community Supervision Tool (CST), The Community Supervision Screening Tool (CSST), The Prison Intake Tool (PIT), and the Reentry Tool (RT). Between the five tools, there are a total of 63 risk items, with 24 of those items appearing on two or more of the instruments (Latessa et al., 2010). The ORAS has been tested and found to have predictive validity concurrent with other established tools like the LSI-R (Latessa et al., 2009).

Latessa et al. (2010) utilized a prospective design to conduct a validation study against four of the ORAS tools: the PAT, CST, PIT, and RT. The total sample of 1,830 Ohio offenders was broken down by tool: 452 (pretrial), 678 (community supervision), 423 (prison intake), and 277 (community reentry). The outcome measure of recidivism was arrest for new crime, and follow-up data were collected on the participants for one year from the point of study inclusion (Latessa et al., 2010). Results demonstrated all assessment tools were able significantly distinguish between risk levels, with r value ranging from .22 to .44. These findings provided strong preliminary support for the ORAS, with generalizability limitations, as the sample underrepresented female and Hispanic offenders (Latessa et al., 2010).

Building on Latessa et al. (2010), Lovins, Latessa, May, and Lux (2018) validated the ORAS against a Texas population. In doing so, this study addressed developments in risk assessment literature suggesting that fourth-generation tools may be particularly vulnerable to threats to validity when tested against novel populations (Lovins et al., 2018). The ORAS CST was tested against 5,482 offenders across 13 different Community Corrections and Supervisions Departments in Texas. Additionally, several items were added to the tool resulting in a Texas-specific edition (TRAS). The TRAS added clarifying language to the criminal history section to account for differences in legal definitions between Ohio and Texas. Additionally, the TRAS included an item about legal employment status to address undocumented population differences between Texas and Ohio. Recidivism was measured as any new arrest for a criminal act or revocation for law/technical violation, and participants were tracked between 12-18 months from the time of assessment. Results indicated that while the ORAS and TRAS both significantly predicted risk of failure, the TRAS had a higher AUC value that the ORAS (.679 to .664, respectively), supporting the conclusion that risk assessment tools should be validated across populations and modified where necessary (Lovins et al., 2018).

PRCA

The Federal Post-Conviction Risk Assessment (PCRA) instrument is a fourth-generation tool developed to predict recidivism for offenders on federal supervision (Cohen, Lowenkamp, and Robinson, 2018). The PCRA is scored through two sections, the officer assessment and the offender assessment. While the offender assessment informs officers about an offender's criminogenic thinking patterns, it is not used to calculate the risk score. The officer assessment consists of 15 scored factors that measure risk across several domains: criminal history, education/employment, substance abuse, social networks, and cognitions (Cohen et al., 2018).

Scores range from 0 to 18 and fall into one of four risk categories: low (0–5), low/moderate (6–9), moderate (10–12), or high (13 or above). There have been a few major efforts to validate the PCRA (Cohen et al., 2018).

Lowenkamp, Holsinger, and Cohen (2015) conducted a validation study of PCRA using assessments completed by U.S. probation officers on a sample of 113,281 offenders during the course of supervision. The recidivism outcome was measured as re-arrest for any new criminal conduct and re-arrest for violent offenses at six, 12, 18, and 24-month intervals Lowenkamp et al., 2015). Findings generally supported the PCRA as a valid predictor for recidivism. PCRA categorization significantly predicted re-arrest for violence, with AUC values ranging from .70 to .77 depending on the length of follow-up time observed (Lowenkamp et al., 2015).

Luallen, Radakrishnan, and Rhodes (2016) also conducted a study on the PCRA's predictive validity and found similar results to Lowenkamp et al. (2015). Luallen et al. (2016) used a sample of 139,239 offenders received into federal community supervision from 2004 to 2013, with the PCRA readministered on 6-month and 12-months windows. Recidivism was measured as any new arrest within 6-months of PCRA administration in the following areas: Violent, Property, Drug, Sex Offense, Firearms (e.g., possession of weapon), Escape/Obstruction (e.g., perjury), Public Order (e.g., drunk and disorderly), Technical (e.g., failing to appear in court), Immigration (e.g., facilitating an illegal entry), and Other (e.g., refusal to pay court fines). Results showed that the PCRA had strong predictive validity, such that at 12-month follow-up AUC values averaged between .73 and .74 (Luallen et al., 2016). Additionally, the PCRA demonstrated significant predictive validity for multiple recidivism outcomes including offenses including drug, violent, and property offenses (Luallen et al., 2016).

LSI-R

The Level of Service Inventory – Revised (LSI-R) is a third-generation risk assessment tool which, as previously described, utilizes both static and dynamic risk factors to determine offender needs and which treatment are to be deployed. The LSI-R is particularly effective as a tool because it gives equal importance to the measurement of both types of risk factors and because it accounts for offender risk and need through its items (Andrews, 1982). Additionally, the LSI-R is theoretically based in multi-theory frameworks built upon general personality and cognitive social learning perspectives, which perform well when assessed for theory adequacy (Andrews and Bonta, 2003).

Flores et al. (2006) studied the predictive accuracy (validity) of the LSI-R on a sample of 2,107 male federal probationers. The individuals' LSI-R scores served as the independent variable while recidivism—operationalized as incarceration—served as the dependent variable A cross-sectional correlational design was used to measure the predictive validity of the LSI-R on recidivism by comparing the participants' LSI-R scores to subsequent incarceration measured once at a two-year follow-up. This method of calculating a correlation coefficient is common in validity research but does not allow for comparison of predictive validity across samples. To address this issue, Flores et al. (2006) conducted a receiver operator characteristic (ROC) analysis which produces a statistic without sample-specific selection ratios and base rates. In both analyses Flores et al. (2006) found similarly robust predictive validity for the LSI-R, with there being a 68.9 percent chance that a randomly selected recidivist would have a higher LSI-R score than a randomly selected nonrecidivist. In a final multivariate analysis, the LSI-R maintained strong predictive validity when controlling for age, sex, and ethnicity. Flores et al. (2006) utilized a cross-sectional design, so valuable data such as length of time until failure were unavailable. The recidivism measure was limited to only incarceration at the two-year follow-up, which may be conservative compared to other recidivism measures like re-arrest or reconviction. Additionally, the recidivism measure did not provide information on what the offense was when recidivism occurred. Also, this study did not test the LSI-R against any subgroup populations.

Folsom and Atkinson (2007) attempted to expand the literature on the predictive validity of LSI-R for an understudied subgroup, female offenders. Folsom and Atkinson (2007) utilized a cross-sectional correlational design to measure the predictive validity of a self-report version of the LSI-R on a sample of 100 serious female offenders serving sentences of at least 2 years, with an average sentence length of 8 years. Participants had released and at risk to recidivate for an average of 6 years at the time of study follow-up. Building on the limited measures of previous validations studies, recidivism was operationalized any new conviction, and differentiated into violent and nonviolent categories. Like Flores et al. (2006), Folsom and Atkinson (2007) also used ROC analysis and found that the LSI-R self-report had strong predictive validity, with a 67 percent chance that a randomly selected recidivist will have a higher LSI-R score than a randomly selected nonrecidivist. A survival analysis found a significant difference between the low LSI-R scoring individuals when compared to the moderate and high scoring individuals. After logistic regression, the two criminal history items (age at first arrest and number of previous convictions) of the LSI-R largely accounted for the strength in predicting recidivism, as is consistent with previous literature (Folsom and Atkinson, 2007; Quinsey et al., 2006). While the recidivism measure was broken down into multiple categories, Folsom and Atkinson (2007) only utilized the first re-offense, meaning that the follow-up period ceased and no further data were collected on recidivism, even if the individual was rereleased. More recent studies have addressed this issue by measuring the prevalence and incidence of recidivist behavior (Chenane et al., 2015; Ellison et al., 2016).

While the self-report data from this study were corroborated through official file information, Future studies should consider the weakness of using self-report data if not substantiated, due to the possibility of deception, in the present study this was address through the corroboration of information through official files (Folsom and Atkinson, 2007). Additionally, self-report versions of actuarial assessment may mitigate recurring issues found in other methods of assessment, such as interviewer variability and inconsistent record keeping across jurisdictions (Folsom and Atkinson, 2007).

Schlager and Simourd (2007) completed a validation study of the LSI-R among a sample (n = 446) of Black (75%) and Hispanic (25%) parolees enrolled in one of two New Jersey DOC halfway house or one day reporting center community corrections program between 2000-2001. Data were secondary in nature and the recidivism outcome measures were dichotomous rearrest and reconviction variables in a two-year follow-up. Schlager and Simourd (2007) did not find significant correlations between LSI-R scores and either re-arrest or reconviction outcomes for neither Black nor Hispanic participants. However, upon reviewing the predictive validity of the LSI-R subcomponents, education/employment and family/marital scores were significantly higher for Black recidivists than nonrecidivists (Schlager and Simourd, 2007).

Fass et al. (2008) conducted a validation study of the LSI-R and another risk assessment tool, the Correctional Offender Management Profiling for Alternative Sanctions (COMPAS), on a sample of 975 male offenders (COMPAS n = 276; LSI-R n = 696) released from two New Jersey prisons into community halfway houses between 1999-2002. Fass et al. (2008) utilized a retrospective known-groups validation design, similar to the type of data used in other validation studies (Chenane et al., 2015; Ellison, 2016). The participants were randomly selected from a list of all male offenders exiting the two prisons during the 1999-2002 period. Compared to previous validation studies, with majority White samples, this sample consisted of primarily (71.4%) Black and Hispanic or Latino (15%) participants compared to White (13.6%) participants (Fass et al., 2008). This allowed for further investigation of the generalizability claims of the LSI-R. Recidivism was measured as a dichotomous re-arrest variable within 12-months of release into the community, derived from official DOC records. Fass et al. (2008) found the LSI-R score was able to accurately predict re-arrest for the entire LSI-R sample within 12-months and Black participants, but that the LSI-R score was not reliable in predicting re-arrest for White and Hispanic or Latino participants. The COMPAS tool did not yield significant predictive validity for the full COMPAS sample or when broken down by racial/ethnic subgroups (Fass et al., 2008). Fass et al. (2008) utilized a sample drawn from a community setting, rather than a prison setting, which does not directly aid in the correct delivery of services to incarcerated individuals (Chenane et al., 2015). This sampling issue is addressed in other research, such as the validation study by Chenane et al. (2015), which utilized an incarcerated sample. Fass et al. (2008) also employed a short outcome period at 12 months and measured the recidivism outcome as re-arrest which may not capture actual behaviors like self-report data.

Chenane et al. (2015) conducted a validation study for the LSI-R and its ten subcomponents against a Midwest statewide sample of 2,778 male inmates who received custody classification in 2009. This the resulting sample was composed of 1,910 Whites, 672 Blacks, and 196 Hispanics. This study built upon previous validation research through multiple efforts. An incarcerated sample was used, expanding on previous samples which utilized offenders in the community rather than those incarcerated. Then, the usual outcome recidivism measures of re-arrest or reconviction were exchanged for violent and nonviolent misconduct from the first 2-years of incarceration. Chenane et al. (2015) also built on previous studies by measuring the incidence of misconduct, rather than measuring misconduct only as a prevalence measure (Flores et al., 2006; Folsom and Atkinson, 2007). Additionally, the predictive validity of the LSI-R and each of its ten subcomponents were measured through individual equations for each racial/ethnic group.

Chenane et al. (2015) found mixed results for the generalizability of the LSI-R validation across racial/ethnic groups. While the composite LSI-R score had similar predictive validity for violent misconduct across racial/ethnic groups (White [64%], Black [67%], and Hispanic [67%], it was only predictive of White inmates for nonviolent misconduct, failing to predict for Black and Hispanic inmates. There were differences in the predictive validity of individual components by race for both prevalence and incidence of misconduct. In predicting prevalence, Chenane et al. (2015) found that only three subcomponents (financial, accommodation, and motional/personal) predicted the violent misconduct for Black inmates, and only two (companions and emotional/personal) were related to violent misconduct for Hispanic inmates. Similarly for nonviolent misconduct, only one subcomponent (emotional/personal) had significant predictive validity for Black inmates, and no such significance was found of any subcomponents for the Hispanic inmates. In predicting incidence, Chenane et al. (2015) found further disparity. The composite LSI-R score and five subcomponents (criminal history, financial, family/marital, peers/companions, and attitudes/orientations) had stronger predictive validity for White inmates relative to Black inmates. When comparing White and Hispanic inmates, three subcomponents (financial, leisure/recreation, and attitudes/orientation) had greater effect for White versus Hispanic inmates, while two subcomponents (accommodation and emotional/personal) were more predictive for Hispanic than for White inmates. Differences in Black and Hispanic effects were also demonstrated, with the composite LSI-R score, and four subcomponents (criminal history, education/employment, alcohol/drug problems, and the emotional/personal) having weaker predictive validity for Black inmates compared to Hispanic inmates and the leisure/recreation subcomponent demonstrating a stronger effect for Black relative to Hispanic inmates (Chenane et al., 2015).

In a companion study to Chenane et al. (2015), Ellison et al. (2016) conducted a predictive validity test of the LSI-R across different age groups on the prevalence and incidence of inmate misconduct. Age groups were distinguished based on developmental psychology literature, resulting in four classifications: inmates aged 18-24 years old at the time of admission to prison were categorized as emerging adults (n = 789), those aged 25-34 years old were classified as young adults (n = 943), individuals aged 35-44 years old comprised the group in middle adulthood (n = 616), and inmates aged 45-64 made up the group in the late adulthood (n = 456). As with Chenane et al. (2015), misconduct data were collected from official disciplinary data, which may underreport actual misconduct but generally have good criterion validity (Ellison et al., 2016). And as with Chenane (2015), all outcomes were controlled for by length of time spent in prison at time of data collection.

When considering the prevalence of violence and nonviolent misconduct, the LSI-R and each of its ten subcomponents had similar predictive validity for each age group category. While the LSI-R composite score had strong predictive validity for the prevalence and incidence of violent misconduct, and prevalence of nonviolent misconduct, there were differences in predictive validity by subcomponents. Ellison et al. (2016) found that four subcomponents (criminal history, education/employment, alcohol/drug problems, and emotional/personal) differed between inmates in the emerging adulthood category compared to the young adulthood category for predicting the incidence of violent misconduct. Additionally, the predictive validity of the LSI-R composite score and its subcomponents varied for the prevalence and incidence of nonviolent misconduct by inmate age categorization. In particular, the family/marital subcomponent was significant in predicting the prevalence and incidence of nonviolent misconduct for the emerging and middle adulthood age ranges, but not for inmates in the young or late adulthood categories. Conversely, the attitudes/orientation subcomponent had strong predictive validity for inmates in the young and late adulthood categories but did not for inmates under the emerging or middle adulthood categories (Ellison et al., 2016).

