DYNAMIC CHANGES IN THE LEVEL OF SERVICE INVENTORY-REVISED (LSI-R) AND THE EFFECTS ON PREDICTION ACCURACY

by

Thomas K. Arnold

B.A., St. Cloud State University, St. Cloud, 1983

A Thesis

Submitted to the Graduate Faculty

of

St. Cloud State University

in Partial Fulfillment of the Requirements

for the Degree

Master of Science

St. Cloud, Minnesota

March, 2007

This thesis, submitted by Thomas K. Arnold in partial fulfillment of the requirements for the Degree of Master of Science at St. Cloud State University, is hereby approved by the thesis evaluation committee.

Chairperson: Richard Lawrence

Robert Prout

Michael Mayhew

Dean
School of Graduate Studies

DYNAMIC CHANGES IN THE LEVEL OF SERVICE INVENTORY-REVISED (LSI-R) AND THE EFFECTS ON PREDICTION ACCURACY

Thomas K. Arnold

The Level of Service Inventory-Revised (LSI-R) is one of the more popular and effective risk/needs instruments in corrections, and while it is well known that the LSI-R scores may change over time, there are few studies on the dynamic predictive validity of the LSI-R. This study used LSI-R scores from a Central Minnesota County Community Corrections Agency to replicate earlier studies on dynamic changes in LSI-R scores and also looked at several other aspects of changes in LSI-R scores. An attempt made to answer several questions with this analysis. 1) Are changes in LSI-R scores accompanied by corresponding changes in outcome, as measured by arrests resulting in conviction? 2) Do the magnitude and direction of the starting and ending score values have any relationship with the outcome? 3) Does the length of time between assessments make a difference in the accuracy of the outcome? 4) What is actually changing? Is the offender actually changing or is the new score just a reflection of some newer, more accurate, information that was obtained by the caseworker? 5) Are the properties of score changes between assessments 1 and 2, replicated between assessments 2, 3, and 4? The results of this analysis show clearly that the scores generated in assessments 2, 3, and 4 are much more predictive of outcome than the scores for assessment 1. Changes between assessments 2, 3, and 4 do not seem to affect prediction outcome, however this could be do to the effects of repeat sampling and regression toward the mean. There is no definitive answer to whether the magnitude or directions of score changes affect the outcome. The length of time between assessments does not seem to affect predictive accuracy. Changes in the Criminal History, Alcohol/Drug, and Emotional/Personal sub scale scores seem most predictive of outcome.

Month Year

Approved by Research Committee:

Richard Lawrence Chairperson

ACKNOWLEDGEMENTS

I would like to thank Dr. Lawrence, Dr. Prout and Dr. Mayhew for their help in preparing this thesis.

TABLE OF CONTENTS

Chapter		Page

LIST OF TABLES

Table		Page

LIST OF FIGURES

Chapter 1

INTRODUCTION AND PROBLEM STATEMENT

Thesis Purpose

The purpose of this thesis is to add to the understanding of the process and prediction of change as it occurs in a criminal justice setting. Specifically, this research will look at how the scores generated by the Level of Service Inventory-Revised (LSI-R; Andrews & Bonta, 1995) change over time, and whether those changed scores are a more accurate reflection of the offender's risk of future criminal activity than the original scores.

The LSI-R was initially developed as the Level of Supervision Inventory (LSI; Andrews, 1982). The LSI-R uses a structured interview to answer 54 yes/no questions about subjects that have been shown to relate to criminal activity (Andrews & Bonta, 2006). Many of the predictive properties of the LSI-R have been heavily researched, prompting Hollin (2002) to comment that the LSI-R had the strongest 'research pedigree' of any risk assessment instrument.

There is a property of the LSI-R that has not been as well researched. It is understood that the scores on LSI risk assessments are dynamic and may change between assessments and previous research has shown that the LSI scores on a subsequent assessment are better predictors of outcome than the scores on the first LSI assessment (Andrews & Robinson, 1984). More research is needed in this area however, as the previous studies were small and few in number (Andrews and Robinson, 2003).

1

A recent review of popular risk assessment instruments found that the LSI-R was one of the more accurate assessments (Andrews, Bonta, & Wormith (2006).

A recent survey of correctional institutions found that theperformed

This analysis will attempt to answer several questions. 1) Are changes in risk assessment scores accompanied by corresponding changes in arrests resulting in conviction? 2) Do the starting and ending score values have any relationship to the outcome? That is, are high scores that change more or less accurate than low scores that change, and does the size or the direction of the change in score affect the predictive accuracy? 3) Does the length of time between assessments make a difference in the accuracy of the outcome? 4) Are the properties of score changes between assessments 1 and 2, replicated between assessments 2, 3, and 4?

Why This Research was Needed

The reasons for this research can be categorized under general need and specific need. The general need for research into dynamic risk assessment can be further classified using the utilitarian approach to punishment that Jeremy Bentham (1748-1832) proposed and Hollin (2004:2-3) paraphrased, “punishment should achieve four outcomes; 1. To prevent crime; 2. If prevention is not achieved, then convince a criminal to commit a less serious crime; 3. To reduce the harm inflicted during a crime; and 4. To prevent crime as cheaply as possible.” Risk assessment is important for 1,2 and 4, crime prevention, crime reduction, and cost control.

The General Need For Risk Assessment Research

Prevention of crime. The prevention of crime is so central to the subject of risk assessment that it is often not stated since it is an underlying assumption. Bonta (2000) writes, “There are few activities in corrections as important as the assessment of offenders.” He goes on to point out some of the dangers associated with inaccurate risk assessment such as placing a dangerous offender on parole. Byrne (2006) categorized risk assessment as “critical to the success of community corrections.”

Convince a criminal to commit a less serious crime. Convincing criminals to commit less serious crimes is the role of treatment. Two studies have shown rather convincingly that accurate risk assessment is essential to effective treatment (Andrews, Bonta, & Hoge, 1990; Andrews, Zinger, Hoge, Bonta, Gendreau, & Cullen, 1990).

Preventing crime as cheaply as possible. According to Austin (2004), the need for accurate risk assessment is greater now that it has been in the past due to the rising costs of corrections. The annual corrections costs in the United States in 2001 were in excess of $38 billion and were increasing at an average annual rate of over 6% (Stephan, 2004). Of the $38 billion spent, about $30 billion was used for the prison system. Almost 1.5 million offenders were incarcerated in 2005 and this number has been growing at 2% to 3% per year (Harrison & Beck, 2006). There were over 4 million offenders on probation and 780 thousand on parole in 2005, and these numbers have been growing at 2.5% and 1.4% annually (Glaze & Bonczar, 2006). From the 2001 corrections costs, which show that the total prison cost is over three times the cost for probation and parole, and the 2005 figures that show that the number of offenders on probation and parole is three times the number of prisoners, it can be estimated that the cost of keeping an offender in prison is nine times the cost of keeping an offender in the community.

Keeping an offender in the community is also less costly in other ways. Over 55% of the incarcerated offenders in 1999 had minor children, with over 2% of the male and 10% of the female offender’s children placed in foster care (Mumola, 2000), which puts an addition burden on public services. Incarceration of parents can cause a variety of problems for young children (Parke, & Clarke-Stewart, 2001), which may lead to further problems in the years ahead. Certainly, placing the parents on probation in the community would help ameliorate this problem. Probation is only practical if the risk of doing so can be estimated so as to assure the public safety (Austin, 2004).

The Specific Need for Research on the LSI-R

Research into the dynamic properties of the LSI-R was long overdue. Andrews and Bonta (2003), the developers of the LSI-R, had indicated that further research was needed in this area. The LSI-R is a popular risk assessment instrument that is currently used by many corrections agencies. The results of a National Institute of Corrections (2003) survey of corrections departments in the United States revealed that 22% of agencies reporting use the LSI-R for risk assessment.

Many of the corrections agencies use the LSI-R for reassessment of offenders. The practice of periodic reassessment is a recommended practice for corrections agencies (Center for Substance Abuse Treatment, 2005) and reassessment after any change in status, with at least an annual reassessment, is suggested as the criteria to use with the LSI-R (National Institute of Corrections, 2004).

Previously, there were only four known studies into the dynamic properties of the LSI-R. Three were done in Canada on small samples (50-60) of Canadian offenders (Andrews & Robinson, 1984; Motiuk, 1991; Motiuk, Bonta & Andrews, 1990) and one was done in England, Wales, and Jersey with a sample of 360 offenders (Raynor, In Press). Research into the dynamic properties of the LSI-R had not been replicated in the United States and this, in and of itself, is a concern, since research has shown that risk assessment instruments developed in one locale do not always perform as well in other locations (Wright, Clear, & Dickson, 1984).

In addition to the benefit derived from studying the dynamic properties of the LSI-R, this study adds to the available information on the predictive properties of the LSI-R in Minnesota. Only one known study had been done previously on the predictive validity of the LSI-R in Minnesota (Jenson, T.D., 1998).

This research addresses two issues that have not been studied before. The first is a study of the performance of the LSI-R after multiple reassessments. The data available for this study provided the opportunity to study multiple assessments and reassessments that were taken over the period of several years. This provided the chance to determine how dynamic changes in LSI-R scores affected the ongoing predictive accuracy of the LSI-R. The second is the effect of various time periods between LSI-R assessments on prediction accuracy. Both J. Bonta (personal communication, February 21, 2007) and C. Lowenkamp (personal communication, March 24, 2007) had expressed concerns that insufficient time periods between assessments could affect the predictive ability of the LSI-R.

Thesis Overview

Chapter 2 provides an in depth exploration of the history of risk prediction through the ages. This is followed by a review of recent developments and research in risk prediction efforts, including an in-depth review of the instrument used in this research. Next, an examination of personality and change theory as it pertains to offender change is provided. Finally, there is an analysis of some of the methodological considerations that could affect the reliability and generalizability of this research.

Chapter 3 shows how the data collection and manipulation was done. The statistical tools that were used are explained. Finally, there is an examination of the variables used and some of the assumptions that were made.

Chapter 4 shows the results of the analysis. The data is explained in a somewhat linear fashion, with a general progression from the largest sample to the smallest for each stage of the research.

Chapter 5 provides an in depth analysis of the results. Possible explanations for the outcomes found are given.

The Appendices follow the reference section. There is a copy of the 54 LSI-R questions and information on the makeup of the LSI-R.

Chapter 2

LITERATURE REVIEW

THE HISTORY OF RISK ASSESSMENT

A Risk Assessment Classification System

A useful schema for the categorization of risk assessment instruments was used by Bonta (1996) and has become somewhat of a standard when describing such tools. Bonta describes three generations of risk assessment, and alludes to a fourth. He points out that the term "generation" is somewhat of a misnomer as any or all of the four generations of risk assessment may be in use at any given time. The term generation is used because there has been a sequential development of the different forms of offender risk assessment. His categorization schema will be used in this project.

First Generation Risk Assessment. Bonta describes the first generation of risk assessments (1G) as subjective assessment, professional judgment, intuition, or gut-level feelings. First generation risk assessments usually involve collection of relevant information and possibly an unstructured interview with the offender. The worker uses his or her professional judgment to determine the best course of action. Andrews et. al. (2006) report that first generation risk assessments have a low (.12) prediction rate.

Second Generation Risk Assessment. In second-generation risk assessment (2G), factors that have been statistically associated with greater risk of recidivism are used to create a scoring system. The offender is assigned a value for each item and the magnitude of the sum of the values is used to assign a risk level to the offender. Second generation risk assessment instruments generally measure historical items which are static, or unchangeable, in nature, making them of little use in determining treatment options, since treatment assumes that there is some possibility of change.

The Third Generation of Risk Assessment. Third generation risk assessment (3G), which includes the LSI-R, measures both offender risks and criminogenic needs. Risks are often similar to those items used in second generation instruments and may include fairly static items such as number of previous incarcerations. Criminogenic needs are factors that are related to recidivism but are dynamic, or changeable. One criminogenic need is employment, since unemployed offenders are more likely to recidivate than employed offenders. (Andrews et. al., 2003). By identifying criminogenic needs, the caseworker is able to target treatment where it can be most effective.

The Fourth Generation of Risk Assessment. Andrews, et. al. (2006) describe fourth generation assessments (4G) as being guides to service delivery. These tools combine third generation risk assessment methods with case management tools that help practitioners make decisions about service plans and service delivery. The Level of Service/Case Management Inventory (LS/CMI; Andrews, Bonta, & Wormith, 2004) is a fourth generation assessment instrument.

Built in Risk Assessment

It does not take a great stretch of the imagination to suppose that the first generation of risk assessment, gut level feelings, was developed as a survival mechanism. Ancient man must have developed some way to identify which people to trust and which to not trust. There is some scientific basis to believe this is the case. Cosmides (1989) found evidence that people have built-in algorithms for detecting cheaters. Chiappe, Brown, Dow, Koonz, Rodriguez, & McCulloch (2004) found that people look more closely at cheaters and are more likely to remember them.

One of the first recorded instances of (1G) risk assessment in corrections is found in the mid 1800s with the work of John Augustus, who is considered the father of probation. John, a shoemaker by trade, sat in the local courtroom watching trials. He posted bail for people that he thought might be worthy candidates for rehabilitation. Since John had limited funds, he needed to make sure that his "client" had some likelihood of successfully completing treatment and paying him back before he posted his or her bail. His method of risk assessment was based on observation and questioning and amounted to trusting his gut as to the risk level of a client (Panzarella, 2002).

Generally, (1G) risk assessment is not recommended as the sole method of predicting risk. Grove and Meehl (1996), Grove, Zald, Lebow, Snitz, & Nelson, (2000), and Meehl (1954) found that structured clinical assessments are usually more accurate than unstructured clinical assessments. Additional research in correctional settings, cited by Andrews, et. al. (2006) found that clinical judgment alone, is considerably less accurate than structured methods of risk assessment.