Ostermann and Salerno (2016) conducted a validation study of the LSI-R for different combinations of race and gender. This was an expansion on previous attempts to build validation literature, which focused either on race or gender (Schlager and Simourd, 2007; Vose et al., 2009; Chenane et al., 2015). The sample included randomly selected 9,454 offenders (50% male, 50% female) released from New Jersey correctional facilities between 2004 and 2010. Racial/ethnic categories were defined as Black, White, and Hispanic. Ostermann and Salerno (2016) utilized a 1-year follow-up period from time of release and operationalized the recidivism outcome variable recidivism as a re-arrest or a parole violation resulting in revocation within 1-year of release. Data were collected from official state records. Ostermann and Salerno (2016) found that across all races, males had higher recidivism rates when compared with females. When considering combinations of gender and race, the LSI-R had low to moderate predictive validity (fairing below the .70 AUC guidelines for a good distinguisher). In particular, the LSI-R had very low ability to discern Black recidivists from nonrecidivists (Ostermann and Salerno, 2016). Ostermann and Salerno (2016) also selected a short (1-year) follow-up time period in order to include more participants in the sample. While the shortened follow-up helped to increase the subgroup populations of the study, it also limited the researchers’ ability to sufficiently measure recidivism, as with Fass et al. (2008). Additionally, the dichotomous nature of the recidivism variable missed the opportunity to measure the LSI-R's predictive validity of frequency of reoffending (Chenane et al., 2015; Ellison et al., 2016).

Vose et al. (2009) carried out a validation of the LSI-R on a sample of 2,849 male (85.9%) and female (14.1%) probationers and parolees who were given the LSI-R assessment at two points in time. This design is an informative departure from cross-sectional validation studies, as its longitudinal design allows for the analysis of the LSI-R's predictive validity in three ways: at time 1, time 2, and the through the change in scores between Time 1 and 2 (Vose et al., 2009). Additionally, this study built on previously gender homogenous validation literature by comparing the effect of the LSI-R across gender categorizations. The independent variables are the LSI-R scores at Time 1 and 2, with the participants having received the LSI-R twice within the 2000-2005 study period. This resulted in a mean number of 1,384 days at risk for Time 1 and 1,724 days at risk for Time 2. The outcome recidivism measure was operationalized as a dichotomous variable, any new misdemeanor or felony conviction after the first LSI-R administration. Vose et al. (2009) found that the LSI-R was a valid predictor for males and females, and that this did not vary between Time 1 and Time 2. Additionally, the percentage change in LSI-R score between Time 1 and Time 2 was a significant predictor of recidivism for males and females. While Vose et al. (2009) looked at the generalizability of the LSI-R across genders, the study was limited in that its sample was majority male. This resulted in estimates that were likely less precise for females compared to males, thus future studies considering gender comparisons should strive for greater balance in the sample used (Vose et al., 2009).

Utilizing the same sample, data, and variable operationalizations from Vose et al. (2009), Vose et al. (2013) carried out a validation of the LSI-R on 2,849 probationers and parolees who were given the LSI-R assessment at two points in time within the 2000-2005 study period. Having these multiple assessment points allows the researcher to determine the extent to which an individual's risk level changes over time (Vose et al., 2009; Vose et al. 2013). The independent variables are the LSI-R scores at Time 1 and 2 and the dependent variable was recidivism, measured as a dichotomous variable for any new misdemeanor or felony conviction after the first LSI-R administration. The mean time at risk—or days in the follow-up period—for assessment at Time 1 was 1,385 and the mean time at risk for the reassessment was 1,019. Vose et al. (2013) found that the LSI-R was a valid predictor for recidivism, and that a change in the LSI-R score between Time 1 and Time 2 was correlated with recidivism. Additionally, time at risk and LSI-R score were significant predictors of recidivism at both Time 1 and Time 2.

This review of the current LSI-R validation literature indicates that the tool has moderate to strong predictive accuracy general samples of offenders (Flores et al., 2006; Folsom and Atkinson, 2007). Studies that have looked into demographic variations in predictive validity revealed that the LSI-R is particularly valid for male and white offenders (Holsinger, Lowenkamp, and Latessa, 2006; Chenane et al., 2015; Ostermann and Salerno, 2016.) For the female offenders, the LSI-R has demonstrated moderate predictive ability (Folsom and Atkinson, 2007; Vose et al. (2009), although the strength of the tool's validity is more variable for female offenders of different racial groups (Ostermann and Salerno, 2016). The tool demonstrates lower predictive accuracy in minority samples such as those comprised largely of Black or Hispanic offenders (Schlager and Simourd, 2007; Chenane et al., 2015; Ostermann and Salerno, 2016). In the following section, validation literature considering the LSI-R's predictive limitations by racial category is further explored for the present study's racial category of interest, Native Americans.

Risk assessment and Native Americans

Ethnic minorities are overrepresented in correctional populations in the United States, thus it is important that risk assessment instruments are tested on representative populations, to minimize classification errors of minorities (Gutierrez, Helmus, and Hanson, 2016; Whiteacre, 2006). Particularly in the United States, there is a scarcity of research relating to the predictive accuracy of risk tools for indigenous populations (Myer, 2019), as most research in this area has been conducted in Australia (Hsu, Caputi, and Byrne, 2010; Shepherd, Adams, McEntyre, and Walker, 2014) or Canada (Bonta, LaPairie, and Wallace-Capretta, 1997; Jung and Rawana, 1999; Wilson and Gutierrez, 2014; Wormith, Hogg, and Guzzo, 2015). While these studies found limited general predictive validity, they also noted that individual components of the tools were differentially important for indigenous populations.

In a meta-analysis of 32 reports utilizing 12 datasets, Gutierrez et al. (2013) found that tool components relating to criminal history, substance use, and antisocial characteristics were more predictive for non-indigenous individuals when compared to indigenous individuals of the sample. For the LSI-R in particular, Holsinger, Lowenkamp, and Latessa (2006) found that the LSI-R did not have predictive validity for future offending in Native Americans, while it had moderate predictive validity for the Caucasians included in the sample. Based on the existing research available on indigenous populations, there is space and need for further LSI-R validation against Native Americans (Holsinger, Lowenkamp, and Latessa, 2006; Olver, Stockdale, and Wormith, 2014).

Summary and current project

Risk assessment instruments allow correctional agencies to classify offenders and correctly allocate treatments and services to these criminal justice-involved clients based on criminogenic needs and risk of recidivism (Andrews and Bonta, 2003). In order for these tools to be best utilized, they must be repeatedly validated over time, and on local populations to ensure generalizability (Latessa and Lovins, 2010; Lovins et al. 2018). Research on the LSI-R has demonstrated continued predictive validity over time on national samples across different stages of the criminal justice system (Flores et al., 2006; Chenane et al., 2015; Ellison et al., 2016; Olver, Stockdale, and Wormith, 2014; Ostermann and Salerno, 2016; Vose et al., 2009; 2013). However, validation studies of the LSI-R on local samples of minority populations have demonstrated less consensus on the instrument's predictive accuracy (Vose et al., 2009; Chenane et al., 2015; Holsinger, Lowenkamp, and Latessa, 2006; Ostermann and Salerno, 2016; Wilson and Gutierrez, 2014). Native Americans are a subpopulation which continues to reveal mixed to non-significant predictive validity against the LSI-R (Gutierrez et al., 2013). Therefore, the present study intends to revalidate the LSI-R for a state agency, while adding to the minimal existing literature on risk assessment for indigenous populations. Thus, the next chapter outlines the methods used to carry out the revalidation of the LSI-R for a Midwest state agency, utilizing a sample including Native Americans to allow for predictive validity across racial groups.

CHAPTER 3: METHODOLOGY

Sample and data

The sample frame for this study included all individuals either placed on probation or released from in 2016 from an NDDOCR facility. Participants were deemed eligible for inclusion into the sample if they received LSI-R assessment by the NDDOCR. The reasoning for utilizing DOCR releases from 2016 was to allow for a three-year follow-up period, as was used in the previous validation study (Austin, 2011). The sample was separated into two cohorts: 1) community cohort (those placed directly into supervised probation at sentencing) and 2) institutional cohort (those who are discharged from an institution and return to the community on supervision). These cohorts were selected based on cohorts from the previous validation study and selected because additional tracking and data collection is possible for each type of cohort (Austin, 2011). 4,165 individuals were included in the final sample, with 2,700 in the community cohort and 1,465 comprising the institutional sample.

Data for this study were provided by the NDDOCR as part of data that have already been collected by the state as part of their day-to-day procedures. Data were delivered via secure electronic transfer by the NDDOCR for all LSI-R assessed individuals placed on probation or released from an institution in 2016. The LSI-R assessments were administered by NDDOCR staff trained and certified as LSI-R raters. Data included: demographic information (e.g., age, race, gender, etc.), offense information (e.g., offense type), assessment information (LSI-R responses), State ID Number, DOCR ID Number, movements in and out of NDDOCR facilities, and previous criminal history. Recidivism data were collected using the publicly available data from the North Dakota Court Records Inquiry website. With IRB approval, data collection began in October 2020 and lasted for a period of two weeks.

Variables

Independent variables

There are two primary independent variables: 1) The risk score as assigned by LSI-R, which can range from 0-54, measured as a ratio variable and 2) the risk level as assigned by the LSI-R, measured as an ordinal variable. For supplemental analyses on the LSI-R's validity by racial categories, race is also used as a nominal predictor variable (0 = White, 1 = Native American, 2 = Other).

Dependent variables

The dependent variable was recidivism, with multiple outcomes within a two-year follow up period: new charge, new conviction, return to prison, and revocation. The two-year follow up period follows the NDDOCR's definition of recidivism. New charge and new conviction were selected as outcome measures as they are more stringent than new arrests which may overestimate actual offending (Myer, 2019). Return to prison was a measure provided in data provided by the NDDOCR and is included as a recidivism outcome because the NDDOCR defines recidivism as an offender returning to prison. Revocation occurs when an offender has their parole or probation revoked due to the individual violating the terms of their parole or probation conditions through the commission of a new offense or by absconding from supervision (Steen, Opsal, Lovegrove, and Mckinzey, 2013). Traffic events and certain B misdemeanors, listed below, were not recorded as recidivism events because they do not capture the nature of reoffending seen in previous LSI-R validation studies (Schlager, 2005; Folsom and Atkinson, 2007) or were not considered recidivism events by the NDDOCR. B misdemeanors excluded from the present study include misdemeanors classified as "no level;" Driving Under a Suspended/Revoked License; Failure to transfer title-Owner or transferor – traffic; Failure to transfer title – traffic; Licenses To Hunt, Trap, Or Fish

Required Of Nonresidents; School Attendance; Unlawful use of license plate/tabs – traffic; Drove With Cancel/Revoked Altered Plate/Card – traffic; Licenses To Hunt, Trap, Or Fish Required Of Residents – traffic; Game & Fish-Licenses to hunt, trap, or fish required by residents – Fishing; Duty Upon Striking Fixtures; Unlawful Registration; Operate A Motor Boat W/O Adequate Life Preservers.

Research design

The design utilized in this study is a retrospective correlational or "known-groups" design, which is frequently used in validation studies (Folsom and Atkinson, 2007; Fass et al., 2008; Chenane et al., 2015; Ellison, 2016). This design is frequently utilized in validation studies because evidence for the validity of an assessment is determined through the tool's ability to discriminate between the groups that are expected to demonstrate differences in the measure (Fass et al., 2008).

Analytic strategy

Data analyses included univariate tests for demographic purposes and also included bivariate correlations and Receiver Operator Characteristic (ROC)/Area Under the Curve (AUC) analysis. Correlation coefficients are useful for establishing associations between predictors and outcomes (Andrews and Bonta, 2003). However, for measuring predictive accuracy, more meaningful statistical measures have to be considered in addition to basic regression models. As explained by Smith and Smith (1998), selection ratios (the proportion of people in a given group predicted to participate in a future event) base rates (the number of a group who actually participated) greatly influence the rates of false positives/negatives and true positives/negatives derived from a risk assessment, as well as the magnitude of the association between a risk predictor and a criminal outcome. In the context of risk management, false negatives or false positives can yield major issues. If there are too many false negatives for example, an agency may reallocate resources away from high-risk interventions toward low-risk offenders due to the increase of offenders classified as low-risk. Similarly, if there were to be an inflation of false positives, more resources than needed would be allocated to treat and service offenders deemed high-risk but are not actually at such a risk-level. Either an inflation of false negatives or positives result in a system which is underperforming, with some individuals receiving unneeded intervention and others missing much-needed interventions (Schlager, 2005). Utilizing a statistical measure like ROC which is minimally affected by base rate and selection ratio issues, provides the ability to plot for each LSI-R score the proportion of hits (true positives) and false positives (Andrews and Bonta, 2003). A curve illustrating these proportions is then drawn, and an additional calculation, the AUC is also made. Based on the ROC curve, the LSI-R can be evaluated as better than chance. The AUC value, which ranges from 0-1, reports the probability that a randomly selected recidivist would have a higher LSI-R score than a randomly selected nonrecidivist. The closer to 1, the greater predictive accuracy the tool has. In evaluating the strength of AUC values, there are three thresholds indicating small, moderate and large effect sizes: 0.56, 0.64, and 0.71, respectively (Rice and Harris, 2005).

CHAPTER 4: RESULTS

Tables 1 through 4 present descriptive statistics for the full sample (n =4,165), as well as the community (n =2,700) and institutional cohorts (n =1,465). In Table 1, demographic characteristics of the samples are presented. Males comprised 75.8 percent of the full sample, 72.9 percent and 81.1 percent of the community and institutional cohorts, respectively. Whites made up 69.6 percent of the full sample, Native Americans comprising the second largest majority at 19.2 percent. In the full sample, most of the sample fell into the 43+ age bracket (25.7%) followed by the 28-32 age bracket (23.9%).

Table 1

Demographic Information on Sample

	Community (n = 2,700)		Institutional (n = 1,465)		Combined (n = 4,165)	
	n	%	n	%	n	%
Gender						
Male	1,969	72.9	1,188	81.1	3,157	75.8
Female	731	27.1	277	18.9	1,008	24.2
Race						
White	1,927	71.4	973	66.4	2,900	69.6
Native American	473	17.5	325	22.2	798	19.2
Black	208	7.7	91	6.2	299	7.2
Hispanic	86	3.2	71	4.8	157	3.8
Asian	6	0.2	2	0.1	8	0.2
Multiracial	0	0.0	3	0.2	3	0.1
Age						
18-22	25	0.9	1	0.1	26	0.6
23-27	474	17.6	164	11.2	638	15.3
28-32	639	23.7	356	24.3	995	23.9
33-37	507	18.8	318	21.7	825	19.8
38-42	382	14.1	230	15.7	612	14.7
43+	673	24.9	396	27.0	1,069	25.7

Table 2 provides information on the raw LSI-R scores for the sample. LSI-R raw scores range from 2-50. Table 3 provides further LSI-R information on the samples, broken into the tool's five risk levels. For the full sample, 13.4 percent of individuals in this study were classified as low risk, 9.5 percent as low-moderate risk, 20.3 percent as moderate risk, 38.9 percent as moderate-high risk, and 17.8 percent as high risk. In both the community and institutional cohorts, the largest majority of individuals fall under the moderate-high risk category (33.5% and 48.8%, respectively). A Chi-square test run on the community and institutional cohorts revealed a significant chi-square value of 353.812 with four degrees of freedom and a p-value of .000, indicating a significant difference between the community and institutional cohort, with the institutional sample presenting as significantly higher risk.

Table 2

LSI-R Raw Score Information on Sample

	Community (n = 2,700)		Institutional (n = 1,465)		Combined (n = 4,165)	
	n	%	n	%	n	%
Raw Score						
2	3	0.1	0	0.0	3	0.1
3	4	0.1	0	0.0	4	0.1
4	7	0.3	0	0.0	7	0.2
5	4	0.1	0	0.0	4	0.1
6	4	0.1	0	0.0	4	0.1
7	8	0.3	0	0.0	8	0.2
8	20	0.7	0	0.0	20	0.5
9	23	0.9	0	0.0	9	0.6
10	17	0.6	1	0.1	10	0.4
11	28	1.0	3	0.2	31	0.7
12	29	1.1	3	0.2	32	0.8
13	48	1.8	3	0.2	51	1.2
14	36	1.3	3	0.2	39	0.9
15	52	1.9	2	0.1	54	1.3
16	43	1.6	13	0.9	56	1.3
17	68	2.5	9	0.6	77	1.8
18	55	2.0	10	0.7	65	1.6
19	55	2.0	9	0.6	64	1.5
20	75	2.8	19	1.3	94	2.3
21	87	3.2	18	1.2	105	2.5
22	85	3.1	16	1.1	101	2.4

Table 2

LSI-R Raw Score Information on Sample (Continued)

	Community (n = 2,700)		Institutional (n = 1,465)		Combined (n = 4,165)	
	n	%	n	%	n	%
Raw Score						
23	78	2.9	18	1.2	96	2.3
24	89	3.3	23	1.6	112	2.7
25	99	3.7	22	1.5	121	2.9
26	91	3.4	45	3.1	136	3.3
27	96	3.6	44	3.0	140	3.4
28	100	3.7	53	3.6	153	3.7
29	124	4.6	61	4.2	185	4.4
30	110	4.1	53	3.6	163	3.9
31	101	3.7	68	4.6	169	4.1
32	107	4.0	75	5.1	182	4.4
33	117	4.3	74	5.1	191	4.6
34	105	3.9	79	5.4	184	4.4
35	94	3.6	85	5.8	179	4.3
36	98	3.6	98	6.7	196	4.7
37	88	3.3	92	6.3	180	4.3
38	84	3.1	91	6.2	175	4.2
39	80	3.0	64	4.4	144	3.5
40	67	2.5	76	5.2	143	3.4
41	51	1.9	56	3.8	107	2.6
42	44	1.6	47	3.2	91	2.2
43	42	1.6	49	3.3	91	2.2
44	34	1.3	35	2.4	69	1.7
45	14	0.5	23	1.6	37	0.9
46	16	0.6	7	0.5	23	0.6
47	9	0.3	11	0.8	20	0.5
48	8	0.3	4	0.3	12	0.3
49	3	0.1	0	0.0	3	0.1
50	0	0.0	3	0.2	3	0.1

Table 3

LSI-R Risk Level Information on Sample

	Community (n = 2,700)		Institutional (n = 1,465)		Combined (n = 4,165)	
	n	%	n	%	n	%
Risk Level						
Low risk (0-19)	504	18.7	56	3.8	560	13.4
Low-moderate risk (20-23)	325	12.0	71	4.8	396	9.5
Moderate risk (24-29)	599	22.2	248	16.9	847	20.3
Moderate-high risk (30-38)	904	33.5	715	48.8	1619	38.9
High risk (39-54)	358	13.6	375	25.6	743	17.8

A Chi-square test was run on the Community and Institutional samples: $\chi 2$=353.812; df = 4; p=.000

Recidivism information is presented for the sample in Table 4. In the full sample, 52.4 percent of individuals had a new charge in the two year follow up period and 47.1 percent had a new conviction. However, only 19.2 percent returned to prison during the follow up period. The revocation outcome was relatively constant across the full sample and two cohorts, with a 41.4 percent revocation rate in the institutional cohort and 41.3 percent for the community cohort.

Table 4

Recidivism Information on Sample

	Community (n = 2,700)		Institutional (n = 1,465)		Combined (n = 4,165)	
	n	%	n	%	n	%
Two-year recidivism						
New Charge	1366	50.6	816	55.7	2182	52.4
New Conviction	1223	45.3	740	50.5	1963	47.1
Return to Prison	507	18.8	292	19.9	799	19.2
Revocation	1116	41.3	607	41.4	1723	41.4

Table 5 presents the results from the AUC analysis. The findings indicate that overall, the LSI-R had moderate success in discriminating between recidivists and non-recidivists. For the full sample, the AUC shows there is a 62.2 percent chance that a randomly selected recidivist will have higher LSI-R score than a non-recidivist for the new charge and new conviction outcomes. For the institutional cohort, the LSI-R had a smaller effect size suggesting lower discriminant ability, with a 59.5 percent chance at distinguishing between recidivists and non-recidivists for the new charge outcome, 59.3 percent chance for the new conviction outcome, 50 percent chance for the return to prison outcome, and 56.8 percent for the revocation outcome. In contrast, the LSI-R fared better in discriminating between recidivists and non-recidivists for the community cohort, with a 73.2 percent chance at distinguishing for the return to prison outcome and 74.9 percent chance for the revocation outcome.

Table 5

Area Under Curve Information on Sample

	Community (n = 2,700)			Institutional (n = 1,465)			Combined (n = 4,165)		
		95% CI			95% CI			95% CI	
	AUC	LL	UL	AUC	LL	UL	AUC	LL	UL
Two-year recidivism									
New Charge	0.665	0.645	0.686	0.595	0.565	0.624	0.622	0.605	0.639
New Conviction	0.665	0.645	0.686	0.593	0.564	0.622	0.622	0.605	0.639
Return to Prison	0.732	0.710	0.755	0.536	0.500	0.573	0.626	0.607	0.645
Revocation	0.749	0.731	0.767	0.597	0.568	0.626	0.641	0.625	0.658

Table 6 provides a comparison of the present study's AUC and bivariate correlation analyses against findings from previous LSI-R validation studies. In terms of the AUC results, the present study fared similarly to past studies, with the 62.2 percent likelihood of distinguishing between a recidivist and non-recidivist comparable to the 60 percent chance in Fass et al. (2008), the 68.9 percent chance in Flores et al. (2006), 67 percent chance in Folsom and Atkinson (2007), and 61.78 percent chance in Ostermann and Salerno (2016). In the present study, the *r*-values for the bivariate correlations are statistically significant at 0.01 for all four recidivism outcomes, with an *r*-value of .253 for the new charge and new conviction outcomes. These results are similar to the findings presented in Flores et al. (2006), Folsom and Atkinson (2007), and Ostermann and

Salerno (2016), with *r*-values of .283, .300, and .2034 (significant at the $p \leq .001$ level), respectively.

Table 6

Comparison of Current Results for LSI-R and Recidivism Outcomes Against Prior Research

			95% CI		
	n	AUC	LL	UL	r
LSI-R study sample					
Current sample					
Two-year new charge	4,165	0.622	0.605	0.639	.253**
Two-year new conviction	4,165	0.622	0.605	0.639	.253**
Two-year return to prison	4,165	0.626	0.607	0.645	.230**
Two-year revocation	4,165	0.641	0.625	0.658	.333**
Fass et al. (2008)					
One-year new arrest	696	0.600	--	--	--
Flores et al. (2006)					
Reincarceration	2,107	0.689	--	--	.283
Folsom and Atkinson (2007)					
Seven-year new conviction	100	0.670	0.550	0.790	.300
Seven-year new nonviolent conviction	100	0.620	0.500	0.740	--
Seven-year new violent conviction	100	0.670	0.490	0.850	--
Ostermann and Salerno (2016)					
One-year general recidivism	9,454	0.6178	0.6062	0.6294	.2034***

Fass et al. (2008) and Flores et al. (2006) did not present confidence intervals when presenting ROC/AUC results in their respective studies. Flores et al. (2006) did not indicate a follow-up time period for recidivism events. Ostermann and Salerno (2016) constructed their general recidivism outcome from two measures: arrest for a new crime or a return to prison due to a parole revocation.

**. Correlation is significant at the 0.01 level (2-tailed).

Note. AUC = area under the curve; CI = confidence interval.

Figures 1 through 4 present findings for the bivariate analyses on the full sample. Figure 1 and 2 show that there was a significant correlation ($r = .253$; $p < .000$) between the LSI-R risk level and new charge/new conviction outcomes at the two year follow up. Additionally, the return to prison ($r = .230$; $p < .000$) and revocation ($r = .333$; $p < .000$) outcomes were also significant at the 0.01 level. Across all recidivism outcomes, those in the lower risk categories were less likely to recidivate than those individuals in the higher risk categories, as indicated by the stair-step pattern seen in the figures.