Early Attempts to Classify Offenders

The first structured attempts at classification of people date back to ancient times. These early attempts at offender classification are generally not well documented, however O'Connor (2006) has compiled a brief and accessible overview.

These early methods are included here, partly for completeness, and partly because, as O'Connor points out, there is an element of accuracy in these early methods. Since this paper looks at evidence for the dynamic nature of offending, and many of these older methods suggest that offending is static in nature, it would be remiss to bypass this investigation. By examining methods that were used in the past, we can often find information that pertains to our present situation. With even the best prediction instruments topping out around at a 50% accuracy rate (Andrews et. al., 2006), it would be worthwhile to explore factors that may explain some of the other 50% of the variance.

Using Facial Features in Classification. The use of facial features to classify people is called Physiognomy. Physiognomists, who developed one of the earliest structured attempts at classification, suggest that there is a link between unusual physical appearance (mostly the face, ears, and eyes), and criminal behavior. O'Connor lists J. Baptiste della Porte (1535-1615) as the modern founder of this idea, with Johan Kaspar Lavater (1741-1801) as being another early practitioner. While these may be the best-known contemporary practitioners, there are indications that Physiognomy has been practiced for over two thousand years. Evans (1969) reports that the practice of Physiognomy was used in ancient Greece by Pythagoras (582-507 B.C) in the sixth century B.C. and was also used by Aristotle (384-322 B.C.).

While O'Connor (2006) notes that Physiognomy has been largely discredited, and is not used in contemporary practice, recent research suggests that there may actually be some basis for the proposition that facial features are related to criminal behavior. Research done by Agnew (1984) indicated that adolescents with unattractive faces had a 9% greater chance of being involved in delinquent behavior in school.

Facial disfigurement may also be related to criminal behavior. In a study done by Kurtzberg, Mandell, Lewin, Lipton, and Scuster (1978), offenders with facial disfigurement were given plastic surgery to correct their condition. The recidivism rate for the group that had plastic surgery was 50% vs. 79% for the control group. These finding suggests that there may be a modest link between facial features, which are normally rather static features, and criminal behavior.

<u>Phrenology in Offender Classifiction.</u> Phrenology is the study of the shape of a person's skull. O'Connor lists Franz Joseph Gall (1758-1828) and his pupil John Gaspar Spurzheim (1776-1832) as the two most eminent early practitioners of this science. O'Connor points out that some of the bumps that were classified as problematic by Gall and other Phrenologists have been found to be related to brain features such as the amygdala and hippocampus. Amen (1998), using Positron Emission Tomography to examine the brain activity of people with behavior problems, has found that unusual activity in certain parts of the brain, including the amygdala and hippocampus, is associated with violent or other undesirable behavior. Without treatment, these problems would also be considered static factors that may influence a person's behavior and could be associated with criminal activity. Again, an element of truth is found in old science.

Criminal Anthropology in Offender Classification. O'Connor credits Cesare Lombroso (1835-1909) with being one of the first practitioners of Criminal Anthropology. Criminal Anthropologists attributed much of criminal behavior to hereditary physical and psychological features that they considered to be indications of a primitive and brutish nature. Lombroso (1876/2006) wrote the book "Criminal Man" in several editions in which he categorized offenders by various characteristics such as the shape of their hands, skulls, hair, etc. Lombroso and Ferrerro (1893/2004) also wrote a book called "Criminal Woman, the Prostitute, and the Normal Woman" to classify females into criminal and non-criminal types. Enrico Ferri (1856-1929), a confederate of Lombroso created the term, "born criminal" to describe certain types of offenders (Sellin, 1958). The term seems to imply that criminality was their destiny. O'Connor points out that these theories were largely discredited in the early 1900s, but Rafter (2004) reports that as late as the 1930s and 1940s, Earnest A. Hooton (1887-1954) was working to develop a criminal typology based on physical characteristics.

The work of Lombroso is worth noting for two reasons. The first is that, while Lombroso's typology of physical characteristics has largely been forgotten, his methods of classification are at the heart of modern criminology (Wolfgang, 1961). The second reason Lombroso's work is important that we now know that many predispositions towards criminal behavior are, in fact, heritable (Walsh, 2002). Of course, this evidence for heritability of criminal predispositions is not quite the same evidence that Lombroso was using. Walsh reports that modern researchers use sophisticated genetic studies. It does suggest however that perhaps Lombroso was on to something.

<u>Actuarial Risk Prediction</u>

Actuarial risk prediction is classified by Bonta (1996) as a (2G) risk prediction instrument. Burgess (1928) is one of the first researchers to suggest the use of statistical methods to predict parolees' risk of reoffense (Bonta, 1996; Latessa & Allen, 2003). This method was discovered by Hornell Hart (1923) when he reanalyzed data collected by Warner (1923) on 680 prisoners released on parole in Massachusetts. Hart identified several factors that could help guide parole decisions and suggested the use of statistical inference to determine which offenders were most likely to recidivate. Burgess, using the methods that Hart had suggested, studied 3,000 paroled men in Illinois offenders in and developed a list of 22 static traits that were common to offenders. This list included items such as type of offense, marital status, number of previous offenses, age, etc. He developed a scoring system where an offender was scored a 1 if a particular trait was present and a 0 if it was not. Some of these traits, and this scoring method, are still used today in the LSI-R. Oddly, Burgess found that criminals who committed crimes alone were more likely to recidivate, while the LSI-R scores them more likely to offend if they commit crimes with other offenders.

Bonta (1996) credits Glueck and Glueck (1950) as also being pioneers in the art of statistical prediction. Glueck and Glueck studied 500 delinquent boys and 500 non-delinquent boys who lived, in Massachusetts. They studied the two groups for quite some time and developed detailed actuarial tables comparing many different behavior attributes between the two groups. The work of the Gluecks is still used to guide research (Sampson & Laub, 1993; 2005).

<u>The Development of the LSI-R.</u>

The immediate predecessor of the LSI is the Wisconsin Case Management Classification System (CMC). The CMC was one of the first classification systems that included dynamic, or changeable factors called needs, and so the CMC is one of the first 3G risk assessment instruments. The CMC was developed in Wisconsin in 1975 and became a model system recommended for use by the National Institute of Corrections (Latessa & Allen, 2003). The CMC was an improvement over previous systems for measuring risk, but there were problems with the needs component of the CMC due to a lack of research into the relationship between the needs and criminal activity (Andrews & Bonta, 2003).

In order to address some of the issues with the CMC,) the Level of Supervision Inventory (LSI; Andrews, 1982) was developed in Canada during the late 1970s with funding provided by the province of Ontario (Bonta & Motiuk, 1987). The LSI was later revised as the Level of Service Inventory-Revised (LSI-R; Andrews and Bonta, 1995). The LSI-R is a structured interview with 54 yes/no response items that are scored as either a 1 or a zero. The result is a score from 0 to 54 with low scores indicating a low probability of criminal activity and high scores indicating a high probability of criminal activity. The LSI-R has ten sub-scales, which are related to the primary factors associated with risk of criminal conduct. One of the advantages of the LSI-R over other risk assessment instruments is the ability to guide treatment decisions (Andrews & Bonta, 2003). The LSI-R will be discussed in more detail later in this chapter.

The Fourth Generation of Risk Assessment

Fourth generation (4G) assessments guide and follow the supervision process from offender intake through release (Andrews, Bonta, & Wormith, 2006). The 4G assessments assess risks, strengths, needs, and offender responsivity to treatment. They can be linked to service plans and service delivery and measured through intermediate outcomes. They are designed to maximize adherence principles of effective treatment and to provide information that can improve treatment outcome in the future. The most well known 4G systems are the Wisconsin Correctional Assessment and Intervention System (CAIS), the Correctional Offender Management Profiling for Alternative Sanctions (COMPAS), the Offender Intake Assessment (OIA), and the Level of Service/Case Management Inventory (LS/CMI).

Andrews and Bonta (2003:86) refer to these factors as the "big eight" risk factors of criminal conduct: "antisocial attitudes, antisocial associates, a history of antisocial behavior, antisocial personality pattern, problematic circumstances at home (family/marital), problematic circumstances at school or work, problematic leisure circumstance, and substance abuse."

Andrews and Bonta (2003) report that the LSI-R is primarily based upon Social Learning Theory (Akers, 1973; Bandura, 1977), and many of its constructs are related to the offender's relationship to others and society. To some extent, it may be thought of as a measurement of a person's 'stake in conformity' (Toby, 1957). A person who doesn't break the law, has a job, saves money, has stable family relationships, etc. would score low on the LSI-R and would, presumably, be less likely to break the law. A person who commits crimes, doesn't work, has no money, and has few friends, etc. would score high on the LSI-R and would be expected to be more likely to break the law.

The

In a meta-analysis of LSI-R research, Gendreau, Goggin and Smith (2002) found that the LSI-R had a weighted effect size of .39. In a comparison of risk assessment instruments, Andrews, Bonta, and Wormith (2006) found that most structured risk prediction tools have a prediction accuracy of between .20 and .40. The LSI-R appears to perform as well or better than most of the risk assessment instruments reviewed. Latessa & Allen (2003) called the LSI-R one of the most accurate risk assessment tools available, and Hollin (2002) said the LSI-R had the strongest 'research pedigree' of any risk assessment instrument.

The items on the LSI-R have been shown to relate to the risk of recidivism, criminogenic and non-criminogenic needs, and general responsivity to treatment (Gendreau, Goggin, & Little, 1996; Lowenkamp and Latessa, 2001; Girard and Wormith, 2004). The LSI-R has been found to be both reliable and valid as a risk assessment instrument with many offender populations (Simourd & Malcolm, 1998; Gendreau, Goggin & Smith, 2002; Simourd, 2004).

Not all factors in the LSI-R are dynamic. Some factors on the LSI-R are static factors such as the number of prior arrests, which will never go down. Other factors, such as drug abuse, or employment status are considered dynamic factors because they can change over time (Andrews and Bonta, 2003).

Bonta (1996) writes that the advantage of the LSI-R over (2G) actuarial type prediction is its ability to inform treatment decisions. Since (2G) instruments measure historic variables that generally cannot change, there is no opportunity for change. The dynamic variables used on the LSI-R such as employment status, for example, can change. Employment can then become a treatment target for the caseworker. In theory, get the offender working, and he or she will not offend as frequently.

Previous Research on the Dynamic Properties of the LSI-R

There have been few studies on the dynamic properties of the LSI-R (Andrews and Bonta, 2003). The most notable studies to date were done by Andrews and Robinson (1984), Motiuk (1991), Motiuk, Bonta and Andrews (1990), and Raynor (In Press). The study by Raynor used sample of 360 offenders in Great Britain, and the other studies used small samples of between 50 to 60 Canadian offenders. As far as is known, this research has not been replicated in the United States.

Andrews and Robinson (1984) studied 57 offenders in Canada who had both a test and retest done with the original Level of Supervision Inventory (LSI). The offenders had a follow-up period of at least 18 months. They divided the offenders into four levels by LSI score. Offenders with a score of 0-7 were placed in the low risk category, LSI scores of 8-11 were moderate risk, LSI scores of 12-23 were high risk and LSI scores of 24+ were placed in the very-high risk category. They totaled the recidivism and outcome rates for offenders using the initial and retest LSI categories and compared the results in a table. Recidivism was measured by reports from probation officers and self-reports by offenders. The outcome criteria was based on a score from 0 to 2 in which early termination or closure without recidivism was coded as a 0, regular termination without recidivism was coded as a 1, and recidivism was coded as a 2.

Andrews and Robinson found that the offenders were more accurately placed in risk categories by the retest LSI than the initial LSI. They concluded that the retest LSI predicted recidivism and outcome better than the initial LSI.

When they measured the outcome for the LSI subtotals using both the initial and retest values, they found that all of the retest LSI subtotals except Attitudes more strongly predicted recidivism than the initial LSI subtotals. The retest LSI subtotals most strongly linked with recidivism were Companions and Leisure/Recreation. The retest LSI subtotals with the weakest links to recidivism, besides Attitudes, were found in the Family and Emotional/Personal subtotals.

Motiuk (1991) studied the post-release outcomes of 54 Canadian offenders given an LSI assessment upon intake to prison and then given a follow-up LSI before release. He used two LSI risk categories, low (0-19) and high (20-54). He found that when retest risk level increased to the high level from initially being low, post–release remands, incarceration, recidivism, and parole violation increased. When retest risk level decreased to the low level from initially being high post–release remands, incarceration, recidivism, and parole violation decreased. When outcomes for violent recidivism, violent re-offense, and Federal sentence were compared with LSI intake/retest risk levels, the results were mixed. When LSI risk levels decreased from the first to second assessment, the levels of violent recidivism, violent re-offense, and Federal sentence decreased. When LSI risk level increased, the incidence of violent recidivism, violent re-offense, and Federal sentence did not show a corresponding increase.

Motiuk, Bonta and Andrews (1990) studied 55 adult male offenders in Ontario Canada who were administered both intake and retest LSI assessments. They separated the offenders into three LSI risk level groups for both the initial and retest LSI. The groups used were low (0-14); moderate (15-20); and high (21+). They found that the retest LSI more accurately classified the risk levels of the offenders. The outcome criteria were recidivism and reincarceration.