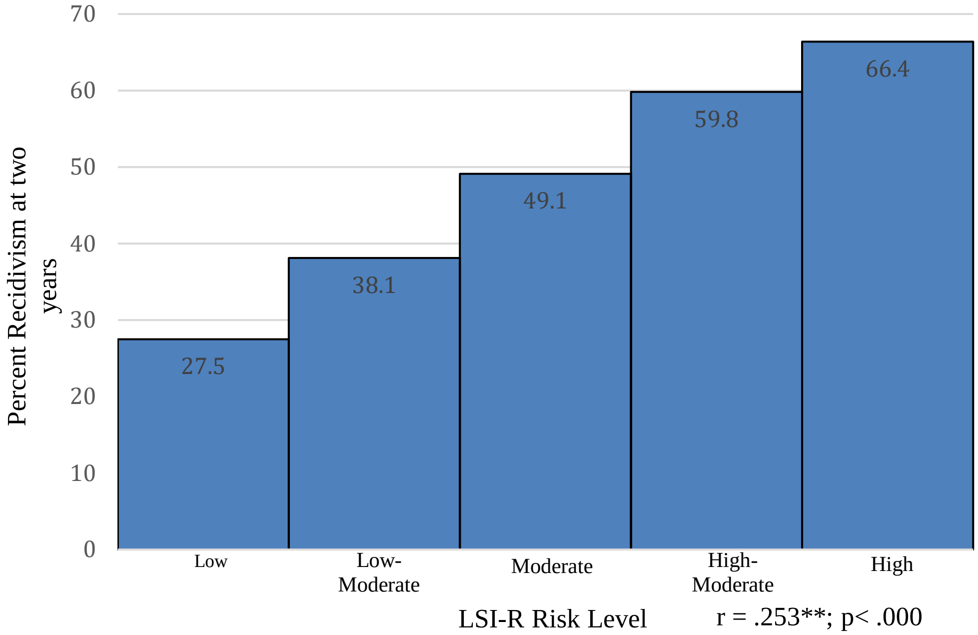

Figure 1. New Charge Rate in Combined Sample

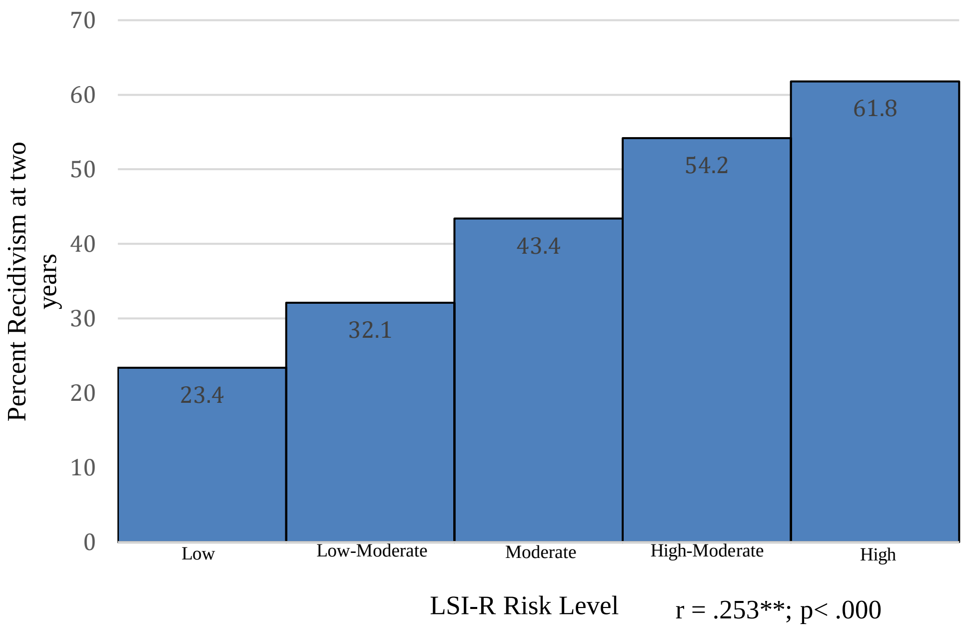

Figure 2. New Conviction Rate in Combined Sample

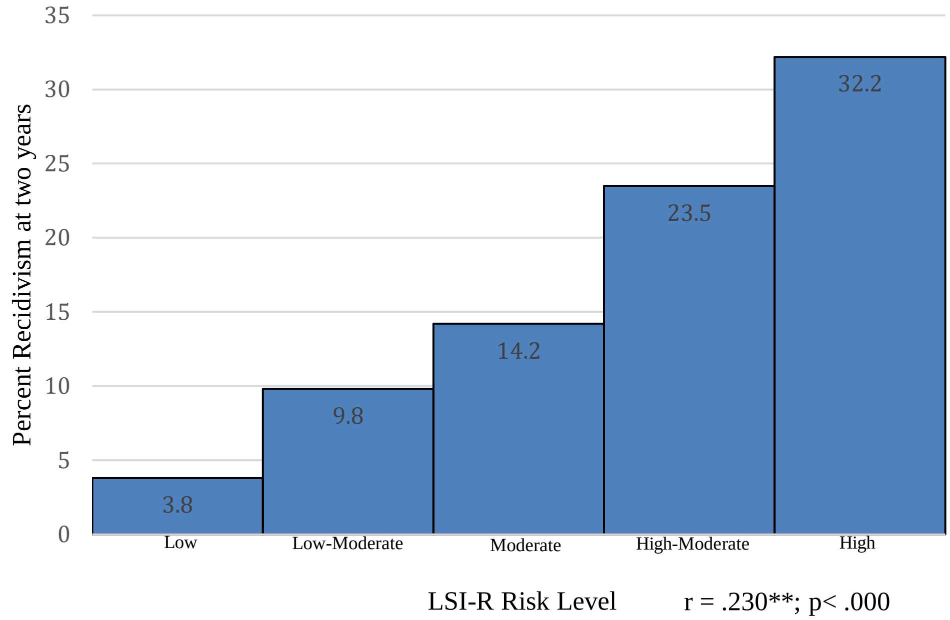

Figure 3. Return to Prison Rate in Combined Sample

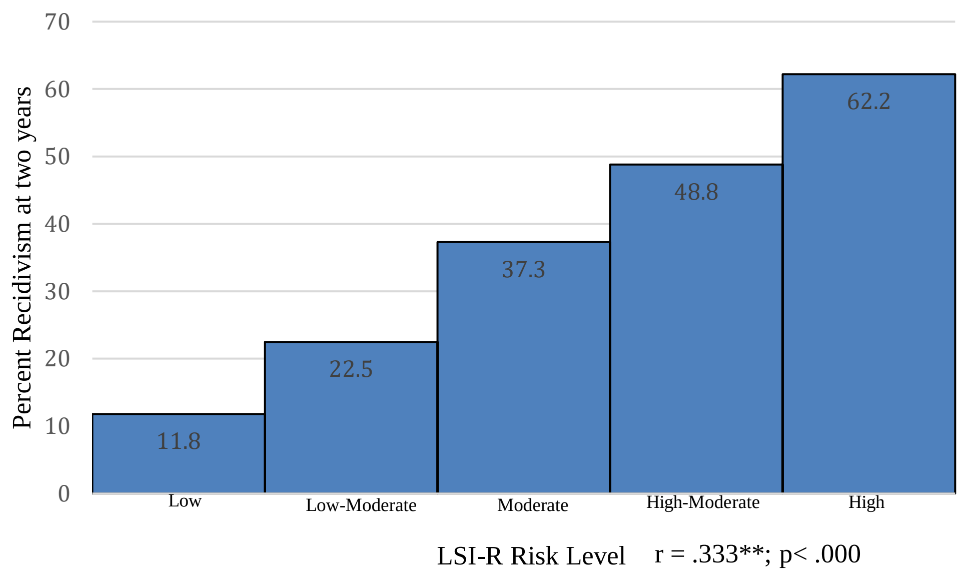

Figure 4. Revocation Rate in Combined Sample

Figures 5 through 8 illustrate results on the bivariate analyses for the community cohort. The correlation between LSI-R risk level and recidivism was significant for each of the four outcomes. New charge in the community cohort had an *r* value of .280 ($p < .000$), while new conviction had an *r* value of .283 ($p < .000$), return to prison had an *r* value of .315 ($p < .000$), and revocation had an *r* value of .424 ($p < .000$); all were significant at the 0.01 level. Similarly to those shown in 1 through 4, figures 5 through 8 also follow a stair-step pattern with respect to the recidivism outcomes for each risk category. However, for the return to prison outcome, the recidivism rate increase is less consistent across the risk categories compared to the other recidivism outcomes.

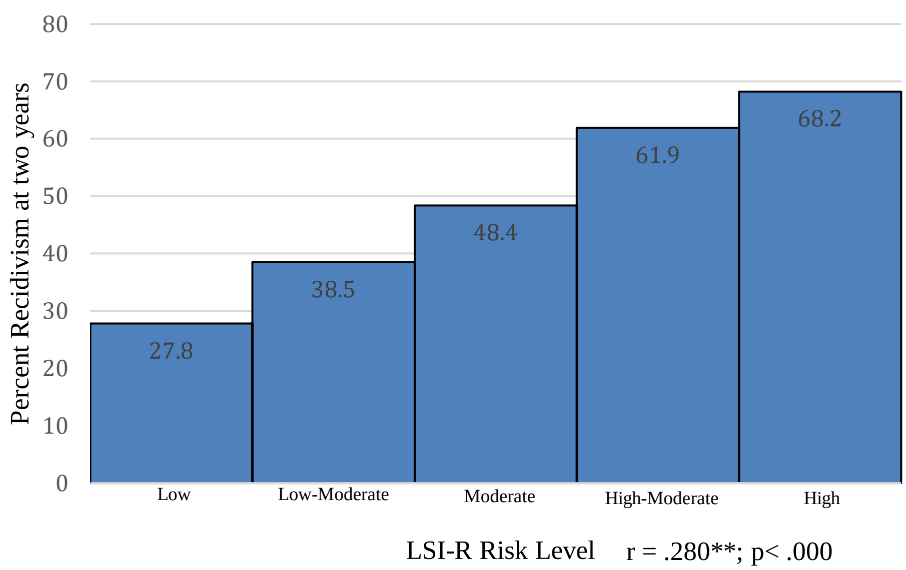

Figure 5. New Charge Rate in Community Sample

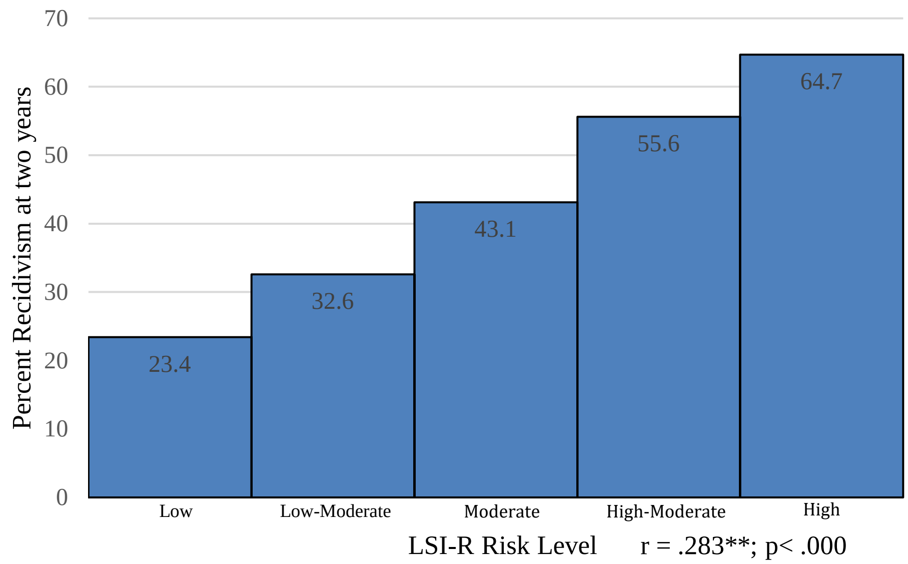

Figure 6. New Conviction Rate in Community Sample

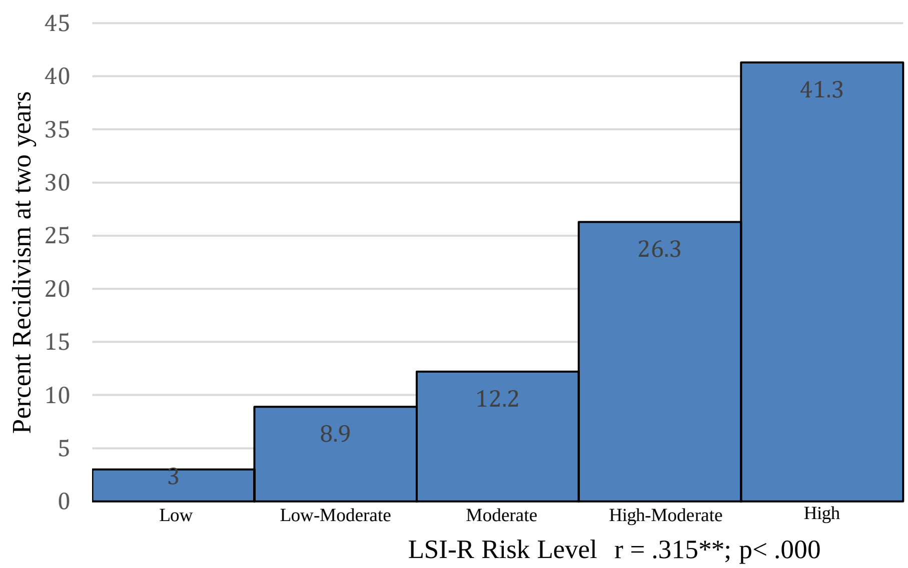

Figure 7. Return to Prison Rate in Community Sample

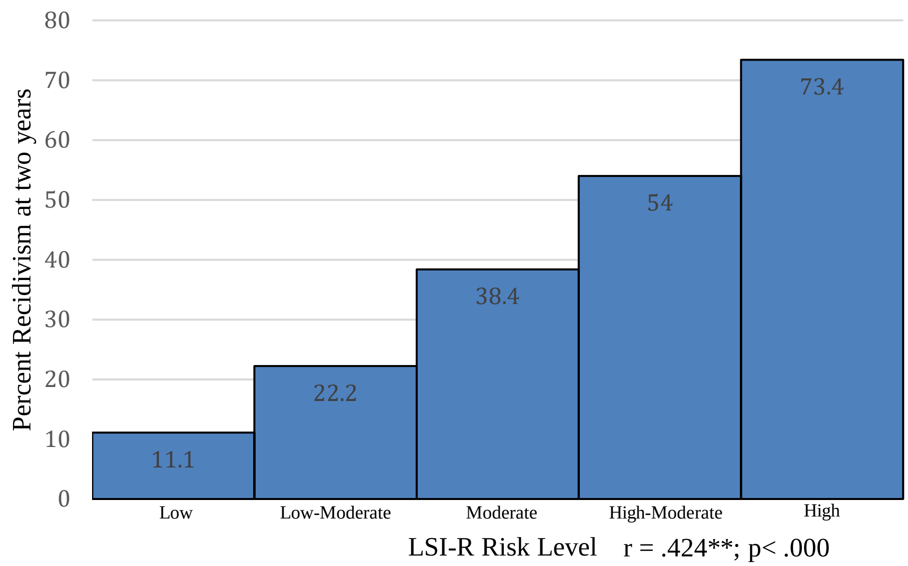

Figure 8. Revocation Rate in Community Sample

Figures 9 through 12 show the findings for the bivariate analyses of the institutional cohort. New charge (*r* = .181; *p*< .000), new conviction (*r* = .178; *p*< .000), and revocation (*r* = .175; *p*< .000), were all significant at the 0.01 level. The return to prison outcome (*r* = .059; *p*= .024) was comparatively weaker, and significant only at the 0.05 level. This can be seen in stair-step pattern of figure 11, where the gradation between risk categories is more modest compared to those seen in figures for the other recidivism outcomes.