When they analyzed the correlation rates between the intake LSI and retest LSI sub scales and reincarceration, they found that retest scores for Education/Employment, Accommodations, and Drug/Alcohol were more highly correlated with a negative outcome, while the initial scores for Financial, Family/Marital, Leisure, Companions, and Attitude sub scales were more predictive than retest scores. The Emotional/Personal sub scale was more predictive at retest of incarceration and more predictive at the initial test of recidivism. A regression analysis was done on the test and retest LSI scores to determine which test was better able to explain variance between the scores and outcome. It was found that the retest scores had a 107% percentage gain in explained variance (PGV) in predicting incarceration and a 64% PGV in predicting general recidivism.

Raynor (In Press) studied 360 offenders in the British Isles who had follow-up assessments done with the LSI-R. Due to concerns with regression towards the mean in the follow-up LSI-R scores he split two samples of offenders, one from England and Wales, and the other from Jersey, into increasing and decreasing categories for both the above and below average offenders. He compared the recidivism rates for offenders from both locations whose scores were below average on the first LSI-R assessment and decreasing, below average on the first LSI-R assessment and increasing, above average on the first LSI-R assessment and decreasing, and above average on the first LSI-R assessment and increasing. He found that for England and Wales offenders with above average scores on the first assessment, offenders with increasing scores on the second assessment had significantly larger reconviction rates ($p<.01$) than offenders with decreasing scores. For Jersey offenders with above average scores on the first assessment, offenders with increasing scores on the second assessment had larger reconviction rates than offenders with decreasing scores, but the difference did not reach the level of significance ($p=.06$). He speculated that the lack of significance for the Jersey offenders may have been due to a small sample size of 21 offenders. Offenders from both groups of offenders who started with below average score on the first assessment, had significantly higher ($p<.05$) reconviction rates if their scores on the second LSI-R were increasing than if their scores were decreasing. When he combined all offenders with increasing LSI-R scores and compared them with offenders with decreasing scores, he found that the offenders with increasing scores had higher reconviction rates (67%) than offenders with decreasing scores (42%).

CURRENT RESEARCH

If a caseworker can help an offender find employment, which generally

Current Actuarial Risk Prediction.   Bonta (1996) mentions several other pioneers in the statistical prediction field and reports that the instruments they have developed perform adequately as risk prediction tools. He does not favor their use since actuarial methods do not provide a direction for treatment.   generally use historical data that measures static variables that do not change, they.

Basic Premises

In doing this research, there are two basic underlying premises. The first premise is that offenders can and do change their offending behavior, and the second premise is that this change can be measured. If these two premises are in fact valid, then the question of whether the LSI-R can measure this change is something that is worth an attempt at answering.

The first premise, “offenders can and do change their offending behavior”, is not without its detractors. Some researchers, despite the rather vociferous arguments of Andrews and Bonta (2006), are still using static risk prediction tools that measure

historical facts, rather than current facts. A study done by Glover, Nicholson, Hemmati, Bernfeld, and Quinsey (2002) that compared 10 risk assessment tools, used all static risk instruments. This brings up the question, “Why were no dynamic risk predictors tested?” The authors claim that dynamic predictors are only suitable for making short-term predictions. Their assertion calls into question the premise of whether offenders change.

The ultimate answer to the question of whether offenders stop offending is not really in doubt. As Moffit (1993) has pointed out, in the long term, over a lifetime, almost all offenders stop offending. This has been verified through records that go back over a hundred years and span several countries. We know that most offenders eventually stop offending and the first premise, “offenders can and do change their offending behavior”, is generally true.

Saying that offenders usually stop offending is not to say that this change occurs quickly. Sampson and Laub (2005) argue that we need to look at changes in offending over the life course. They claim that offenders go through a developmental process where they move in to, and out of, a criminal lifestyle. Andrews and Bonta (2006) point to a growing body of research that indicates many factors that lead to offending behavior are personality factors. This brings up the question that was explored by Heatherton and Weinberger (1994), “Can personality change?” and if it can, how will we know it has?

Or, it is possible that offenders with offending type personalities can stop offending, and if so, how will we know that this has occurred? We don't really know.

We do know that the majority of offenders do not stop offending in the short term. Langan and Levin (2002) report that 67.5% of prisoners from 15 states that were released in 1994 were rearrested for a serious crime within three years. One could hope that the other 37.5% of the offenders were no longer offending, but this is almost certainly not true. The Minnesota Bureau of Criminal Apprehension (BCA) (2005) reported that only 50% of the crimes committed in Stearns County in 2005 resulted in an arrest and conviction. What percentage of those unsolved crimes were committed by the 37.5% of offenders who were not arrested within three years? Again, we don't really know.

The state of the art in determining whether a person has stopped committing crimes is still rather limited. There have been attempts at discovering how desistance from crime occurs. Maruna (2001) studied offenders who stopped offending and found that those offenders could usually point to some event in their lives that caused them to change lifestyles. The presence of a story about why they were not offenders anymore, which he called "redemption script", was seen as an indicator that they had turned their lives around. In theory, a test for the presence of a redemption script would be one way to tell if offenders have changed. It is not known whether this has ever been tried, and the methodological obstacles to using this method could prove insurmountable.

The fact is that determining with 100% accuracy whether an offender has stopped committing crimes is difficult, if not impossible, to do. This uncertainty in predicting future behavior is the whole basis of risk prediction. According to Andrews, Bonta, and Wormith (2006), most risk prediction tools have a prediction accuracy of between 20% and 40%. This is the current state of the art in risk prediction.

The Case for the LSI-R

The benefit of the LSI-R, according to Andrews and Bonta (2006) is that it measures dynamic (changeable) factors, called criminogenic needs, which are associated with increased criminal behavior. Gendreau, Goggin and Smith (2002) have found the LSI-R to be an accurate predictor of the risk of offending. Andrews, Bonta and Hoge (1990) have shown that helping offenders reduce criminogenic needs will reduce the incidence of criminal behavior. From these facts, it is a short step to take to conclude that changes in the LSI-R scores would indicate changes in offending.

There have been three studies done in Canada with small samples of Canadian offenders (Andrews and Robinson, 1984; Motiuk, 1991; Motiuk, Bonta, and Andrews; 1990) and one study in Great Britain with a larger sample of offenders from England, Wales, and Jersey (Raynor, In Press) to support the contention that changes in LSI-R scores between assessments are associated with corresponding changes in offending behavior. The results so far have generally shown that when LSI-R scores change, behavior changes. The nature of the changes in LSI-R sub scale scores and the relationship with recidivism rates was been studied in more depth by Motiuk (1991) and Motiuk, Bonta, and Andrews (1990). Andrews and Bonta (2003) had suggested that more research was needed in this area.

This study will replicate earlier research that examined the relationship between dynamic changes in offender LSI-R scores over time and the subsequent ability of the changed scores to predict arrest rates for offenders. By using larger sample sizes, it is hoped to gain a better understanding of the dynamic nature of the LSI-R.

Offender Assessment

Overview. Bonta (1996) describes four generations of offender risk assessment instruments. First generation (1G) instruments are unstructured and workers use “gut level feelings” based upon past experience and education to make assessments. Second generation (2G) instruments use actuarial methods to measure historical and generally static information such as criminal history. Third generation (3G) instruments measure both static and dynamic factors. The dynamic factors, called criminogenic needs, can be used to guide treatment decisions. Fourth generation (4G) assessment tools measure both static and dynamic factors and also measure factors such as the personality profile or reading ability that can guide the treatment process.

Lowenkamp and Latessa (2005) recommend the use of 3G or 4G assessment tools that use a structured approach. They cite several benefits of this approach, reporting that structured assessments help classify offenders by risk of recidivating and by danger level, they help determine who needs intervention and which interventions are appropriate, they improve the utilization of resources, and they help remove bias by using objective, rather than subjective criteria. There are three types of assessment tools available, screening tools, comprehensive risk/needs instruments, and specialized assessments for specific offender populations. They point out that training in the proper use of offender assessment instruments is essential if the assessments are to work as expected.

Use of Risk/Needs Assessment. The use of structured risk assessment instruments in corrections is an almost universal practice. According to the National Institute of Corrections (2003), 97% of agencies responding to a National survey of corrections departments were using some form of structured risk assessment instrument. Of that total, 45% were using a solution designed by agency staff, 35% were using a system based on the Wisconsin model (Baird, Heinz, & Bemus, 1979), and 22% were using the Level of Service Inventory-Revised (LSI-R; Andrews and Bonta, 1995).

The Wisconsin Model and the LSI-R are 3G assessment tools that measure both static risks and dynamic needs of offenders (Latessa & Allen, 2003). Risks are those historical facts, such as criminal history, that have been shown to be correlated with higher rates of offending. Needs, often referred to as criminogenic needs, are dynamic or changeable facts that are risk factors also, and have been shown to relate to higher rates of offending, but can be used as targets for treatment. An example of a criminogenic need is employment. Job status is statistically related to crime rate, and offenders with no job are more likely to offend. Job status is listed as a need, and not just a risk factor, because offenders who become employed are less likely to commit crimes, which causes their risk level to go down (Andrews & Bonta, 2003).

The overall risk level can be useful in two ways. The first is classification for prediction and safety, and the second is classification for treatment matching (Andrews, Bonta, & Hoge, 1990). In classification for treatment matching, Andrews et. al. showed that high intensity treatment is most effective for high risk offenders and low intensity treatment was most effective for low risk offenders. High intensity treatment for the low risk offender was shown to be counter productive and low intensity treatment for the high-risk offender was shown to be ineffective. Proper classification and treatment was shown to be directly related to treatment outcome.

<u>Criminogenic Needs is Assessment.</u> An analysis of the assessment factors that were most predictive of adult offender recidivism found that dynamic factors were some of the best predictors (Gendreau, Little, and Goggin, 1996). The factors with the highest effect sizes were Antisocial Personality, Criminal Companions, and Criminogenic Needs, which was a composite of antisocial attitudes, antisocial lifestyle, and behavior regarding education and employment.

<u>Targeting Treatment to the Need.</u> Criminogenic needs can be targeted for treatment. Research into factors present in effective treatment programs has shown that programs that focus efforts at helping offenders reduce criminogenic needs are more effective at reducing criminal behavior (Andrews, et. al., 1990; Andrews, Zinger, Hoge, Bonta, Gendreau, & Cullen, 1990).

The Level of Service Inventory-Revised (LSI-R) (Andrews, & Bonta, 1995) is a dynamic risk/needs instrument that was designed to both predict the risk of criminal offending and help guide treatment decisions (Andrews, & Bonta, 2006). The LSI-R is able to accomplish both functions because it uses dynamic, or changeable factors to measure the risk of criminal behavior. These changeable factors, which Andrews and Bonta call criminogenic needs, can become targets for treatment intervention.

The LSI-R was initially developed in Canada as the Level of Supervision Inventory (LSI) by Andrews (1982). The LSI-VI had 58 items and was validated with 341 Canadian offenders. Andrews and Robinson (1984) working with LSI reassessment scores from 57 Canadian offenders were able to show that the reassessment scores were

better predictors of recidivism than the initial LSI scores.

One of the features of the LSI-R that differentiates it from some of the other risk prediction instruments is the use of dynamic, or changeable factors. This capability

one of the more accurate risk

What motivates people to change their behavior?

The past 30 years have been somewhat of a roller coaster ride for the field of corrections. The often cited, but seldom read, Martinson (1974) article lead to a great debate as to whether there is anything that can be done to encourage offenders to stop committing crimes. On the one hand, this is somewhat ironic since, as Moffitt (1993) has pointed out, almost all offenders stop committing crimes, as they get older. On the other hand, this is not an easy question to answer. It is one thing to say that offenders will eventually age out of crime and another to tell for certain when that will occur.

The costs of not coming up with a solution are high.

1

The problem is that we have a difficult time predicting what people are going to do. As

The problems in corrections are many and varied.

The point of this debate though is not whether they stop, but when.

If we could tell which offenders were going to

The purpose of this research project is to provide information for practitioners and fellow researchers that will enable them to better understand and utilize the information they are collecting from offenders. Much of what was done is simple replication of old research with new offenders, however some new material has been added.

, and the earlier studies were all done with smaller size samples of 50 to 60 offenders. The research done by Raynor, using a larger sample size, did not fully replicate the earlier Canadian studies. By replicating the earlier research, this study will add significantly to our current understanding of the dynamic properties of the LSI-R, which will in turn, help our corrections departments make more informed decisions.

Although the LSI-R itself has been thoroughly researched, being cited by Hollin (2002) as having the strongest 'research pedigree' of any risk assessment instrument, the dynamic properties of the LSI-R, which are the effects of score changes between assessments on prediction, are not as well researched. At the time, there were only three studies of the dynamic properties of the LSI-R. All three were done with small samples of between 50 to 60 Canadian offenders.

A more recent report on the dynamic properties of the LSI-R is in the process of being published (Raynor, In Press). The report by Raynor was based on earlier research done on a larger sample (360) of offenders from England, Wales, and Jersey (Raynor, Kynch, Roberts, and Merrington, 2000). The Raynor report did not go as far in depth in its analysis of the dynamic score changes as the Canadian studies.

Even though the effects of changes in assessment scores has not been very thoroughly researched,

Previous research has shown the LSI-R to be a relatively accurate risk assessment instrument. In a meta-analysis of LSI-R studies, Gendreau, Goggin, and Smith (2002) found a weighted effect size of .39 for the LSI-R. This is comparable or superior to other risk assessment instruments reviewed by Andrews, Bonta, and Wormith (2006).

According to Andrews and Bonta (2003), the ten LSI-R sub scales are based upon some of the most well established factors related to criminal behavior. These factors, which they call criminogenic needs, have been shown to have a high correlation with criminal activity. From these two items, scores predict behavior, and working to reduce factors related to the scores improves outcome, it is a short step to conclude that changes in the LSI-R scores would indicate changes in offending, and this is what the previous studies (Andrews, et. al., 1984; Motiuk, 1991; Motiuk, et. al., 1990; Raynor, In Press), have found.