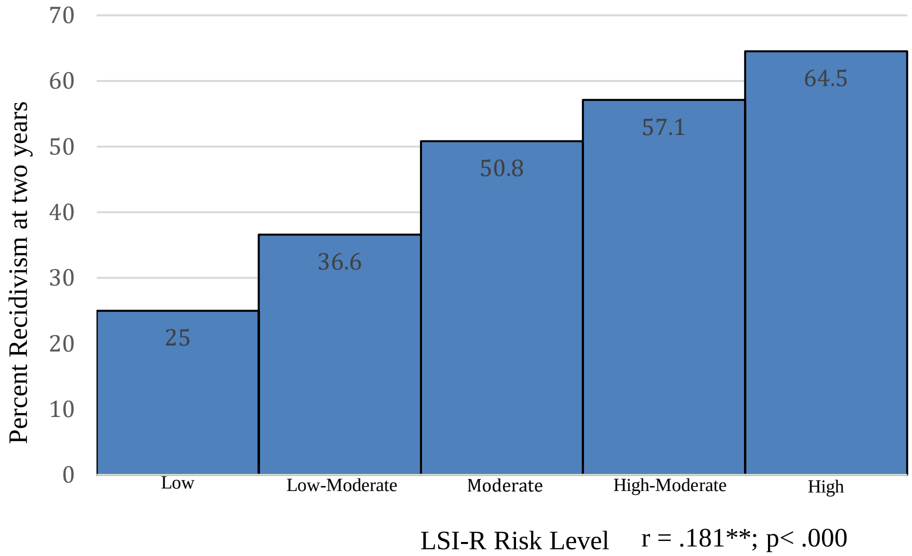

Figure 9. New Charge Rate in Institutional Sample

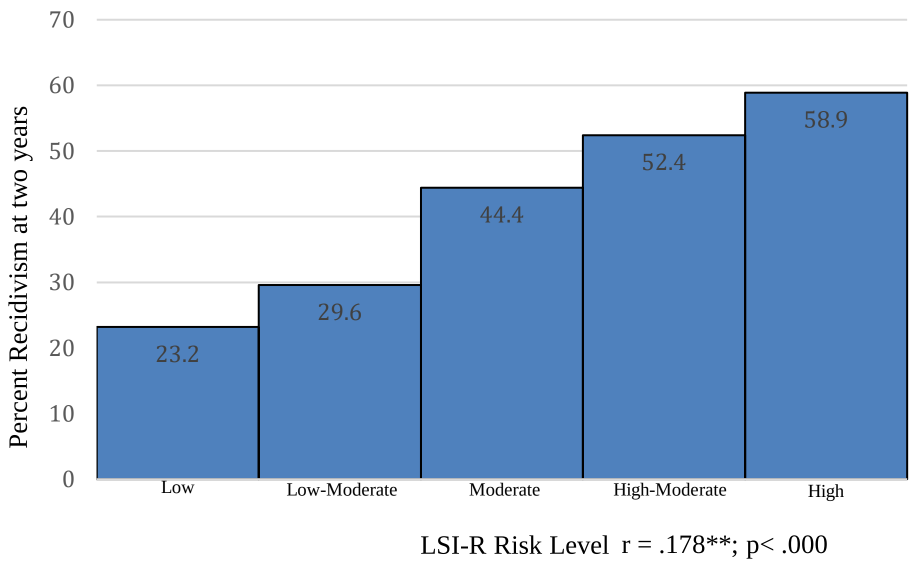

Figure 10. New Conviction Rate in Institutional Sample

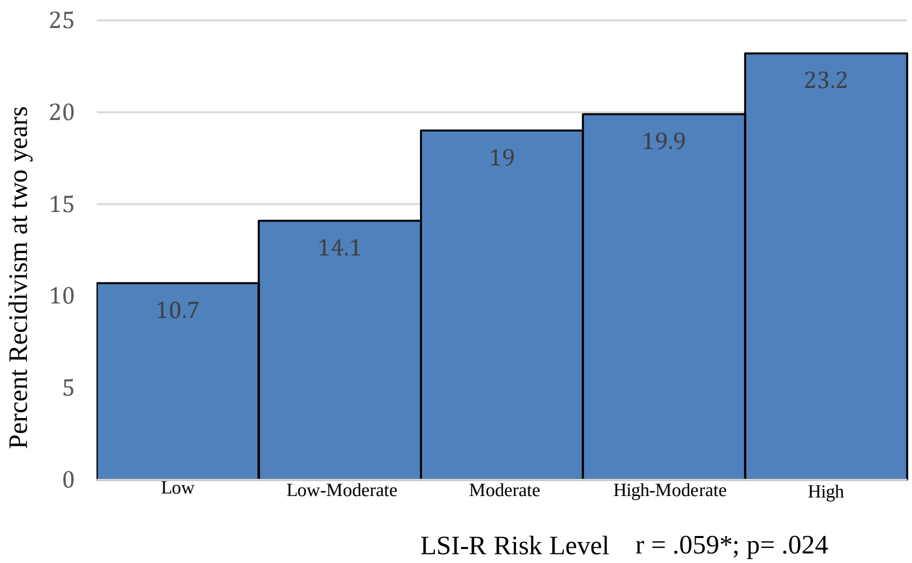

Figure 11. Return to Prison Rate in Institutional Sample

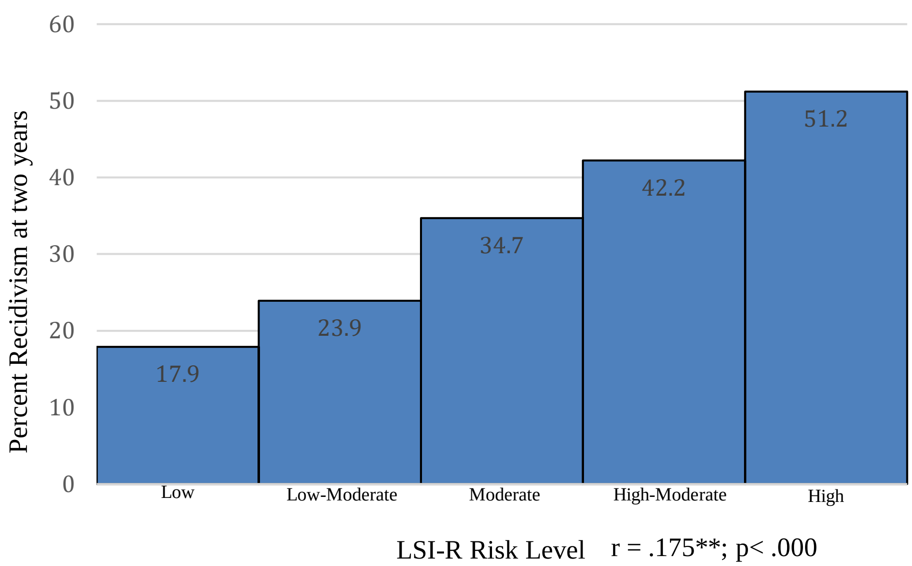

Figure 12. Revocation Rate in Institutional Sample

Supplemental analyses on race

Additional analyses were conducted on the sample by race, with the sample separated into three cohorts for analyses: White (n= 2,900), Native American (n= 798), and Other (n =467). These cohorts were created by recoding the existing race variable into three levels. Consolidating Black, Asian, Hispanic, and Multiracial individuals into the "Other" category allowed for easier comparison of the two racial demographics of primary interest to the present study, White and Native American. Table 7 presents information on the raw LSI-R scores for the three race cohorts. LSI-R raw scores range from 2-50. Table 8 provides LSI-R information on the cohorts by risk levels. For all three cohorts, the largest majority of individuals fell under the moderate-high risk level, with 38.6 percent for the White cohorts, 41.2 percent for the Native American cohort, and 36.4 percent for the Other cohort. 14.9 percent of White cohort and 18.4 percent those in the Other cohort were classified as low risk, compared to just 5.1 percent of the Native American cohort. Conversely, 31.3 percent of the Native American cohort were classified at high risk compared to 15 percent and 12.2 percent for the White and Other cohorts, respectively. A Chi-square test run on White and Native American cohorts revealed a significant difference. Specifically, the Native American cohort was shown to be significantly higher risk than the White cohort.

Table 7

LSI-R Raw Score Information by Race

	White (n = 2,900)		Native American (n = 798)		Other (n = 467)	
	n	%	n	%	n	%
Raw Score						
2	2	0.1	0	0.0	1	0.2
3	3	0.1	0	0.0	1	0.2
4	5	0.2	0	0.0	2	0.4
5	4	0.1	0	0.0	0	0.0
6	2	0.1	1	0.1	1	0.2
7	6	0.2	0	0.0	2	0.4
8	16	0.6	1	0.1	3	0.6
9	18	0.6	1	0.1	4	0.9
10	13	0.4	2	0.3	2	0.6
11	23	0.8	3	0.4	5	1.1
12	25	0.9	1	0.1	6	1.3
13	41	1.4	5	0.6	5	1.1
14	27	0.9	1	0.1	11	2.4
15	41	1.0	7	0.9	6	1.3
16	43	0.4	4	0.5	9	1.9
17	61	2.1	6	0.8	10	2.1
18	55	1.9	5	0.6	5	1.1
19	48	1.7	4	0.5	12	2.6
20	67	2.3	8	1.0	19	4.1
21	85	2.9	6	0.8	14	3.0
22	73	2.5	14	1.8	14	3.0
23	77	2.7	10	1.3	9	1.9
24	101	3.5	4	0.5	7	1.5
25	81	2.8	23	2.9	17	3.6
26	89	3.1	29	3.6	18	3.9
27	109	3.8	17	2.1	14	3.0
28	100	3.4	34	4.3	19	4.1
29	129	4.4	33	4.1	23	4.9
30	124	4.3	20	2.5	19	4.1
31	123	4.2	29	3.6	17	3.6
32	135	4.7	34	4.3	13	2.8
33	128	4.4	39	4.9	24	5.1
34	121	4.2	39	4.9	24	5.1
35	134	4.6	34	4.3	11	2.4
36	133	4.6	48	6.0	15	3.2
37	107	3.7	46	5.8	27	5.8
38	115	4.0	40	5.0	20	4.3

Table 7

LSI-R Raw Score Information by Race (Continued)

	White (n = 2,900)		Native American (n = 798)		Other (n = 467)	
	n	%	n	%	n	%
Raw Score						
39	102	3.5	31	3.9	11	2.4
40	84	2.9	48	6.0	11	2.4
41	66	2.3	32	4.0	9	1.9
42	50	1.7	37	4.6	4	0.9
43	47	1.6	33	4.1	11	2.4
44	38	1.3	28	3.5	3	0.6
45	19	0.7	14	1.8	4	0.9
46	13	0.4	9	1.1	1	0.2
47	11	0.4	6	0.8	3	0.6
48	4	0.1	8	1.0	0	0.0
49	1	0.0	2	0.3	0	0.0
50	1	0.0	2	0.3	0	0.0

Table 8

LSI-R Risk Level Information by Race

	White (n = 2,900)		Native American (n = 798)		Other (n = 467)	
	n	%	n	%	n	%
Risk Level						
Low risk (0-19)	433	14.9	41	5.1	86	18.4
Low-moderate risk (20-23)	302	10.4	38	4.8	56	12.0
Moderate risk (24-29)	609	21.0	140	17.5	98	21.0
Moderate-high risk (30-38)	1120	38.6	329	41.2	170	36.4
High risk (39-54)	436	15.0	250	31.3	57	12.2

A Chi-square test was run on the White and Native American samples: $\chi 2=162.905$; df = 4; p=.000

In Table 9, recidivism results are presented for the three race cohorts, along with the results of chi-square tests on each of the four recidivism outcomes. In the White cohort, the new charge rate was 50.9 percent, compared to 57.3 percent for the Native American cohort and 53.5 percent for the Other cohort. The new conviction rate was 45.8 percent for the White cohort, 51.5 percent for the Native American cohort, and 47.8 percent for the Other cohort. However, only 19.2 percent returned to prison during the follow up period. The revocation outcome differed between the cohorts, with a rate of 38.1 percent for the White cohort and 37.9 percent for the Other cohort, compared to a rate of 55.4 percent for the Native American cohort. The chi-square tests for the White and Native American cohorts revealed that Native Americans were significantly more likely to have recidivated across each of the recidivism outcomes when compared to White individuals.

Table 9

Recidivism Information by Race

	White (n = 2,900)		Native American (n = 798)		Other (n = 467)		Chi-Square (White and Native American)		
	n	%	n	%	n	%	$\chi 2$	df	p
Two-year recidivism									
New Charge	1,475	50.9	457	57.3	250	53.5	10.293	1	.001
New Conviction	1,329	45.8	411	51.5	223	47.8	8.093	1	.004
Return to Prison	521	18.0	196	24.6	82	17.6	17.419	1	.000
Revocation	1,104	38.1	442	55.4	177	37.9	77.159	1	.000

Table 10 presents the findings from the AUC analysis for the race cohorts. The results suggest that LSI-R had small to moderate discriminant validity across racial demographics. For the new charge outcome, the chance that a randomly selected recidivist had a higher LSI-R score than non-recidivist was 62.7 percent for the White cohort, 61.6 percent for the Native American cohort, and 58 percent for the Other cohort. Further comparison of the White and Native American cohort shows that the tool does not predict similarly for Whites and Native Americans, the return to prison and revocation AUC scores are significantly different (64.2% vs. 52.6% and 64.5% vs. 56.9%). The significant was indicated by the non-overlapping confidence intervals between Whites (return to prison, CI = [.619, .665], revocation, CI = [.625, .665]) and Native Americans (return to prison, CI = [.484, .569], revocation, CI = [.528, .610].

Table 10

Area Under the Curve Information by Race

	White (n = 2,900)			Native American (n = 798)			Other (n = 467)		
		95% CI			95% CI			95% CI	
	AUC	LL	UL	AUC	LL	UL	AUC	LL	UL
Two-year recidivism									
New Charge	0.627	0.607	0.647	0.616	0.577	0.656	0.580	0.528	0.632
New Conviction	0.629	0.609	0.649	0.619	0.580	0.658	0.569	0.517	0.620
Return to Prison	0.642	0.619	0.665	0.526	0.484	0.569	0.658	0.606	0.711
Revocation	0.645	0.625	0.665	0.569	0.528	0.610	0.643	0.594	0.692

Table 11 presents a comparison of the present study's AUC and bivariate correlation analyses for its Native American cohort against results from previous LSI-R validation studies that used Native American and Indigenous samples. For the AUC analyses, the present study's sample fared similar to worse compared to the sample in Wormith, Hogg, and Guzzo (2014), which reported a 72 percent chance in discriminating between recidivists and non-recidivists for its general recidivism outcome and a 64 percent chance for its violent recidivism outcome, which are moderate to large in effect size compared to the small-approaching-moderate effect sizes seen in the AUC range of 61.6 percent to 64.5 percent reported for the four recidivism outcomes of the present study. For the bivariate analyses, the present study reported *r*-values on its Native

American cohort for new charge (r =.250), new conviction (r =.254), return to prison (r =.169), and revocation (r = .285). In comparison, Holsinger, Lowenkamp, and Latessa (2006) reported an r-value of .18 for its seventeen-month new arrest outcome.

Table 11

Comparison of AUC results for current sample against prior LSI-R studies on Native Americans/Indigenous Samples

			95% CI		
	n	AUC	LL	UL	r
Native American Samples					
Current sample					
Two-year new charge	798	0.616	0.577	0.656	.250**
Two-year new conviction	798	0.619	0.58	0.658	.254**
Two-year return to prison	798	0.642	0.619	0.665	.169**
Two-year revocation	798	0.645	0.625	0.665	.285**
Holsinger, Lowenkamp, & Latessa (2006)					
Seventeen-month new arrest	403	--	--	--	.18**
Indigenous Samples					
Wormith, Hogg, & Guzzo (2015)					
Four-year general recidivism	1,692	0.72	0.69	0.74	--
Four-year violent recidivism	1,692	0.64	0.62	0.68	--

Note. AUC = area under the curve; CI = confidence interval. *$p \leq .05$. **$p = .01$. ***$p \leq .001$.

Figures 13 through 16 present the findings for the bivariate analyses on the White cohort. For each of four outcomes, the correlation between LSI-R risk level and recidivism was significant at the 0.01 level. New charge in the White cohort had an r value of .266 ($p< .000$), while new conviction had an r value of .268 ($p< .000$), return to prison had an r value of .230 ($p< .000$), and revocation had an r value of .311 ($p< .000$).

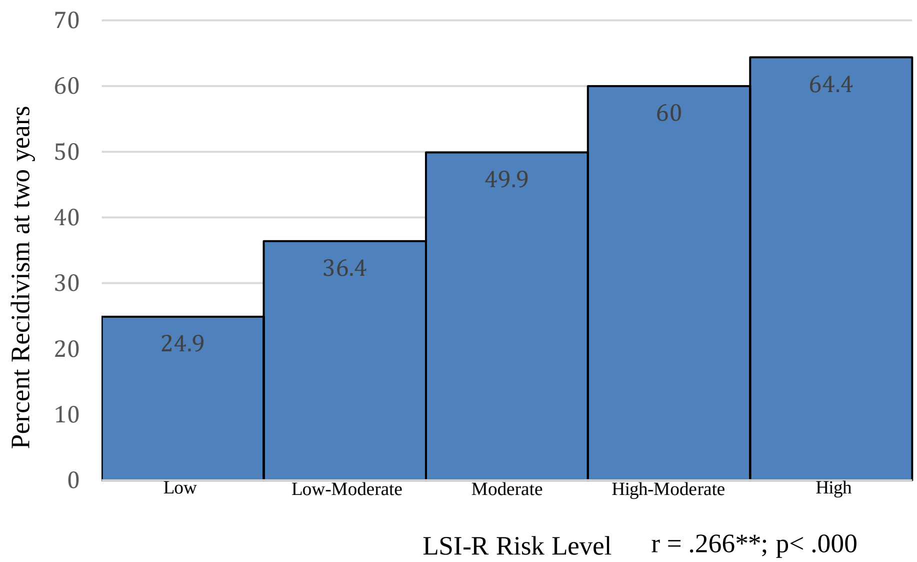

Figure 13. New Charge Rate in White Sample

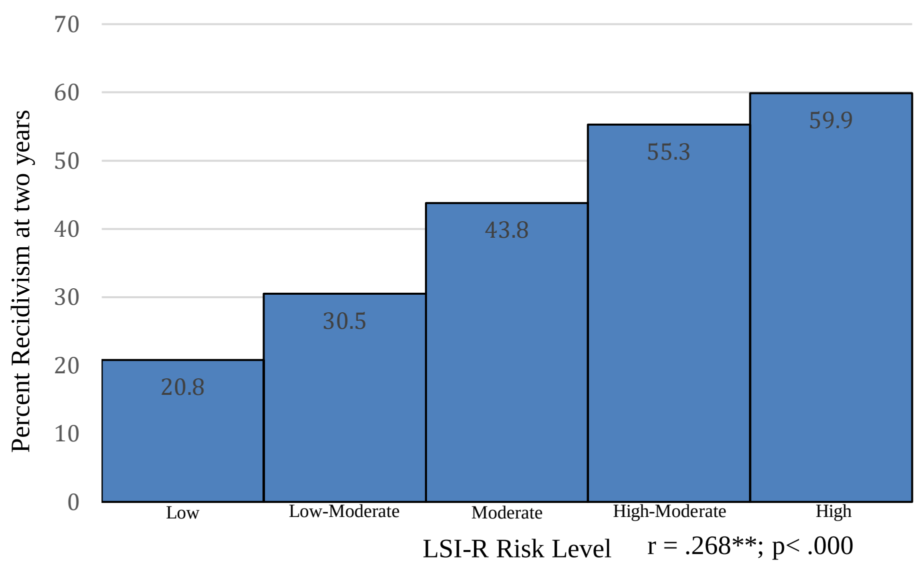

Figure 14. New Conviction Rate in White Sample

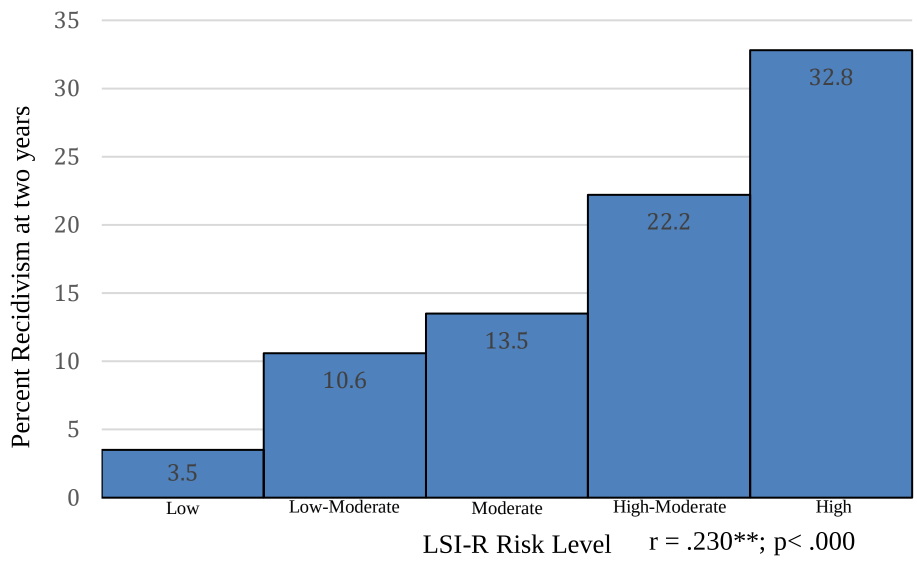

Figure 15. Return to Prison Rate in White Sample

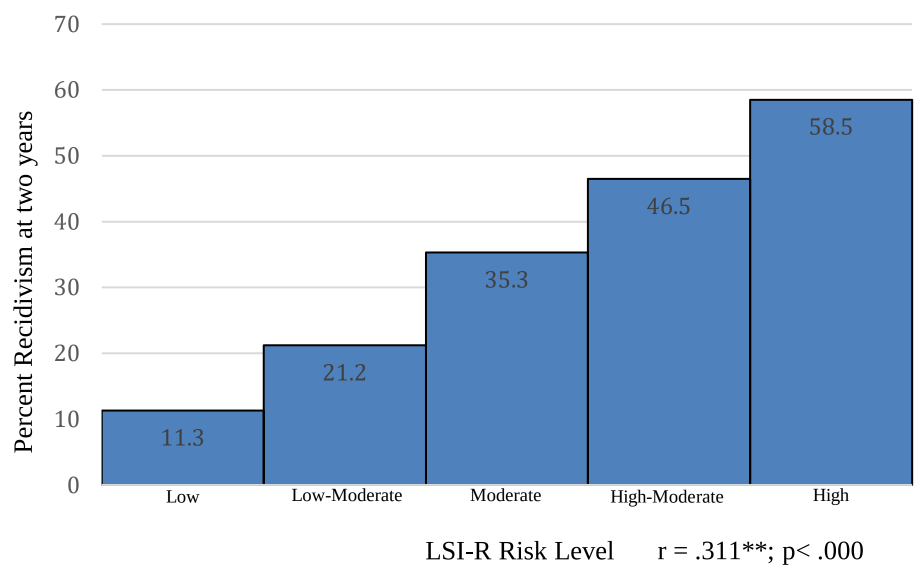

Figure 16. Revocation Rate in White Sample

Figures 17 through 20 show the bivariate analyses results for the Native American cohort. As with the White cohort, all correlations were significant at the 0.01 level. New charge in the Native American cohort had an *r* value of .250 (*p*< .000), while new conviction had an *r* value of

.254 ($p < .000$), return to prison had an r value of .169 ($p < .000$), and revocation had an r value of .285 ($p < .000$).

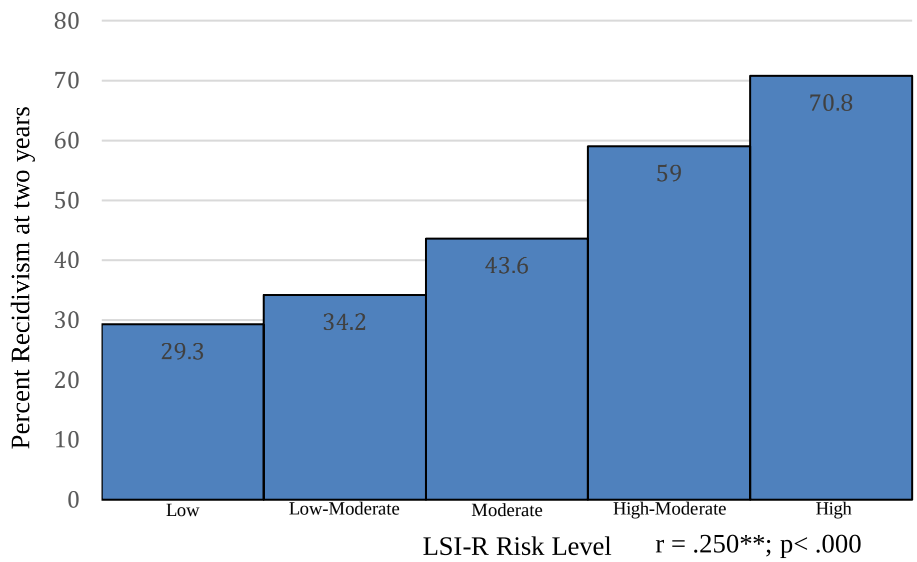

Figure 17. New Charge Rate in Native American Sample

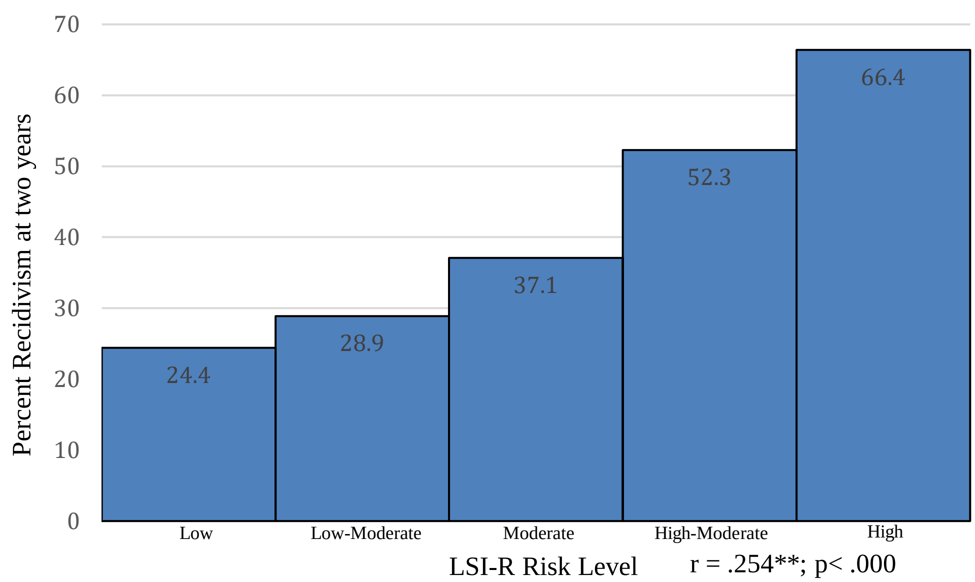

Figure 18. New Conviction Rate in Native American Sample

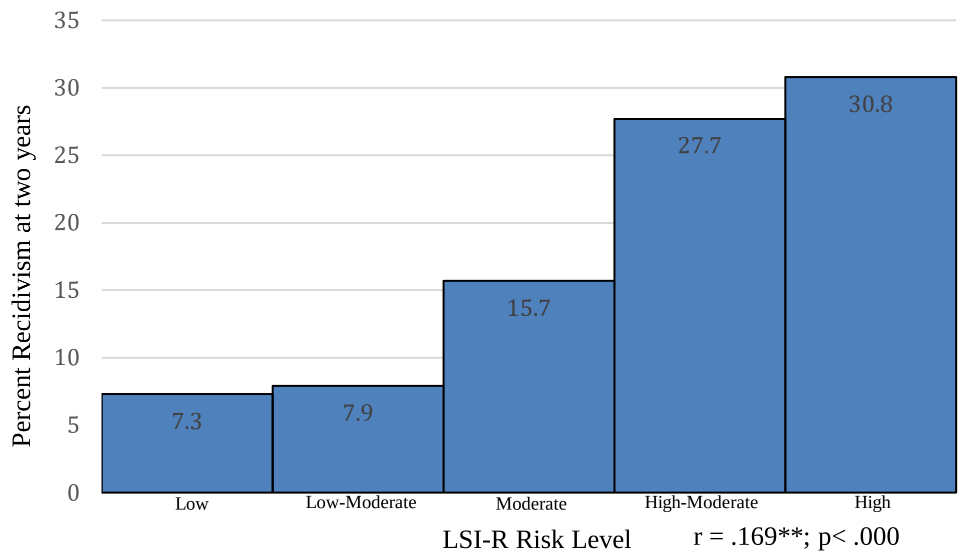

Figure 19. Return to Prison Rate in Native American Sample

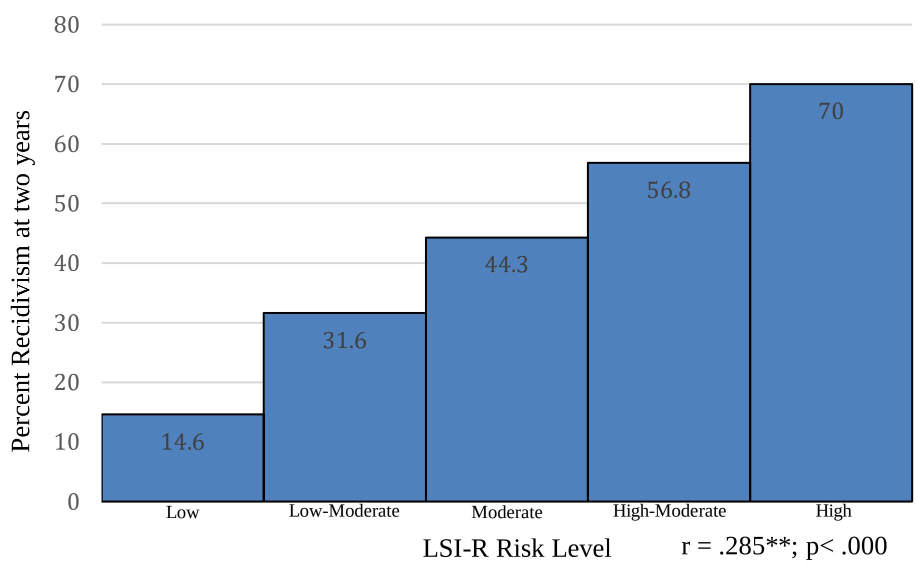

Figure 20. Revocation in Native American Sample

The final set of figures, 21 through 24, present the results for the bivariate analyses conducted on the Other cohort. As with the White and Native American cohorts, all correlations between LSI-R risk level and recidivism outcome were significant at the 0.01 level. New charge in the Other cohort had an *r* value of .152 ($p< .000$), while new conviction had an *r* value of .126 ($p< .000$), return to prison had an *r* value of .251 ($p< .000$), and revocation had an *r* value of .304 ($p< .000$).