Chapter 3

METHODS

Subjects and Data Sources

Community Corrections Data. The LSI-R records used in this study were obtained from the Stearns County Community Corrections Department with the provision that all identifying characteristics of individual offenders be kept confidential. The records obtained were a subset of data kept in a larger LSI-R database that was maintained by the State of Minnesota for all corrections departments in Minnesota that use the LSI-R. This subset contained all of the records created by Stearns County Community Corrections for offenders placed in community corrections from 2002 through the latter part of 2006. The records were provided in a Microsoft Access Database format and included names, birthdates, LSI-R scores, sub-scale totals, and overall score totals.

The data provided by the County contained 8,860 separate LSI-R assessment records including initial LSI-R results on 5,111 individual offenders, and at least one follow-up assessment on 1,866 offenders. The total number of follow-up LSI-R assessments was 3,749 with the number of assessments per offender varying in number from 1 to 8. Ten of the individual records were excluded from the data set because they were not completed, leaving 5,101 individuals for analysis.

Minnesota BCA Data. The names and birthdates were exported from the LSI-R records to a Microsoft Excel file and sent to the State of Minnesota Bureau of Criminal Apprehension (BCA). The BCA matched 4,918 of the original 5,101names and birthdates with offender arrest and conviction records and returned the resulting data in a text file on a CD. The BCA records were provided with the provision that all records remain confidential with regards to individual characteristics.

Data Variables

The independent variables used in this study were race, gender, offender age at assessment completion, LSI-R item scores, LSI-R sub-scale scores, total LSI-R score, rater id, and LSI-R completion date. The dependent variable used was the arrest date. Since arrests did not get entered into the BCA database unless the offender was subsequently convicted, this variable was recoded as arrests resulting in conviction.

The dependent variable, arrest resulting in conviction, had the advantage of being the strongest and most reliable indicator of recidivism. The disadvantage of this variable was that it is one of the weaker indicators of offending. There are two reasons that arrest is a weak indicator of offending. The first is shared by many measures of offending, and that is the fact that not all offenders get caught. The Minnesota BCA (2005) reports that only 50% of the crimes reported in Stearns County in 2005 resulted in an arrest. The second issue is that not all arrests result in a conviction. The conviction rate is generally about 50% lower than the arrest rate (Langan & Levin, 2002)

Sample Selection. The data provided by the BCA only included arrest data for offenders who had been convicted of a crime. Since there is often a lag time of several months between arrest and conviction, due to the process of brining the case to trial, only assessments that were completed before 2005 were used in the study. This allowed the use of arrest data through 2005 with an additional year, (2006) for 2005 arrests to be entered into the database by the BCA. Sample selection was simply done by using the LSI-R assessment number in the order in which they were done. The assessment numbers were used to identify the samples. For instance, Sample 1 covered all first LSI-R assessments made before 2005, Sample 2 covered all second assessments before 2005, etc.

Data Modification. There were three modifications made to the original data for ease of computation. 1) Null fields in the scoring fields were changed to 0 in order to prevent program errors. This was not seen to be a major issue as they had not been included in the totals anyway. 2) About 20 birthdates that were not coded correctly in the original data were collected from the State BCA database by manually matching offender information using the offender names, location of arrest, LSI-R completion date, and arrest date for identification purposes. 3) The incomplete records that were either unfinished LSI-R records or unmatched BCA records were deleted from the working table after the initial demographic information was collected.

Equipment

A fairly fast Pentium Based computer with 2G of RAM was needed for this study, because the database sets were fairly large. The operating system was Microsoft Windows 2000 Server. For general processing, Microsoft Office Premium 2000, which included Word, Excel, and Access, was used. Office functions included word processing and write-up, data calculation, chart creation, and data manipulation. The LSI-R data records from the County and the arrest and conviction data from the BCA were imported into several Microsoft SQL 2000 Server tables using the import function on SQL. Data manipulations were done using a custom program written in Visual Basic 6. After the data were manipulated in the SQL table, the data from the SQL table were then imported into SPSS 13 Graduate Student Version for Windows. The select records function of SPSS was used to select various subsets of the population for further analysis.

Research Design

This was a retrospective study that performed a secondary analysis of data that had already been collected by others (Bachman, & Schutt, 2003). This study used a time series design (Heppner, Kivlighan, & Wampold, 1999), with the LSI-R used for observation and arrest resulting in conviction used as the measurement criteria between assessments. The time between assessments was variable, although the modal time was approximately six months. The treatment used is unknown, although the corrections department is working to implement best practices (The Carey Group, 2005; Lore & Joplin, 2005) and it is known that, in general, low risk offenders received less intensive treatment than high-risk offenders.

Procedure

Custom Data Fields and Calculations. Additional fields were added to the LSI-R SQL table as needed and populated with data from either the BCA or computations from other fields. For example, a set of fields called “A1”, “A2”, “A3”, etc. were created, each representing the number of arrests in a six month period after the completion of the LSI-R. The “A1” field held the number of arrests in the first six months, the “A2” field held the number of arrests in the second six months, etc. Similarly, a set of race fields was added called “raceb”, “racea”, etc. The “raceb” field was populated with a 1 if the race variable in the BCA data was coded as a “B”. This process of culling data from the BCA tables was repeated for all variables of interest.

Some variables in the SQL table were calculated from other variables. A set of fields called “By1”, “By2”, “By3” was created and a 1 or a 0 was placed in that field if an arrest occurred by the end of the 6-month period represented by the number in the variable. If “A1” was 0 and “A2” was greater than 0, “By1” was coded as a 0 and “By2”, “By3”, “By4”, etc. were all coded to contain a 1. To simplify display of data, a “scorecategory” variable was created. This variable was assigned a number from 1 through 5 with 1 representing scores from 0-11, 2 representing scores from 12-18, 3 representing scores from 19,24, 4 representing scores from 25-31, and 5 representing the scores from 32-54.

Statistical Analysis. Once the data was imported from SQL into SPSS, several operations were performed and various statistics were analyzed. For sample selection, the custom fields generated by Visual Basic were used. For instance, if information on all offenders arrested by one year was desired, the records would be selected where "By2" = 1. Some of the output results were transcribed manually into the results section of this report, while other results were copied into a Microsoft Excel file for further manipulation. Microsoft Excel was used to create most of the graphs in the results, with the only exception being the Scatter Plot in Figure 3, which was created using SPSS.

Calculating the Area Under the Curve. The Area Under the Curve (AUC) values in the results were created using an adapted form of a Microsoft Excel spreadsheet developed by Watkins (2000). The ROC curve was initially developed to help improve signal detectability in radio waves (Peterson, Birdsall, & Fox, 1954), and has since been adapted for the measurement of diagnostic accuracy (Hanley and McNeil, 1982). The magnitude of the AUC score indicates the accuracy of the test. An AUC value of 50 would indicate a chance probability of a correct prediction.

Chapter 4

RESULTS

Population Demographics

The Minnesota BCA was able to match the names and birthdates of 4,918 of the 5,101 offenders with arrest and conviction information contained in the state conviction database. The matched data included arrests made through the end of 2006. Since the database only contained records for those offenders who were both arrested and convicted, all arrests used in this study resulted in conviction. The arrest date was used for study purposes because conviction dates generally follow arrest dates by some time.

To avoid undercounting arrests, only data from assessments done from 2002 through 2004 was used in the study phase. This allowed for the analysis of 12-month recidivism rates with an additional year from the end of the study period for any arrests to turn into convictions and be entered into the BCA database. It is assumed that most arrests come to trial and turn into convictions within one year after the arrest date.

There were 183 individual records (3.6% of offenders) that could not be matched with BCA data; the demographic information for those offenders was compared with the offenders in Sample 1 that were matched. The results are shown below in Table 1.

The missing records tended to be more female (40% vs. 20%), have lower LSI-R Scores, (15.9 vs. 22), and fewer prior arrests (2.4 vs. 4.2). The race of these offenders is unknown since the racial characteristics were collected from the BCA database. The mean ages of 31.3 for the missing records and 32.4 for the matched records were not significantly different (p=.184).

Table 1

Mean Demographic Differences Between Offenders Matched With the BCA Database and Those Not Matched With the BCA Database

Data Set	**N**	**Age**	**Gender % Male**	**Race % White**	**LSI-R**	**# of Previous Arrests**
Matched with BCA	4918	32.4	80%	82%	22	4.2
Not Matched	183	31.3	60%	N/A	15.8	2.4
t test results		ns	$p<.001$		$p<.001$	$p<.001$

Sample Information and Reliability Calculations

The samples used in this study consisted of all of the records that were matched with the BCA data and completed in the years from 2002 to 2004. Sample 1 consisted of the first assessment LSI-R records of all of the offenders with at least one assessment before 2005. Sample 2 consisted of the second assessment LSI-R records that were done on the subset of offenders in Sample 1 that had a second assessment before 2005. Similarly, Sample 3 records included all third assessments and Sample 4 records included all of the fourth assessment records that were completed before 2005. The days between assessments varied, although from the modal days between assessments, it appears that the offenders were normally assessed every six months.

Reliability calculations were performed on the 54 items of the LSI-R assessments to determine the Cronbach's alpha score for each Sample and placed in Table 2. The overall reliability was high. The Cronbach's alpha for individual raters was between .78 and .92 except for one rater who had an alpha of .445 for 10 assessments.

Table 2

Sample Counts and Reliability Data

Sample	N	Cronbach's alpha
Sample 1	3190	.896
Sample 2	1173	.867
Sample 3	616	.878
Sample 4	285	.884

Simplification of Data Display

Previous research had divided the LSI-R scores into categories in order to simplify the display. Initial analysis of the offender LSI-R scores in the population suggested that five roughly equal categories could be obtained by dividing the LSI-R scores into groups using 0-11, 12-18, 19,24, 25-31, and 32-54 as the score ranges. The offenders in Samples 2, 3 and 4 are under represented the lower score categories and over represented in the higher score categories, as can be seen in Figure 6, a side by side comparison of the percentages of offenders for the total population and the four samples for each score category.

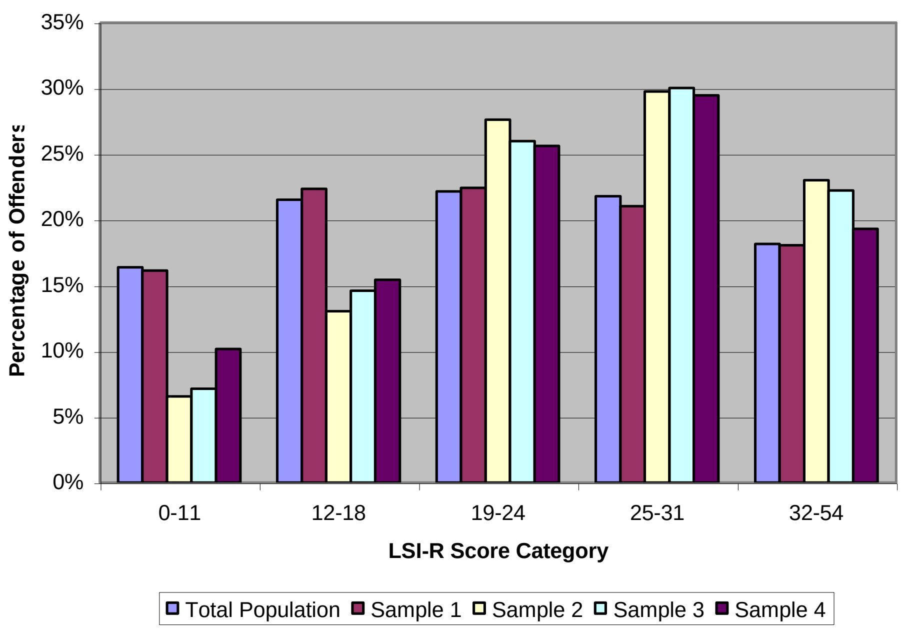

Figure 6

Comparison of Percentages of Offenders per LSI-R Score Category in the General Population and in Samples 1, 2, 3 and 4

Sample 1 Demographic Data

The demographic information for Sample 1 was compiled and is compared to the demographic information of the 4,918 records matched with the BCA in Table 3. A t test was done on the age and the Mann-Whitney U test was used on the categorical values to determine whether the sample statistics for the various demographic items were significantly different from the values of the remaining 1,728 offenders who were assessed in 2005 and 2006. The only significant difference was the gender mix. Sample 1 had 81% male offenders and the excluded offenders were 77% male ($p<.01$).

Table 3

Demographic Information For Offenders Matched with the BCA, Offenders in Sample 1, and Offenders Not Included in Sample 1

	All Records	**Sample 1**	**Records Not Included**	**p**
N	4918	3190	1728	
Mean Age	34.57 S.D.=10.77	32.57 S.D.=10.84	32.09 S.D.=10.92	.145*
Male	80%	81%	77%	.003*
Female	20%	19%	23%	.003*
Race				
White	82% N=4032	82% N=2604	83% N=1428	.380**
Black	13% N=624	13% N=410	12% N=214	.638**
Native	3% N=137	3% N=96	2% N=41	.195**
Asian	2% N=91	2% N=62	2% N=29	.510**
Unknown	1% N=34	1% N=18	1% N=16	.144**

* t-test probability of difference in mean, ** Mann Whitney U probability of difference

<u>Age Distribution For Sample 1</u>

Client ages for the 3,190 offenders in Sample 1 ranged from 17 to 91 for a total range of 74 years. The mean age was 32.57 (StdDev=10.837), the median age was 30 years, and the mode was 23 years. The age frequency distribution for this subset was compiled and graphed in Figure 1. The age distribution was skewed to the left.

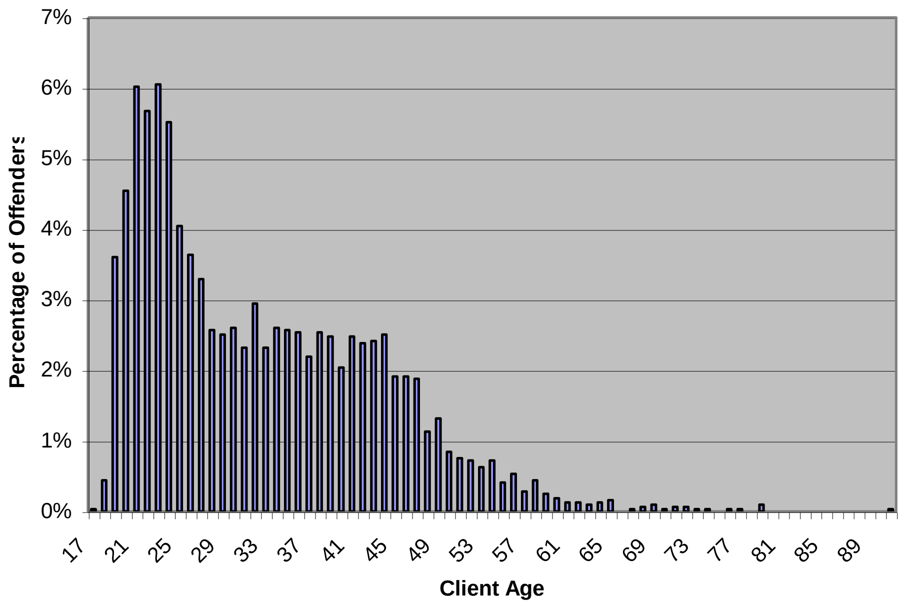

Figure 1

Age Distribution for Sample 1

<u>LSI-R Score Distribution For Sample 1</u>

The LSI-R Scores for this sample of 3,190 offenders had a mean value of 21.92 (StdDev=9.80). The median was 22 and the mode was 23. The minimum score was 1 and the maximum was 48 for a range of 47. A frequency distribution of the scores was compiled and is graphed in Figure 2 below. The shape of the data appears to approximate a normal distribution.

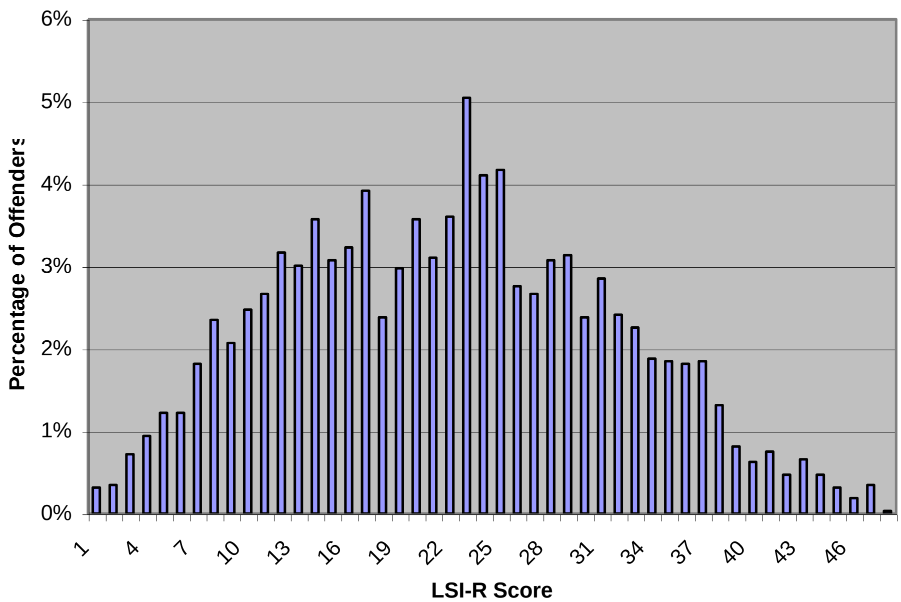

Figure 2

LSI-R Score Distribution For Offenders in Sample 1

<u>LSI-R and Arrest Statistics For Sample 1</u>

The mean, median and modal LSI-R scores for those arrested and those not arrested at six months and one year after assessment are shown in Table 3. A t test done between the mean scores of those arrested vs. those not arrested showed a significant difference at the p<.001 level. The modal value was 23 for all selections.

Table 3

Mean, Median and Modal Values Of The LSI-R Scores by Arrest Status At Six Months and One Year for Sample 1

				LSI-R Scores		
	N	Mean		S.D.	Median	Mode
Six Months						
Arrested	496	25.64		9.32	26	23
Not Arrested	2694	21.24	p<.001*	9.73	21	23
One-Year						
Arrested	778	25.44		9.43	25	23
Not Arrested	2412	20.79	p<.001*	9.64	21	23
Total Offenders	3190	21.92		9.80	22	23

* t-test probability of difference in mean

A correlation analysis was done between LSI-R scores and arrest rates at six-months and one-year following assessment. The Pearson Correlation between LSI-R scores and arrest rates at one year, r=.204 (p<.001) was higher than the at six-month figure r=.163 (p<.001). The Area Under the Curve (AUC) values were calculated for the LSI-R distributions at six-months and one year. The AUC of 63.76% at one-year was found to be somewhat improved over the six-month figure of 63.15%. The Cronbach's alpha for the 54 items in the LSI-R was .896, indicating a high internal consistency.

Sample 2 Demographics

The demographic makeup of Sample 2 was calculated and is shown on Table 4. The offenders with the lowest scores appear to be slightly older, more likely to be male, and White or Asian. The ratio of Black offenders to White offenders shifts dramatically as the score level increases.

Table 4

Demographic Breakdown By LSI-R Score Category For Sample 2

	LSI-R Score Category					
	0-11	**12-18**	**19-24**	**35-31**	**32-54**	**Total**
N	84	242	328	300	219	1173
Age	37.39	37.27	34.00	32.71	32.18	34.25
Male	90%	83%	82%	84%	83%	83%
Female	10%	17%	18%	16%	17%	17%
Race						
White	88%	92%	82%	78%	67%	81%
Black	1%	7%	13%	17%	24%	14%
Native	1%	1%	3%	4%	5%	3%
Asian	10%	0%	2%	1%	4%	2%
Unknown	-	-	-	0% (1)	-	0%

Changes in LSI-R Scores Between Assessments

To determine how dynamic changes in LSI-R scores are related to subsequent arrest rates, the changes in LSI-R scores between assessment 1 and assessment 2 were calculated for Sample 2 and graphed in Figure 3. Changes in LSI-R scores ranged from –21 to 27 with a mean change of –1.43 (Std. Dev. = 6.8), the median change was –1, and the modal score was 0. The mean number of days between assessments was 257 with a significant amount of variation (StdDev= 129). The median number of days between assessments was 222, and the mode was 181, scores ranged from 0 to 996 days.

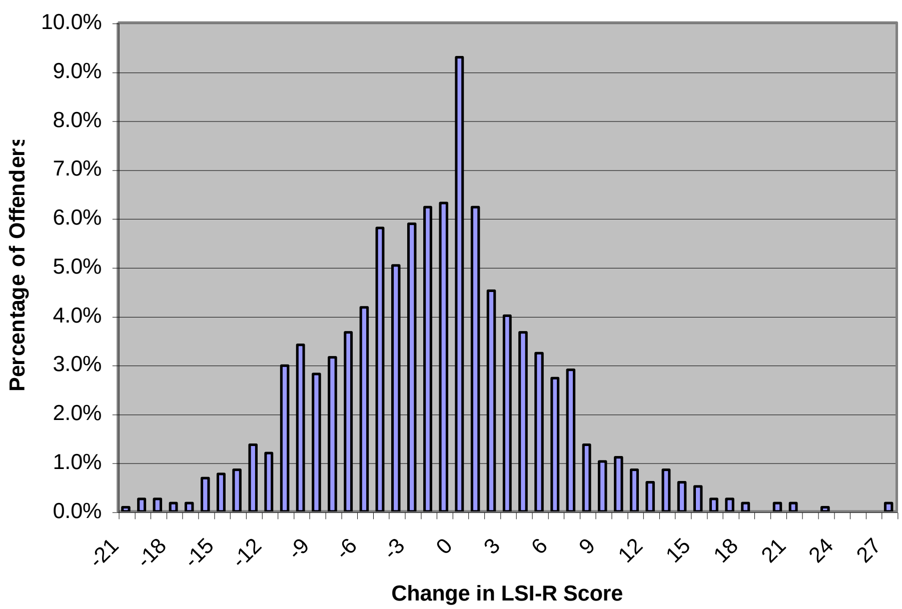

Figure 3

Percentage of Offenders by Change in LSI-R Score For Offenders in Sample 2

<u>Testing For Regression Toward The Mean</u>

Raynor (In Press) had suggested that there might be a regression toward the mean in LSI-R score changes between assessment 1 and assessment 2. To test that hypothesis, the mean score of the 2nd assessments were plotted as a function of the first LSI-R scores and the distribution was plotted in Figure 4.

There was a marked tendency for below average scores on the 1st assessment to be higher on the 2nd assessment and above average scores on the 1st assessment to be slightly lower on the 2nd assessment. This suggests that there is a regression towards the mean in the LSI-R scores between 1st and 2nd assessments.

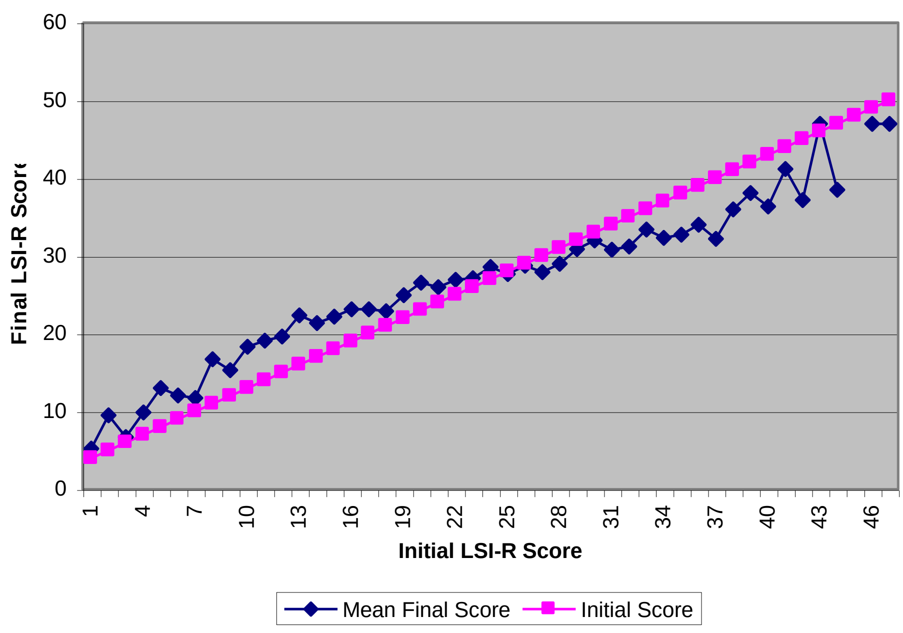

Figure 4

Mean LSI-R Score At Second Assessment Plotted For Each Initial Score
For Offenders In Sample 2 – N=1,173

Plot of Changes in Scores Between Assessments

To determine whether there was a recognizable pattern to the score changes, a scatter plot of the initial scores (FirstScore) vs. the second scores (totalScore) was created and shown in Figure 5. While there does seem to be a general linear trend, many scores appear to be widely dispersed from their initial level.

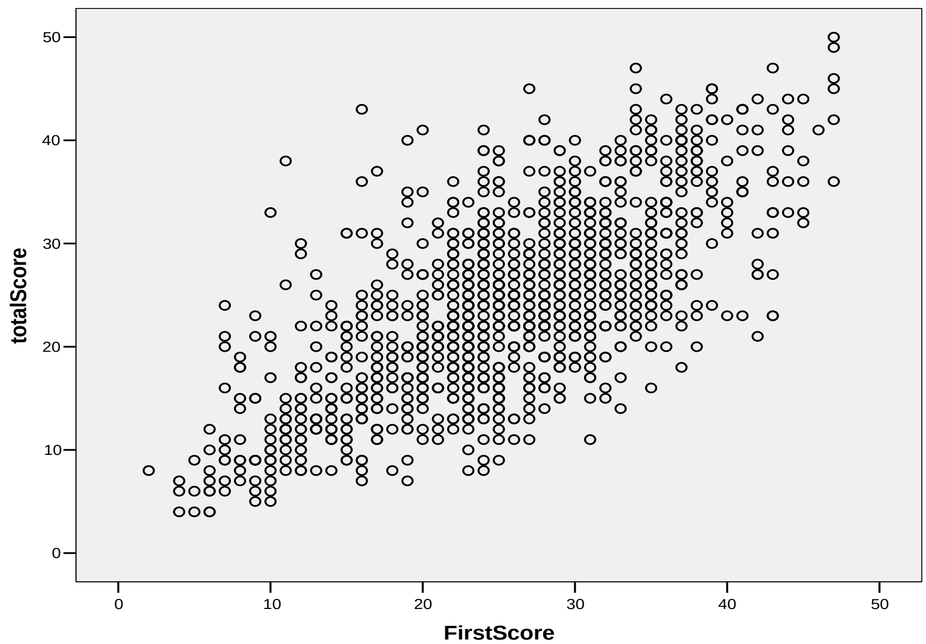

Figure 5

Scatter Plot of LSI-R Scores (totalscore) at Second Assessment by LSI-R Scores (FirstScore) at Initial Assessment For Offenders In Sample 2 – N=1,173

Replicating the 1984 Study by Andrews and Robinson

In the study done by Andrews and Robinson (1984), the score distributions for the first and second assessments were broken down by risk level and compared to see which scores were the more accurate predictors of outcome. The numbers and percentages of arrests at one year were plotted for both the first and second assessments at each risk level (1-11, 12-18, 19,24, 25-31, and 32-54) for Sample 2 and placed in Table 5. There were overall improvements in scoring at the very lowest level (5% vs. 10%), at the 25-31 level (31% vs. 24%), and the highest level (41% vs. 39%). This suggests that the second LSI-R was a better predictor of risk on the second LSI-R due to a more accurate measurement of risk level.