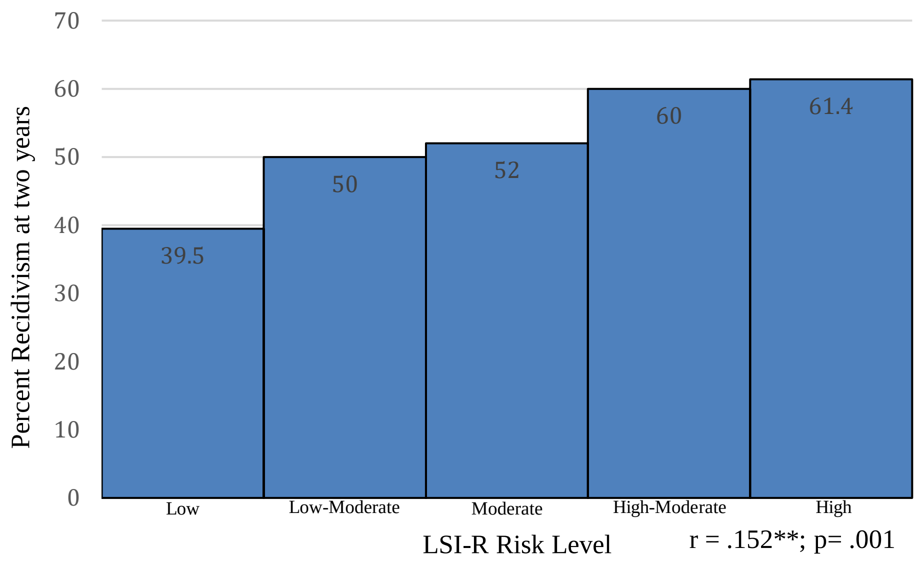

Figure 21. New Charge Rate in Other Sample

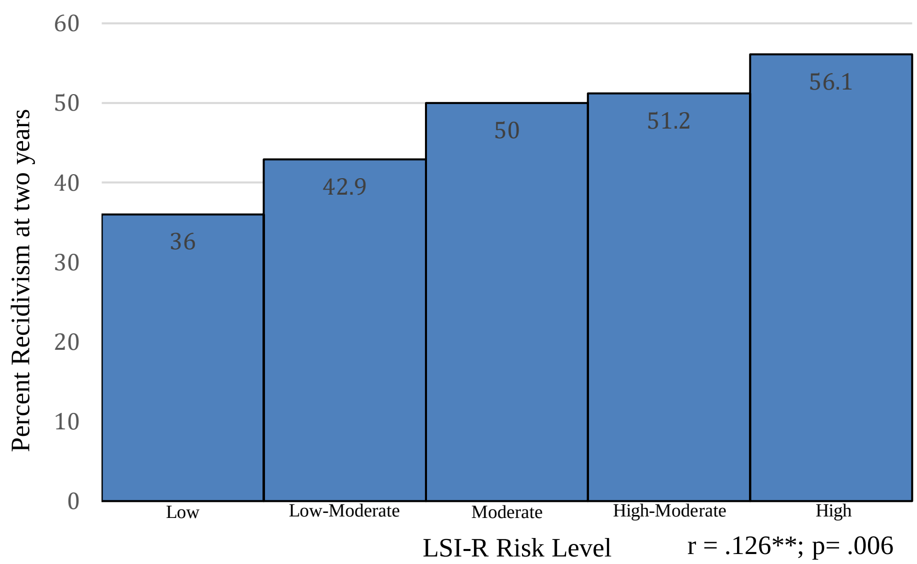

Figure 22. New Conviction Rate in Other Sample

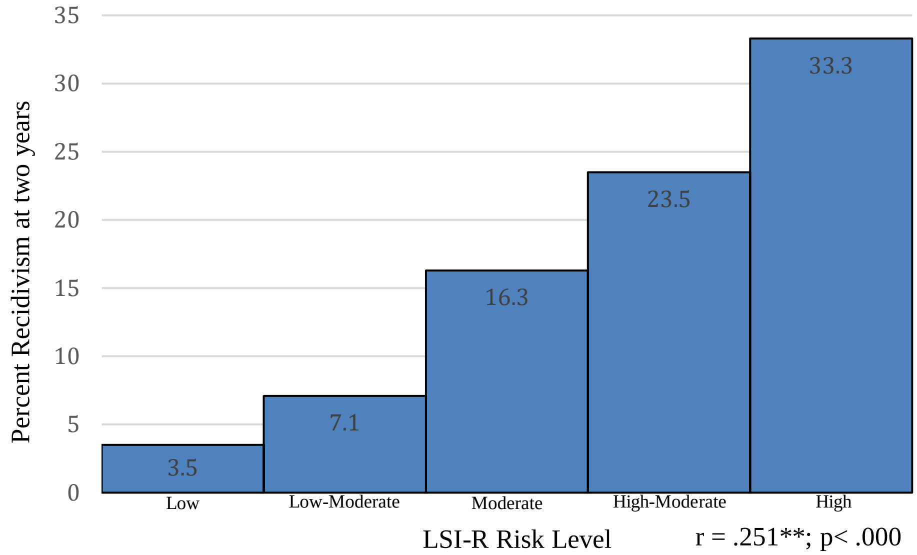

Figure 23. Return to Prison Rate in Other Sample

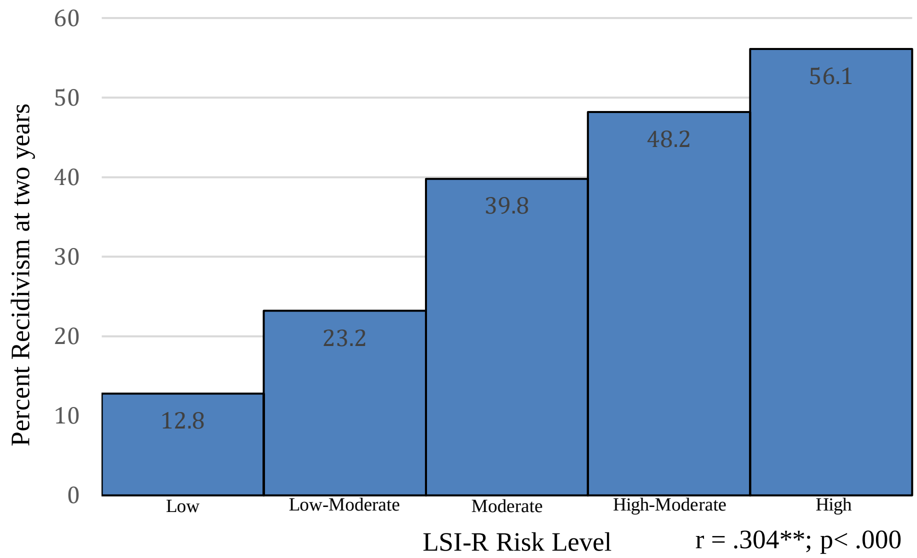

Figure 24. Revocation Rate in Other Sample

CHAPTER 5: DISCUSSION

In the correctional field, risk assessment tools can be invaluable, serving to identify, target, and prevent recidivism in offenders (Latessa and Lovins, 2010). However, risk assessment tools are useful insofar as they are valid predictors of risk. Therefore, it is necessary to conduct validation studies against risk assessment tools, like the LSI-R, so to ensure that resources used in preventing recidivism are correctly allocated to those individuals most at risk for recidivating (Lovins et al., 2018). Additionally, repeated validation of risk assessment tools aids in their refinement, further increasing a tool's predictive ability, especially so when these validations studies are conducted on a wide range of offender populations (Clear, 1991; Lovins et al., 2018). This point becomes particularly salient in the United States, where the general population includes geographically concentrated subpopulations of ethnic minorities, such as Native Americans in Midwestern states (Holsinger, Lowenkamp, and Latessa, 2003).

The LSI-R is one of the most broadly utilized risk assessment tools currently in use (Flores et al., 2006; Folsom and Atkinson, 2007; Schlager and Simourd, 2007; Fass et al., 2008; Austin, 2011; Chenane et al., 2015; Ellison et al., 2016; Olver, Stockdale, and Wormith, 2014; Ostermann and Salerno, 2016; Vose et al., 2009; 2013). Previous efforts to validate the LSI-R have shown support for the LSI-R's general validity, but provided less clarity of the tool's discriminant ability across racial categories (Vose et al., 2009; Chenane et al., 2015; Holsinger, Lowenkamp, and Latessa, 2006; Ostermann and Salerno, 2016). In particular, there is a paucity of research dedicated to validating the LSI-R against Native Americans (Holsinger, Lowenkamp, and Latessa, 2003; Holsinger, Lowenkamp, and Latessa, 2006), with most such studies dealing with indigenous samples from Australia (Hsu, Caputi, and Byrne, 2010; Shepherd, Adams, McEntyre, and Walker, 2014) or Canada (Bonta, LaPairie, and Wallace-Capretta, 1997; Jung and Rawana, 1999; Wilson and Gutierrez, 2014; Wormith, Hogg, and Guzzo, 2015). Thus, the present study augments this sparse area of research by validating the LSI-R on a sample that includes a large cohort of Native American offenders.

There were two research questions in the present study. First, what is the predictive validity of the LSI-R on a Midwest corrections population? Second, what is the predictive validity of the LSI-R on a substantial Native American corrections population? To address these questions, LSI-R data were obtained through the NDDOCR on a sample of 4,165 offenders released to the community in 2016. Recidivism data were gathered for the data through the North Dakota Court Records Inquiry website. There were four recidivism outcomes recorded: new charge, new conviction, return to prison, and revocation. If an event occurred within two years from the time of release to the community, it was coded as a recidivism event. Bivariate correlations and ROC/AUC analyses were run to determine the accuracy of LSI-R scores in predicting offender recidivism.

The first finding of this study is that the LSI-R had small-to-moderate predictive validity for the combined sample of offenders. In this sample, the new charge and new conviction outcomes both had AUC value of .622, while the return to prison and revocation AUC values were .626 and .641. These results for the combined sample are commensurate with some previous studies (Flores et al., 2006; Ostermann and Salerno, 2016), but smaller than the AUC values seen in others (Fass et al., 2008; Folsom and Atkinson, 2007). Bivariate correlations on LSI-R composite score and each recidivism event for the combined sample revealed Pearson's coefficients of .253, .253, .230, and .333 for new charge, new conviction, return to prison, and revocation, respectively; all were significant at the .01 level. These *r*-values were comparable to the .283 coefficient reported Flores et al. (2006), .300 in Folsom and Atkinson (2007), and .203 in Ostermann and Salerno (2016).

There are a couple considerations as to why AUC values for the combined sample are smaller in effect size compared to those reported in some prior validation tests. First, the recidivism measures used in this study were more conservative than some previous studies, utilizing new charge and new conviction rather than new arrest (Fass et al., 2008; Ostermann and Salerno, 2016). Because it is easier for an offender to be arrested than convicted on a new charge, studies using arrest as a recidivism measure cast a wider net under which events can be coded as recidivism. Thus, it is possible that recidivism events captured for offenders through broader measures in previous studies were not captured here. Second, the follow-up time to measure recidivism for the present study was two years, keeping with the NDDOCR’s definition, whereas some previous validation studies have utilized follow-up periods spanning to up to seven years, allowing for more time under which to capture a recidivism event. (Folsom and Atkinson, 2007; Wormith, Hog, and Guzzo, 2015). These variations in the definition of recidivism effect the calculation of AUC values and may be responsible for at least some of the differences in AUC values across studies (Myer, 2019).

The second finding came from the analyses of the community and institutional cohorts. For the ROC/AUC analyses, AUC scores between the community and institutional cohorts for the recidivism outcomes were disparate. In particular, the AUC values for each of the four recidivism outcomes of the community cohort were .665, .665, .732, and .749 (new charge, new conviction, return to prison, and revocation, respectively). As mentioned, these fall within the range of moderate to large effect sizes. Comparatively, AUC values for the recidivism outcomes on the institutional cohort were .595, .593, .536, and .597; here falling just at and/or beneath the .56 threshold for a small effect size. Differences again emerge between the community and institutional cohorts when considering results of the Pearson correlations. While the correlations for each recidivism outcome were significant in both cohorts, the coefficients for the community cohort were larger in size and all reached the .01 significance level (r = .280, r = .283, r = .315, r = .424). In contrast, the institutional cohort had smaller r-values at the .01 significance level for the new charge (r = .181), new conviction (r = .178), and revocation (r = .175) outcomes, and an r-value of .059 at the .05 significance level for the return to prison outcome.

The disparity in AUC values and correlation coefficients between the community and institutional cohorts may be attributed to several explanations. One such reason may be due to the construct validity of the LSI-R's items and/or subcomponents for offenders at different points in the correctional system. For example, scores on in the LSI-R subcomponents such as education/employment, financial, or accommodation may be more valid for an individual from the community cohort than from the institutional cohort, as an offender who is just released from an institution is likely to have employment or accommodation in place than an offender who has remained in the community. Thus, the institutional cohort offender may be flagged as higher risk due to constraints of the tool's definitions rather than real time risk factors at time of assessment.

The third major finding comes from the supplemental analyses conducted on the sample by race. The ROC analyses produced AUC values in the moderate range for Whites and in the small range for Native Americans, with significant differences in these values for the return to prison (.642 and .526) and revocation (.645 and .569) outcomes. These results diverge from the moderate to large AUC values reported by Wormith, Hogg, and Guzzo (2015), which utilized an Aboriginal Canadian sample of offenders compared to the present study's concentration on Native Americans. One reason as to why this disparity exists may be that the LSI-R was originally developed and validated by Andrews and Bonta (1995) against Canadian samples. While Aboriginal offenders do not comprise the majority of Canadians offenders or samples from past validation efforts, their inclusion could mean the LSI-R and its items are better calibrated for Canadian Aboriginal offenders than for Native American offenders, as Native Americans were not adequately considered when the tool was constructed (Wormith, Hogg, and Guzzo, 2015). This possibility would more likely apply to Native American offenders than White American offenders, on which the LSI-R has been repeatedly validated and recalibrated (Flores et al., 2006; Folsom and Atkinson, 2007; Schlager and Simourd, 2007; Fass et al., 2008; Austin, 2011; Chenane et al., 2015; Ellison et al., 2016; Olver, Stockdale, and Wormith, 2014; Ostermann and Salerno, 2016; Vose et al., 2009; 2013).

The bivariate correlations conducted on White and Native American cohorts in the present study found significant correlations between LSI-R composite score and recidivism outcomes at the 0.01 level, for each recidivism outcome. These results (r = .250, r = .254, r = .169, r = .285) for the Native American cohort are commensurate with the r-value reported for the Native American sample in Holsinger, Lowenkamp, and Latessa (2006). The accordance in findings between the LSI-R validation studies that have used Native American samples, compared to less congruent results reported between Native American and Indigenous samples from other countries, underscores the importance of continued validation efforts across racial categories within the context of countries and other jurisdictional considerations. Furthermore, the AUC values reported in the current study for the “Other” race cohort were also in the small to moderate, range. These supplemental findings on race converge with previous LSI-R validation studies that tested the LSI-R across racial categories, suggesting that while the LSI-R has demonstrated discriminant validity in general samples that are primarily white, it has less predictive power for individuals from other racial categories (Schlager and Simourd, 2007; Chenane et al., 2015; Ostermann and Salerno, 2016).