Table 5

Rearrest Rates For Sample 2 After Follow-up For Initial LSI-R and Follow-up LSI-R

Intake Risk Level	**Retake Risk Level**					
	0-11	**12-18**	**19-24**	**25-31**	**32-54**	**Overall**
0-11	4%(2/47)	16% (3/19)	38% (3/8)	0% (0/1)	0% (0/2)	10% (8/77)
12-18	9%(2/23)	8% (6/76)	28%(10/36)	7% (1/15)	100% (3/3)	14% (22/153)
19-24	0% (0/9)	17%(15/90)	17%(21/127)	25% (19/76)	45% (10/22)	20% (65/324)
25-31	0% (0/5)	16% (8/51)	16%(18/113)	30%(37/125)	40% (22/55)	24% (85/349)
32-54	-	17% (1/6)	30% (13/44)	42% (35/83)	40%(55/137)	39%(104/270)
Overall	5%(4/84)	14%(33/242)	20%(65/328)	31%(92/300)	41% (219)	24%(284/1173)

Correlation Analysis of the Dynamic Changes in LSI-R Score

In order to get a better understanding of the differences in predictive value for the two LSI-R assessments, the overall arrest rates were calculated by score level for both the initial and follow-up LSI-R assessments and placed in Table 6 below. The correlation rates and probabilities of correlation between arrest and LSI-R scores were also calculated. The correlation rate was higher for the second assessment than the first (.257 vs. .193).

The AUC values were calculated for the first and second assessment. The AUC values (67.60 vs. 63.62) indicate that the second assessment was better at predicting arrest than the first assessment.

Table 6

Arrest Rate by LSI-R Score Level, AUC, and Correlation Rates Between LSI-R Score and Arrest Rate by OneYear After Second Assessment For Offenders in Sample 2

	LSI-R Score Level								
LSI-R #	**0-11 %**	**12-28 %**	**19-24 %**	**25-31 %**	**32-54 %**	**Total %**	**AUC**	**r**	**p**
LSI#1	10	14	20	24	39	24	63.62	.193	.000
LSI#2	5	14	20	31	41	24	67.60	.257	.000

The Effect of Increases and Decreases in LSI-R Scores

Raynor (In Press) had done a slightly different analysis of change in LSI-R scores. He had broken the score categories into above average and below average to eliminate the possibility of regression toward the mean. He then looked at whether the scores were increasing or decreasing. He also contrasted the arrest rates of all of the increasing and decreasing scores. This method was used to calculate the figures in Table 7 below.

As in the Raynor study, increasing scores at each initial level were associated with higher levels of arrest when compared with decreasing scores. Note that the mean final LSI-R scores also increase and decrease for each category. (The scores that do not change are included for completeness.) When the combined totals for all increasing and all decreasing scores were calculated, the arrest rate was higher for the increasing scores.

Table 7

Arrest Rates and Mean Values For the First and Second LSI-R For Increasing, Decreasing, Static, and Total Sample For Sample 2

LSI-R Change Category	**N**	**1 Yr Arrest Rate**	**Mean LSI-R #1**	**Mean LSI-R #2**
LSI-R #1 <=25 Increasing	277	23%	18.24	24.20
LSI-R #1 <=25 Decreasing	270	13% p<.01*	20.35 p<.001**	15.55 p<.001**
LSI-R #1 > 25 Increasing	139	40%	32.10	36.65
LSI-R #1 > 25 Decreasing	378	29% p<.02*	32.52 ns	25.46 p<.001**
LSI-R #1 Same as LSI-R #2	109	19%	21.83	21.83
Total	1173	24%	25.30	23.87
All Increasing	416	29%	22.87	28.36
All Decreasing	648	22% p<.02*	27.45 p<.001**	21.33 p<.001**

* Mann Whitney U probability of difference, ** t-test probability of difference in mean

<u>Breakdown of Arrest Rates by Change Level For Sample 2</u>

In order to look at how levels of change affected the dynamic validity of the LSI-R scores, the scores were broken down by change level. The mean scores for LSI-R #1 and LSI-R #2, probability of difference in mean scores, arrest rates, correlation rates with arrest and probability of correlation were calculated for each change level and placed in Table 8. The arrest rate for the –21 to –10 change level is slightly higher than for the –9 to –4 change level even though the mean LSI-R score is lower. The correlation rate is also lower for the scores that dropped the most. It is possible that decreasing scores are not as accurate as increasing scores for this sample. The correlation rates improved the most for the scores that changed the most.

Table 8

LSI-R Means, Arrest Rates, and Correlation Rates Between LSI-R Scores and Arrest After Second Assessment For Different Change Levels in Sample 2

	Change in LSI-R Score From LSI-R #1 to LSI-R #2					
	-21 to -10	**-9 to -4**	**-3 to 2**	**3 to 9**	**10 to 27**	**All**
N	143	289	451	222	68	1173
LSI-R #1 Mean	31.51	27.76	23.56	23.60	19.32	25.30
LSI-R #2 Mean	19.12	21.57	23.01	28.86	33.09	23.87
p*	.000	.000	.847	.000	.000	.000
Arrest Rate	22%	21%	21%	32%	37%	24%
r (LSI-R #1)	.155	.262	.236	.231	.261	.193
p	.065	.000	.000	.001	.031	.000
r (LSI-R #2)	.180	.254	.226	.239	.307	.257
p	.032	.000	.000	.000	.011	.000

* t-test probability of difference in mean

Dynamic Changes in LSI-R Sub-Level

The LSI-R sub level scores and the mean differences between the first and second scores were calculated for the first and second assessments and placed in Table 8 below. Not all mean scores changed significantly. The direction of change was also variable. This suggests that changes are not always all in one direction.

Table 8

Values of Mean Sub-Scale Scores for LSI-R #1 and LSI-R #2, Mean Difference, and t-test Probability of Difference in Means For Sample 2

	LSI-R #1		**LSI-R #2**			
	Mean	**SD**	**Mean**	**SD**	**d**	**p**
LSI-R Sub-Scale						
Criminal History	5.06	2.2	5.46	1/9	.40	.000
Education/Employment	4.44	2.9	3.92	2.8	-.52	.000
Financial	1.16	0.7	1.08	0.8	.08	.008
Family/Marital	1.90	1.2	1.84	1.2	-.06	.219
Accommodation	0.83	1.0	0.76	0.9	-.07	.085
Leisure/Recreation	1.32	0.8	1.14	0.9	-.18	.000
Companions	2.06	1.2	2.14	1.1	0.9	.065
Alcohol/Drugs	4.55	2.7	3.90	2.6	-.65	.000
Emotional/Personal	2.45	1.3	2.36	1.2	-.09	.063
Attitude/Orientation	1.53	1.5	1.27	1.4	-.26	.000
Total LSI-R	25.30	8.4	23.87	8.7	-1.43	.000

Notes:_SD = Standard Deviation, d = Change in Score, p = Probability of a difference in absolute value for t-test

Correlation Rates For LSI-R Sub-Scales

The correlation rates between LSI-R sub level scores and arrest rate at one year after assessment were calculated for Sample 2 and listed in Table 9. Second assessment correlation rates were higher than first assessment rates for all scales but the Accommodations sub-scale.

Table 9

Correlation Rates Between First and Second LSI-R Sub Scale Scores and Arrest Rate at One Year After Second Assessment

	LSI-R #1		**LSI-R #2**	
	r	**p**	**r**	**p**
LSI-R Sub-Scale				
Criminal History	.168	.000	.223	.000
Education/Employment	.145	.000	.175	.000
Financial	.100	.001	.130	.000
Family/Marital	.049	.092	.092	.002
Accommodation	.077	.009	.075	.010
Leisure/Recreation	.096	.001	.097	.001
Companions	.084	.004	.138	.001
Alcohol/Drugs	.090	.002	.175	.000
Emotional/Personal	.009	.761	.069	.018
Attitude/Orientation	.149	.000	.178	.000
Total LSI-R	.193	.000	.257	.000

Notes: r = Pearson's Correlation, p=Probability of correlation.

It is interesting that the mean score for the Emotional/Personal sub scale did not change significantly but the correlation with arrest changed from insignificant to significant. This suggests that the mean score change is measuring a composite change which does not reflect individual changes.

Correlation Analysis of Changes in Sub Level Scores

In order to determine which changes in sub level were most correlated with arrest rates, the change in sub level scores was calculated and the correlation between change in score and arrest rate was determined. The results are shown in Table 10. The most significant changes that were related to arrest appear to be in the Criminal History, Alcohol and Drug, and Emotional and Personal sub scales. People with higher arrest rates tend to have larger positive changes in LSI-R sub level scores in these areas.

Table 10

Mean Differences in LSI-R Sub Scale Score Changes Between Offenders Arrested by One Year After LSI-R #2 and Offenders Not Arrested, t-test probability of Difference in Mean Change, and Correlation Rates Between Change and Arrest For Sample 2

	Mean Score Change		**Mean Difference**		**Correlation**	
	Not Arrested	**Arrested**	**d**	**p**	**r**	**p**
LSI-R Sub-Scale						
Criminal History	.35	.54	.19	.008	.083	.004
Education/Employment	-.56	-.39	.17	.347	.029	.323
Financial	-.10	-.04	.06	.177	.040	.175
Family/Marital	-.09	.03	.12	.088	.054	.064
Accommodation	-.07	-.07	-.01	.911	-.003	.907
Leisure/Recreation	-.18	-.17	.02	.757	.009	.761
Companions	.06	.18	.12	.157	.044	.130
Alcohol/Drugs	-.77	-.28	.50	.004	.087	.003
Emotional/Personal	-.13	.03	.16	.039	.061	.035
Attitude/Orientation	-.28	-.20	.08	.453	.023	.434
Total Score	-1.77	-.37	1.40	.005	.088	.002

Dynamic Changes Between Assessment 2 and Assessment 3

Of the 1173 offenders with a second assessment in Sample 2, a group of 616 offenders, which will hereafter be referred to as Sample 3, had a third assessment before 2005. Changes in LSI-R scores for this group of offenders were calculated for the change between LSI-R #1 and LSI-R #3, and between LSI-R #2 and LSI-R #3, and the percentage at each change level was graphed in Figure 7 below. The distribution for changes in LSI-R scores between assessments 2 and 3 is peaked at the score change = 0. The mean change between LSI-R #2 and LSI-R #3 was -0.77 (Std.Dev.=5.3). The mean days between assessments two and three was 215 days (StdDev. = 90), with a mode of 179. The Cronbach's alpha for the 54 items on the LSI-R for Sample 3 was .885.

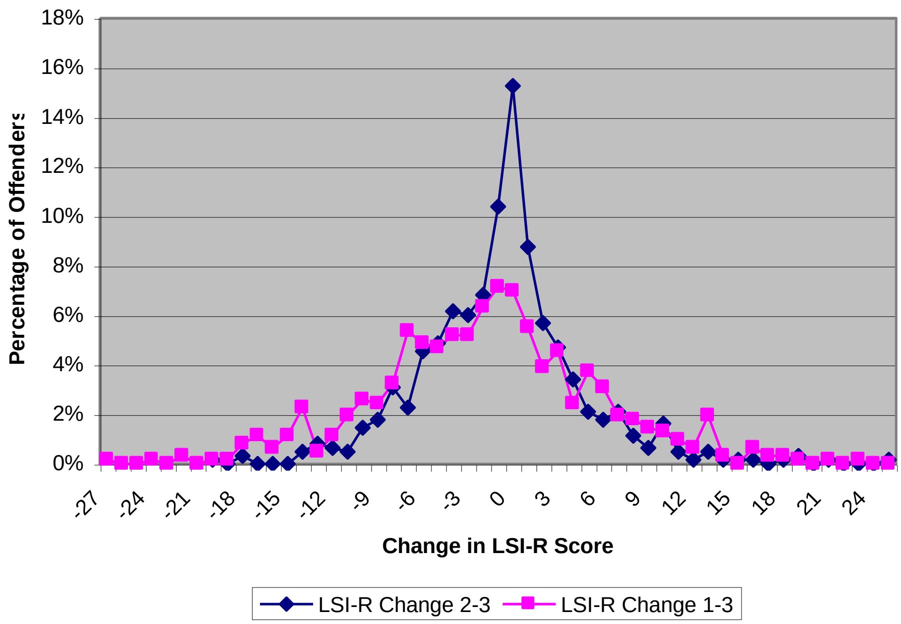

Figure 7

Percentage of Offenders by Change in LSI-R Score for Sample 3

<u>Continued Regression Toward the Mean for Sample 3</u>

The mean scores for assessment 3 were plotted for each score at assessment 2 and plotted in Figure 7 below. Visual observation indicates that either the mean scores remained the same or there was a continued regression toward the mean, with above average scores getting lower and below average scores getting higher. The regression for above average scores seems to be larger than for below average scores.

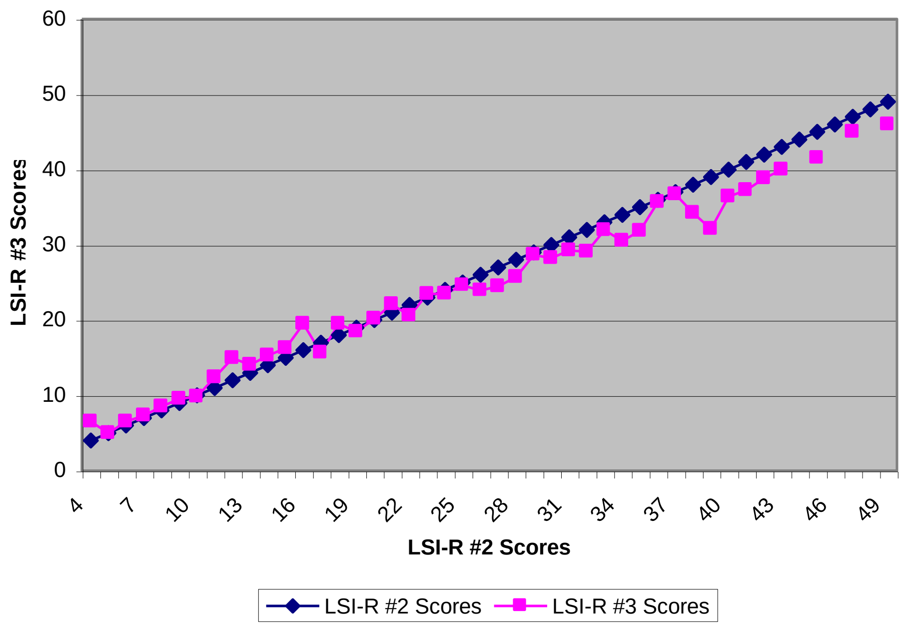

Figure 7

Mean LSI-R Score at Third Assessment Plotted for Each Second Assessment Score
For Offenders In Sample 3 – N=616

Arrest Rates for LSI-R #2 vs. LSI-R #3

The numbers and percentages of arrests at one year were plotted for both the second and third assessments at each risk level for Sample 2 and placed in Table 11.

The overall improvements in scoring were modest. The overall results for LSI-R #3 and LSI-R #2 are very similar with slight improvements for LSI-R #3 over LSI-R #2. An examination of the detail shows that low scorers in LSI-R #2 that were assessed higher in LSI-R #3 tended to be placed in accurate categories but high scorers that were assessed at a lower risk level still tended to be arrested at high rates. This seems to indicate that, for this assessment, increases in LSI-R score are predictive of increased offending, but decreases in LSI-R score don't predict decreases in offending.

Table 11

Arrest Rates by One Year After Assessment 3 For LSI-R #2 and LSI-R #3

LSI-R #2 Risk Level	**LSI-R #3 Risk Level**					
	0-11	**12-18**	**19-24**	**25-31**	**32-54**	**Overall**
0-11	3% (1/34)	13% (1/8)	-	-	-	5% (2/42)
12-18	0% (0/12)	6% (4/72)	10% (2/21)	20% (1/5)	33% (1/3)	7% (8/113)
19-24	0% (0/2)	17% (8/47)	19% (16/85)	32% (10/31)	50% (4/8)	22% (38/173)
25-31	-	7% (1/15)	28% (15/53)	23% (19/81)	38% (10/26)	26% (45/175)
32-54	-	50% (1/2)	33% (1/3)	32% (10/31)	31% (24/77)	32% (36/113)
Overall	2% (1/48)	10% (15/144)	21% (34/162)	27% (40/148)	34% (39/114)	21% (129/616)

Correlation Analysis for Sample 3

To determine whether the score distribution for the third assessment was more predictive than the first or second assessment distributions, an analysis, including calculating the arrest rates by LSI-R Score level, AUC calculation, and correlation analysis, was done to compare the arrest rates for the score distributions of all three LSI-R assessments. The results are shown in Table 11 below.

The scores for the third assessment appear to be significantly more predictive of arrest than the scores from the first assessment, but there is only a slight difference in overall AUC or correlation rates between the scores from the second and third assessments. From the detail shown earlier, the results appear to be affected by the problem with decreasing scores not being associated with decreasing risk. Also, if large percentages of scores did not change much, the results for LSI-R #2 and LSI-R #3 would be expected to be similar.

Table 11

One Year Arrest Rates By LSI-R Score Category, AUC, and Correlation Between LSI-R and Arrest After Third Assessment - N=616

	LSI-R Score Category								
LSI-R #	**0-11 %**	**12-28 %**	**19-24 %**	**25-31 %**	**32-54 %**	**Total %**	**AUC**	**r**	**p**
All Offenders N=616									
LSI#1	7	11	21	25	26	21	60.72	.156	.000
LSI#2	5	7	22	26	32	21	66.49	.225	.000
LSI#3	2	10	21	27	34	21	66.73	.227	.000

<u>Changes in Score Level For Sample 3</u>

To determine how changes in score level affected prediction of arrest, the method used by Raynor (In Press) was applied to the score changes between the second and third assessments. The results are shown in Table 12.

The numbers are as expected. Increasing scores are accompanied by higher arrest rates than are decreasing scores. This suggests that, even though the overall score distribution is not better at predicting arrest, individual scores still predict better at the third assessment than the second.

Table 12

Arrest Rates and Mean Values For the Second and Third LSI-R For Increasing, Decreasing, Static, and Total Sample For Sample 3

LSI-R Change Category	**N**	**1 Yr Arrest Rate**		**Mean LSI-R #2**		**Mean LSI-R #3**	
LSI-R #2 <=25 Increasing	136	20%		18.91		23.71	
LSI-R #2 <=25 Decreasing	154	16%	ns*	19.62	ns**	15.97	p<.001**
LSI-R #2 > 25 Increasing	76	37%		32.13		35.91	
LSI-R #2 > 25 Decreasing	156	24%	p<.05*	32.11	ns**	26.66	p<.001**
LSI-R #2 Same as LSI-R #3	94	13%		19.39		19.39	
Total	616	21%		24.14		23.37	
All Increasing	212	26%		23.65		28.08	
All Decreasing	310	20%	ns*	25.91	p<.01**	21.35	p<.001**

* Mann Whitney U probability of difference, ** t-test probability of difference

Breakdown of Arrest Rates by Change Level For Sample 3

The mean scores for LSI-R #1, LSI-R #2, and LSI-R #3, arrest rates, correlation rates with arrest and probability of correlation were calculated for each change level and placed in Table 8. The arrest rate for the –21 to –10 change level is slightly higher than for the –9 to –4 change level and much higher than expected. The arrest rates for the –21 to -10 and –9 to -4 levels are closer to what would be expected with the mean scores for LSI-R #2 at those levels. The correlation rate for the –21 to –10 change level is higher for LSI-R #2 than for LSI-R #3.

Table 8

LSI-R Means, Arrest Rates, and Correlation Rates Between LSI-R Scores and Arrest After Second Assessment For Different Change Levels in Sample 2

	Change in LSI-R Score From LSI-R #2 to LSI-R #3					
	-21 to -10	**-9 to -4**	**-3 to 2**	**3 to 9**	**10 to 27**	**All**
N	27	140	326	98	25	616
LSI-R #1 Mean	27.96	26.90	23.92	25.16	25.92	25.05
LSI-R #2 Mean	32.00	27.07	22.60	23.52	21.64	24.14
LSI-R #3 Mean	19.63	21.22	22.19	28.47	34.88	23.37
Arrest Rate	26%	24%	18%	24%	28%	21%
r (LSI-R #1)	-.016	-.015	.219	.209	.044	.156
p	.935	.861	.000	.039	.834	.000
r (LSI-R #2)	.379	.081	.262	.315	-.158	.225
p	.051	.343	.000	.002	.452	.000
r (LSI-R #3)	.335	.071	.279	.348	-.255	.227
p	.087	.405	.000	.000	.279	.000

<u>Dynamic Changes Between Assessment 3 and Assessment 4</u>

Of the 616 offenders with a third assessment in Sample 3, a group of 285 offenders, which will hereafter be referred to as Sample 4, had a fourth assessment before 2005. Changes in LSI-R scores for this group of offenders were calculated for the change between LSI-R #1 and LSI-R #4, between LSI-R #2 and LSI-R #4, and between LSI-R #3 and LSI-R #4, and the percentage at each change level was graphed in Figure 8 below. The distribution for changes in LSI-R scores between assessments 3 and 4 is peaked at the score change = 0. The mean change between LSI-R #3 and LSI-R #4 was -0.65 (Std.Dev.=4.9), ranging from –17 to 23. The mean days between assessments three and four was 192 days (StdDev. = 77), with a mode of 187. The Cronbach's alpha for the 54 items on the LSI-R for Sample 4 was .889.

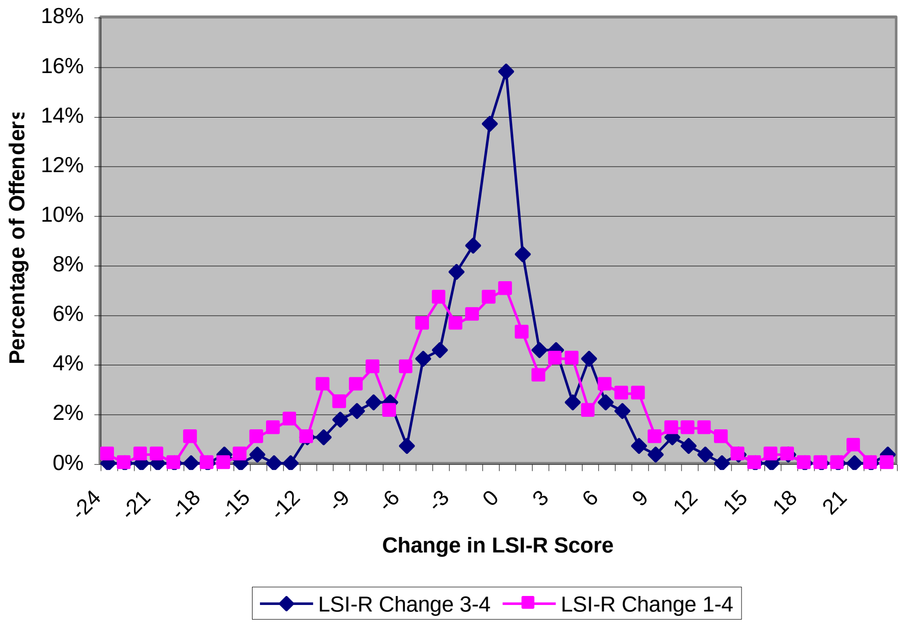

Figure 8

Percentage of Offenders by Change in LSI-R Score for Sample 4

Continued Regression Toward the Mean For Sample 4

The mean scores for assessment 4 were plotted for each score at assessment 3 and plotted in Figure 9 below. Visual observation indicates that either the mean scores remained the same or there was a continued regression toward the mean, with above average scores generally getting lower and below average scores getting higher. The regression for above average scores seems to be larger than for below average scores.

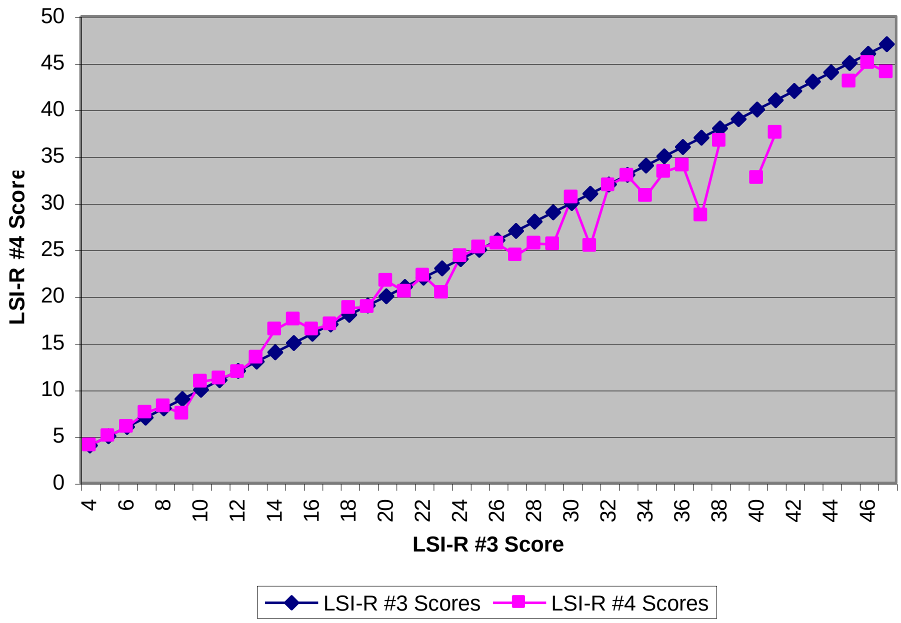

Figure 9

Mean LSI-R Score at Third Assessment Plotted for Each Second Assessment Score For Offenders In Sample 4 – N=285

Arrest Rates for LSI-R #3 vs. LSI-R #4

The numbers and percentages of arrests at one year were plotted for both the third and fourth assessments at each risk level for Sample 4 and placed in Table 13.

It is difficult to tell if there are overall improvements in scoring. The overall results for LSI-R #4 and LSI-R #3 are very similar with a slight improvement for LSI-R #4 over LSI-R #3 in the 24-31 score level and a slight decline in accuracy for LSI-R #4 over LSI-R #3 in the 32-54 score level. An examination of the detail shows that there appear to be problems with both increasing and decreasing scores that do not seem to match the expected arrest rate for that level. The arrest rates should increase from left to right for all risk levels and that is not what happened.