The present study and its findings are limited by several factors. First, the data used herein were secondary in nature and thus constrained the scope of research able to be conducted. For example, the outcome measures of new conviction and new charge were partially selected due to accessibility. Moreover, the LSI-R data were collected as day-to-day procedures by NDDOCR practitioners, rather than for the express purpose of research by the current study's research team. Consequently, some variable information was not available to conduct analyses on, such as individual-item scores for all 54 items. Such information would be useful in future studies to explore whether certain items or domains designated as predictive risk factors for general offender populations are also predictive for Native Americans, and whether there are risk factors exclusive to Native Americans not captured by the LSI-R (Holsinger, Lowenkamp, and Latessa, 2003; Wormith, Hogg, and Guzzo, 2015). Such unaccounted risk factors may exist because Native Americans are a diverse subgroup within themselves to the general population in both physical and intangible contexts. For example, Native Americans are culturally distinct from the dominant Western European culture in the United States, through their concepts of spiritual and communal social institutions (Wormith, Hogg, and Guzzo, 2015). Reservations may account for the geographic segregation of Native Americans from the broader population. The rural nature of reservations combined with the residual distrust and intergenerational trauma associated with colonization and forced resettlement may impact the degree to which Native Americans access resources for mental health and substance use issues (Holsinger, Lowenkamp, and Latessa, 2003; Brzozowski, Taylor-Butts, and Johnson, 2006). More broadly, this cultural and historical considerations may impact the very etiology of crime within the Native American subpopulation (Wormith, Hogg, and Guzzo, 2015). Additionally, because the LSI-R was administered by NDDOCR practitioners, data on inter-rater reliability was not available for the present study.

However, such information would be useful in future studies as a method to determine differences in AUC values (Holsinger, Lowenkamp, and Latessa, 2003; Myer, 2019) and to consider the possibility of rater biases.

This study tested the predictive accuracy of the LSI-R against a Midwest offender population, including a large sample of the critically understudied Native American offender demographic. AUC and bivariate analyses indicated that while the LSI-R had moderate predictive validity for the combined and White samples, it had lower predictive accuracy for the institutional and Native American samples. The differences in predictive accuracy between the institutional and community cohorts may be of particular interest to the NDDOCR. It is possible the lower predictive power for the institutional cohort is due to limitations of the LSI-R to account for an individual's temporal placement in the correctional system process. Thus, NDDOCR practitioners may consider offender status in terms of institutional or community origin when utilizing the LSI-R to determine intervention and resource needs.

In comparison to previous LSI-R validation studies on indigenous samples largely conducted outside the United States, the current study demonstrates that the moderate predictive validity reported for those indigenous samples (Hsu, Caputi, and Byrne, 2010; Wilson and Gutierrez, 2014; Wormith, Hogg, and Guzzo, 2015) is not necessarily transferable to Native Americans. Future research in Native American risk assessment has multiple tasks. First, additional validation studies on the LSI-R and newer risk assessment tools like COMPAS, ORAS, and PCRA should be conducted to augment the still scarce existing literature. Within that, studies should strive for larger cohorts of Native Americans for inclusion in the sample. The present study had an n = 798 for Native American participants, one of the largest samples of Native Americans included in LSI-R validation studies (Holsinger, Lowenkamp, and Latessa, 2003; Holsinger, Lowenkamp, and Latessa, 2006). Future studies should additionally seek validate the LSI-R and other risk assessments by subcomponents and individual items. As previously mentioned, the risk factors established as valid against general populations need to be validated against Native Americans specifically, to determine whether this risk instruments may be missing criminogenic risks unique to Native Americans offenders. For the NDDOCR, the present study's results suggest that the LSI-R has some, but small predictive accuracy for Native Americans. Practitioners should therefore utilize the tool on Native American offenders with consideration that supplemental information on the offender may be necessary when designing treatment and service plans for these individuals. Thus, this study has expanded on the literature pertaining to both the LSI-R's general validity and its validity against Native Americans, while also presenting objectives for future research and implications for current NDDOCR practitioners.

REFERENCES

Andrews, D. A., & Bonta, J. (1995). *The level of service inventory-revised.* Toronto, Canada: Multi-Health Systems.

Andrews, D. A., & Bonta, J. (2003). *The Psychology of Criminal Conduct* (3rd Ed.). Cincinnati: Anderson Publishing.

Andrews, D. A., Bonta, J., & Hoge, R. D. (1990). Classification for effective rehabilitation: Rediscovering psychology. *Criminal justice and Behavior*, *17*(1), 19-52.

Andrews, D. A., Bonta, J., & Wormith, J. S. (2006). The recent past and near future of risk and/or need assessment. *Crime & Delinquency*, *52*(1), 7-27.

Austin, J. (2011). *Evaluation of the North Dakota Department of Corrections and Rehabilitation Risk Assessment System.* National Institute of Corrections.

Blomberg, T., Bales, W., Mann, K., Meldrum, R., & Nedelec, J. (2010). Validation of the COMPAS risk assessment classification instrument. *College of Criminology and Criminal Justice, Florida State University, Tallahassee, FL.*

Bonta, J. (1996). Risk-needs assessment and treatment. In A. T. Harland (Ed.), *Choosing correctional options that work: Defining the demand and evaluating the supply* (pp. 18-32). Thousand Oaks, CA: Sage.

Bonta, J., & Andrews, D. A. (2007). Risk-need-responsivity model for offender assessment and rehabilitation. *Rehabilitation*, *6*(1), 1-22.

Bonta, J., LaPairie, C., & Wallace-Capretta, S. (1997). Risk prediction and re-offending: Aboriginal and non-aboriginal offenders. *Canadian Journal of Criminology*, *39*(2), 127-144.

Brenna, 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.

Brzozowski, J. A., Taylor-Butts, A., & Johnson, S. (2006). *Victimization and offending among the Aboriginal population in Canada*. Ottawa: Canadian Centre for Justice Statistics.

Clear, T. (1997). Developing a universal risk-assessment form: We're closer, but not there yet. *Community Corrections Report, 4*(4), 49–60.

Chenane, J. L., Brennan, P. K., Steiner, B., & Ellison, J. M. (2015). Racial and ethnic differences in the predictive validity of the Level of Service Inventory–Revised among prison inmates. *Criminal Justice and Behavior*, *42*(3), 286-303.

Dowden, C., & Andrews, D. A. (1999, May). What works in young offender treatment: A meta-analysis. In *Forum on corrections research* (Vol. 11, pp. 21-24). Correctional Service of Canada.

Ellison, J. M., Steiner, B., Brennan, P., & Chenane, J. L. (2016). Age Group Differences in the Predictive Validity of the Level of Service Inventory–Revised. *The Prison Journal*, *96*(6), 828–853.

Farabee, D., Zhang, S., Roberts, R. E. L., & Yang, J. (2010). COMPAS validation study: Final report. *Semel Institute for Neuroscience and Human Behavior, UCLA, Los Angeles, CA*.

Fass, T. L., Heilbrun, K., DeMatteo, D., & Fretz, R. (2008). The LSI-R and the COMPAS: Validation data on two risk-needs tools. *Criminal Justice and Behavior*, *35*(9), 1095-1108.

Flores, A. W., Lowenkamp, C. T., Holsinger, A. M., & Latessa, E. J. (2006). Predicting outcome with the Level of Service Inventory-Revised: The importance of implementation integrity. *Journal of Criminal Justice*, *34*(5), 523-529.

Folsom, J., & Atkinson, J. L. (2007). The Generalizability of the LSI-R and the Cat To the Prediction of Recidivism in Female Offenders. *Criminal Justice and Behavior*, *34*(8), 1044–1056.

Gottfredson, S. D., & Moriarty, L. J. (2006). Statistical Risk Assessment: Old Problems and New Applications. Crime & Delinquency, 52(1), 178–200.

Gutierrez, L., Wilson, H. A., Rugge, T., & Bonta, J. (2013). The prediction of recidivism with aboriginal offenders: A theoretically informed meta-analysis. *Canadian Journal of Criminology and Criminal Justice, 55*(1), 55–99.

Gutierrez, L., Helmus, L. M., & Hanson, R. K. (2016). What we know and don't know about risk assessment with offenders of Indigenous heritage. *Journal of Threat Assessment and Management, 3*(2), 97–106.

Holsinger, A. M., Lowenkamp, C. T., & Latessa, E. J. (2003). Ethnicity, gender, and the level of service inventory-revised. *Journal of Criminal Justice, 31*(4), 309-320.

Holsinger, A.M., Lowenkamp, C.T., & Latessa, E.J. (2006). Exploring the Validity of the Level of Service Inventory-Revised with Native American Offenders. *Journal of Criminal Justice, 34*(3), 331-337.

Holtfreter, K., & Cupp, R. (2007). Gender and Risk Assessment: The Empirical Status of the LSI-R for Women. *Journal of Contemporary Criminal Justice*, *23*(4), 363–382.

Hsu, C., Caputi, P., & Byrne, M. K. (2010). Level of Service Inventory–revised: Assessing the risk and need characteristics of Australian Indigenous offenders. *Psychiatry, Psychology, and Law, 17*(3), 35–367.

James, N. (2018). Risk and Needs Assessment in the Federal Prison System. *Congressional Research Service, 10*.

Johnson, J. L., Lowenkamp, C. T., VanBenschoten, S. W., & Robinson, C. R. (2011). The Construction and Validation of the Federal Post Conviction Risk Assessment (PCRA). *Federal Probation*, *75*(2), 16–29.

Jung, S., & Rawana, E. (1999). Risk and need assessment of juvenile offenders. *Criminal Justice and Behavior, 26*(1), 69–89.

Landenberger, N. A., & Lipsey, M. W. (2005). The positive effects of cognitive–behavioral programs for offenders: A meta-analysis of factors associated with effective treatment. *Journal of experimental criminology*, *1*(4), 451-476.

Latessa, E., & Lovins, B. (2010). The Role of Offender Risk Assessment: A Policy Maker Guide. *Victims & Offenders*, *5*(3), 203–219.

Latessa, E. J. (2015). Taking Risk Assessment to the Next Step. *Criminology & Public Policy*, *14*(1), 67–69.

Latessa, E. J., Lemke, R., Makarios, M., & Smith, P. (2010). The creation and validation of the Ohio Risk Assessment System (ORAS). *Fed. Probation*, *74*, 16.

Latessa, E., Smith, P., Lemke, R., Makarios, M., & Lowenkamp, C. (2009). Creation and validation of the Ohio risk assessment system: Final report. *Center for Criminal Justice Research, School of Criminal Justice, University of Cincinnati, Cincinnati, OH.*

Latessa, E., & Lovins, B. (2010). The Role of Offender Risk Assessment: A Policy Maker Guide. *Victims & Offenders*, *5*(3), 203–219.

Lipsey, M. W. (2009). The primary factors that characterize effective interventions with juvenile offenders: A meta-analytic overview. *Victims and offenders*, *4*(2), 124-147.

Lipsey, M. W., Landenberger, N. A., & Wilson, S. J. (2007) Effects of cognitive-behavioral programs for criminal offenders. *Campbell systematic reviews, 3*(1), 1-27.

Lovins, B. K., Latessa, E. J., May, T., & Lux, J. (2018). Validating the Ohio Risk Assessment System community supervision tool with a diverse sample from Texas. *Corrections, 3*(3), 186-202.

Lowenkamp, C. T., Holsinger, A. M., Brusman-Lovins, L. & Latessa, E. J. (2004). Assessing the inter-rater agreement of the level of service inventory revised. *Fed. Probation, 68,* 34.

Lowenkamp, C. T., Latessa, E. J., & Holsinger, A. M. (2006). The risk principle in action: What have we learned from 13,676 offenders and 97 correctional programs?. *Crime & Delinquency*, *52*(1), 77-93.

Luong, D., & Wormith, J. S. (2011). Applying Risk/Need Assessment to Probation Practice and its Impact on the Recidivism of Young Offenders. *Criminal Justice and Behavior*, *38*(12), 1177–1199.

Manchak, S. M., Skeem, J. L., Douglas, K. S., & Siranosian, M. (2009). Does Gender Moderate the Predictive Utility of the Level of Service Inventory—Revised (LSI-R) for Serious Violent Offenders? *Criminal Justice and Behavior*, *36*(5), 425–442.

Myer, A. J. (2019). Examining the Predictive Validity of the Static-99R on Native American Sex Offenders. *Justice Evaluation Journal*, *2*(2), 181-195.

Olver, M. E., Stockdale, K. C., & Wormith, J. S. (2014). Thirty years of research on the Level of Service Scales: A meta-analytic examination of predictive accuracy and sources of variability. *Psychological Assessment*, *26*(1), 156–176.

Ostermann, M., & Salerno, L. M. (2016). The Validity of the Level of Service Inventory–Revised at the Intersection of Race and Gender. *The Prison Journal*, *96*(4), 554–575.

Quinsey, V. L., Harris, G. T., Rice, M. E., & Cormier, C. A. (2006). *Violent offenders: Appraising and managing risk.* American Psychological Association.

Rice, M. E., & Harris, G. T. (2005). Comparing effect sizes in follow-up studies: ROC Area, Cohen's d, and r. *Law and human behavior*, *29*(5), 615-620.

Schlager, M. D. (2005). *Assessing the reliability and validity of the Level of Service Inventory-Revised (LSI-R) on a community correction sample: Implications for corrections and parole* policy*. Rutgers the State University of New Jersey-Newark.

Schlager, M. D., & Simourd, D. J. (2007). Validity of the Level of Service Inventory—Revised (LSI-R) Among African American and Hispanic Male Offenders. *Criminal Justice and Behavior*, *34*(4), 545–554.

Shepherd, S. M., Adams, Y., McEntyre, E., & Walker, R. (2014). Violence risk assessment in Australian Aboriginal offender populations: A review of the literature. *Psychology, Public Policy, and Law*, *20*(3), 281–293.

Smith, W.R. & Smith, D.R. (1998). The consequences of error: Recidivism prediction and civil-liberation ratios. *Journal of Criminal Justice,* 26, 481-502.

Steen, S., Opsal, T., Lovegrove, P., & McKinzey, S. (2013). Putting parolees back in prison: Discretion and the parole revocation process. *Criminal Justice Review*, *38*(1), 70-93.

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*(4), 459–471.

Vose, B., Smith, P., & Cullen, F. T. (2013). Predictive Validity and the Impact of Change in Total LSI-R Score on Recidivism. *Criminal Justice and Behavior*, *40*(12), 1383–1396.

Whiteacre, K. W. (2006). Testing the Level of Service Inventory–Revised (LSI-R) for Racial/Ethnic Bias. *Criminal Justice Policy Review*, *17*(3), 330–342.

Wilson, H. A., & Gutierrez, L. (2014). Does One Size Fit All?: A Meta-Analysis Examining the Predictive Ability of the Level of Service Inventory (LSI) With Aboriginal Offenders. *Criminal Justice and Behavior*, *41*(2), 196–219.

World Prison Brief. (2020). Retrieved from https://www.prisonstudies.org/highest-to-lowest/prison_population_rate?field_region_taxonomy_tid=All

Wormith, J. S., Hogg, S. M., & Guzzo, L. (2015). The predictive validity of the LS/CMI with aboriginal offenders in Canada. *Criminal Justice and Behavior*, *42*(5), 481–508.