Table 13

Arrest Rates by One Year After Assessment 4 For LSI-R #3 and LSI-R #4

LSI-R #3 Risk Level	**LSI-R #4 Risk Level**					
	0-11	**12-18**	**19-24**	**25-31**	**32-54**	**Overall**
0-11	4% (1/24)	0% (0/3)	-	-	-	4% (1/27)
12-18	0% (0/6)	15% (6/41)	22% (2/9)	67% (2/3)	0% (0/1)	17% (10/60)
19-24	0% (0/1)	25% (4/16)	17% (6/35)	50% (9/31)	0% (0/3)	26% (19/73)
25-31	-	14% (1/7)	24% (5/21)	31% (13/42)	25% (2/8)	27% (21/78)
32-54	-	-	100% (1/1)	39% (7/18)	54% (15/28)	49% (23/47)
Overall	3% (1/31)	16% (11/67)	21% (14/66)	38% (31/81)	43% (17/40)	26% (74/285)

<u>Dynamic Changes Between Assessment 3 and Assessment 4</u>

Of the 616 offenders with a third assessment in Sample 2, 285 offenders, which shall hereafter be called Sample 4, had a fourth assessment before 2005. To determine whether the fourth assessment was more predictive than the first, second or third assessments, an analysis, including calculating the arrest rates by LSI-R Score level, and correlation analysis, was done to compare the arrest rates for the score distributions of all four LSI-R assessments. The results are shown in Table 13 below.

The scores for the fourth assessment appear to be significantly more predictive of arrest than the scores from the first assessment, and slightly more predictive of arrest than scores from the second assessment, but slightly less predictive than the scores from the third assessment. This suggests that the total distribution is less predictive. Individual scores did change between assessments with a mean drop of .65 between LSI-R #3 and LSI-R #4 (StdDev. = 4.9). The changes ranged from 17 to 23.

Table 13

One Year Arrest Rates By LSI-R Score Category, and Correlation Between LSI-R and Arrest After Fourth Assessment - N=285

	LSI-R Score Category								
LSI-R #	0-11 %	12-28 %	19-24 %	25-31 %	32-54 %	Total %	AUC	r	p
All Offenders N=285									
LSI#1	14	11	32	29	33	26	60.82	.159	.007
LSI#2	4	16	22	30	47	26	67.09	.264	.000
LSI#3	4	17	26	27	49	26	66.90	.274	.000
LSI#4	3	16	21	38	43	26	67.15	.267	.000

Changes in Score Level For Sample 4

To determine how changes in score level affected prediction of arrest, the method used by Raynor (In Press) was applied to the score changes between the third and fourth assessments. The results are shown in Table 14.

The numbers are not as expected. Increasing scores are accompanied by increasing arrest rates when compared to decreasing scores for the lower half of the distribution but the opposite is true for the high end where increasing scores actually have a lower arrest rate than the decreasing scores. The overall difference between increasing scores and decreasing scores is only 2%, which is much less than that seen for the changes from LSI-R #1 to LSI-R #2, and the changes from LSI-R #2 to LSI-R #3.

Table 14

Arrest Rates and Mean Values For the First, Second, Third and Fourth LSI-R For Increasing, Decreasing, Static, and Total Sample For Sample 4

LSI-R Change Category	**N**	**1 Yr Arrest Rate**	**Mean LSI-R #3**	**Mean LSI-R #4**
LSI-R #3 <=25 Increasing	62	27%	18.63	23.10
LSI-R #3 <=25 Decreasing	78	18% ns*	18.73 ns**	15.72 p<.001**
LSI-R #3 > 25 Increasing	32	31%	31.25	35.22
LSI-R #3 > 25 Decreasing	68	37% ns*	31.59 ns**	26.38 p<.001**
LSI-R #3 Same as LSI-R #4	45	18%	18.07	18.07
Total	285	26%	23.08	22.43
All Increasing	94	29%	22.93	27.22
All Decreasing	146	27% ns*	24.72 ns**	20.68 p<.001**

* Mann Whitney U probability of difference, ** t-test probability of difference

Breakdown of Arrest Rates by Change Level For Sample 4

The mean scores for LSI-R #1, LSI-R #2, LSI-R #3, and LSI-R #4, arrest rates, correlation rates with arrest and probability of correlation between arrest and LSI-R were calculated for each change level and placed in Table 9. The correlation rates for assessment 3 were better across all change levels than for assessment 4. They did not all reach significance, but that may have been due in part to the small numbers of offenders at some levels. The –9 to –4 level had a high arrest rate in comparison to the other levels.

Table 9

LSI-R Means, Arrest Rates, and Correlation Rates Between LSI-R Scores and Arrest After Second Assessment For Different Change Levels in Sample 2

	Change in LSI-R Score From LSI-R #2 to LSI-R #3					
	-21 to -10	**-9 to -4**	**-3 to 2**	**3 to 9**	**10 to 27**	**All**
N	13	47	168	48	9	285
LSI-R #1 Mean	27.85	24.11	22.61	27.02	26.33	23.95
LSI-R #2 Mean	25.15	25.43	21.97	25.85	25.89	23.46
LSI-R #3 Mean	29.69	27.06	21.34	23.77	21.44	23.08
LSI-R #4 Mean	18.08	21.04	20.71	28.69	34.56	22.43
Arrest Rate	.23	.32	.22	.31	.44	.26
r (LSI-R #1)	-.407	.099	.181	.286	.081	.159
p	.167	.508	.019	.049	.835	.007
r (LSI-R #2)	.049	.174	.280	.271	.409	.264
p	.875	.241	.000	.062	.274	.000
r (LSI-R #3)	.472	.256	.294	.209	.073	.274
p	.103	.082	.000	.153	.274	.000
r (LSI-R #4)	.469	.246	.289	.172	-.176	.267
p	.106	.096	.000	.242	.651	.000

<u>The Effect of Days Between Assessments on Second Assessment Correlation Rates</u>

Bonta (Personal Communication, 2/21/2007) had suggested that there might be a lower correlation between changed scores and arrest for assessments done before six months had elapsed. In order to test this, the score changes and correlation between the arrest rate and the LSI-R scores was calculated for assessments made at differing time periods after the initial assessment. The results are shown in Table 15 below. The correlation rates are higher for the second assessment than the first for time periods from 90 to 365 days.

Table 15

Arrest Rate, Mean Values For First and Second LSI-R and Correlation Rate Between LSI-R Score and the Arrest Rate at One Year For Various Time Periods After Second Assessment

Days Between Assessments	**N**	**1st Yr Arrest Rate**	**LSI-R #1**	**LSI-R #2**
0 – 90 (M=53, StdDev=29)	33	27%		
Mean			24.39	22.73
r			.298	.518
p			.092	.002
91 – 180 (M=155, StdDev=25)	233	24%		
Mean			24.76	23.16
r			.193	.237
p			.000	.000
181 – 365 (M=245, StdDev=50)	757	24%		
Mean			25.59	23.72
r			.200	.274
p			.000	.000
More than 365 (M=526, StdDev=137)	150	27%		
Mean			24.88	25.98
r			.153	.145
p			.062	.077
Total (M=257 Days, StdDev=129)	1173	24%		
Mean			25.30	23.87
r			.193	.257
p			.000	.000

<u>The Effect of Days Between Assessments on Fourth Assessment Correlation Rates</u>

The score changes and correlation between the arrest rate and the LSI-R scores was calculated for fourth assessments made at differing time periods after the third assessment. The results are shown in Table 17 below.

Table 17

Arrest Rate, Mean Values For First and Second LSI-R and Correlation Rate Between LSI-R Score and the Arrest Rate at One Year For Various Time Periods After Second Assessment

Days Between Assessments	**N**	**1st Yr Arrest Rate**	**LSI-R #3**	**LSI-R #4**
0 – 90 (M=55, StdDev=25)	14	14%		
Mean			24.14	23.07
r			.245	.434
p			.398	.121
91 – 180 (M=149, StdDev=25)	83	28%		
Mean			23.54	21.94
r			.134	.095
p			.227	.392
181 – 365 (M=223, StdDev=44)	172	26%		
Mean			23.42	22.55
r			.342	.336
p			.000	.000
More than 365 (M=403, StdDev=51)	12	25%		
Mean			22.88	23.06
r			.062	.223
p			.819	.406
Total (M=204 Days, StdDev=77)	285	26%		
Mean			23.46	22.43
r			.264	.267
p			.000	.000

The Effect of Days Between Assessments on Third Assessment Correlation Rates

The score changes and correlation between the arrest rate and the LSI-R scores was calculated for third assessments made at differing time periods after the second assessment. The results are shown in Table 16 below. The correlation rates for LSI-R #2 were higher than the correlation rates for LSI-R #3 for the 0-90 and 180-365 day periods. The correlation rate was higher for LSI-R #3 for the 90-180 day period.

Table 16

Arrest Rate, Mean Values For First and Second LSI-R and Correlation Rate Between LSI-R Score and the Arrest Rate at One Year For Various Time Periods After Second Assessment

Days Between Assessments	N	1st Yr Arrest Rate	LSI-R #2	LSI-R #3
0 – 90 (M=62.7, StdDev=21)	27	30%		
Mean			21.30	18.22
r			.230	.159
p			.249	.427
91 – 180 (M=156, StdDev=24)	179	13%		
Mean			22.71	21.35
r			.234	.293
p			.002	.000
181 – 365 (M=230, StdDev=45)	377	24%		
Mean			24.75	24.04
r			.240	.226
p			.000	.000
More than 365 (M=489, StdDev=95)	33	24%		
Mean			27.24	30.88
r			-.013	-.015
p			.941	.934
Total (M=215 Days, StdDev=90)	616	21%		
Mean			24.14	23.37
r			.225	.227
p			.000	.000

Chapter 5

DISCUSSION

This study attempted to replicate earlier work done to examine how the predictive accuracy of the LSI-R changed from assessment to assessment. The results were mixed. While the second and subsequent LSI-R assessment scores all appear to predict recidivism better than scores from the first assessment, the third assessment did not appear to be significantly more predictive than the second assessment, and the fourth assessment did not appear to be significantly more predictive than the than the second or third assessments.

The results should be interpreted with caution since the follow up assessment records do not include data for all of the original offenders. The second and subsequent assessments were only done on offenders who were selected for closer scrutiny by the caseworkers. The offenders with a fourth assessment were a selection of a selection of a selection of offenders who were presumably problematic and deserving a closer look. It would seem entirely probable that these are the most difficult offenders to classify. They are also more than likely to be the offenders who are the most intractable to change.

While these results do need to be replicated if they are to be relied on, it would seem wise to keep a watchful eye on offenders who come back again and again for assessment. Not all of the offenders in Sample 4 were assessed again because of arrest. A check of arrest status between assessments revealed that 161 (56%) of the 285 offenders in Sample 4 had not been arrested between assessment 1 and assessment 4.

What seemed to be happening in assessment 4 is that there seems to be an upper limit to the prediction of arrest. LSI-R #2, LSI-R #3, and LSI-R #4, all had enough information to accurately predict arrest. The issue that must be considered when using assessments such as the LSI-R is the fact that not all offenders who are still committing crimes get caught. The outcome variable (arrests) could be confounded by the rate of apprehension. The Minnesota BCA (2005) reports that only 50% of the crimes reported in Stearns County in 2005 resulted in an arrest. Even if a measuring tool could predict with 100% accuracy that the person was still offending, the number of offenders who get caught would limit the outcome. While there is a statistical probability that offenders committing crimes will get caught, and with larger sample sizes the more likely this will occur, in the fourth sample, there may not have been enough offenders for accurate measurement.

This problem of an apparent limit to prediction is not new. Kroner, Mills, and Reddon (2005) did an experiment where they took four popular risk assessment tools, the Psychopathy Checklist-Revised, Level of Service Inventory-Revised, Violence Risk Appraisal Guide, and the General Statistical Information on Recidivism and developed four risk assessment tools by randomly combining the scores from individual items on the four instruments. They found that random combination of factors were statistically comparable, as prediction tools, to the original instrument. There was a consistent upper limit to prediction that hovered around a correlation rate of about .30 to .40.

Similarly, Seto (2005) tried to find a statistically better risk scale for Adult sex offenders by combining the results from the Violence Risk Appraisal Guide, the Sex Offender Risk Appraisal Guide, the Rapid Risk Assessment for Sexual Offense Recidivism, and the Static-99. There was no combination of scales that could predict better than the original scales.

From the behavior of the LSI-R on the third and fourth assessments, it would seem that there might be an upper limit on the predictive power that is possible. The correlation rates for the LSI-R scores and arrest in this population seem to top out at .275. No matter how much information is collected, the outcome seems to be limited to the rate of arrest and conviction.

The overall magnitude of change seems to be slightly correlated with arrest, although this may be do to the change in risk level. The changes between assessments that seemed most likely to be consistently associated with recidivism, were changes in the Criminal History, Alcohol/Drug, and Emotional/Personal LSI-R sub scales.

REFERENCES

REFERENCES

Agnew, R. (1984). Appearance and Delinquency. *Criminology*, 22(3), 421-440.

Akers, R.L. (1973). *Deviant Behavior: A Social Learning Approach*. CA: Wadsworth Publishing Company, Inc.

Amen, D.G. (1998). *Change your brain change your life*. New York: Three Rivers Press.

Andrews, D.A. (1982). The Level of Supervisory Inventory (LSI): The first follow-up. Toronto: Ontario Ministry of Correctional Services.

Andrews, D.A., & Bonta, J. (1995). LSI-R: The Level of Service Inventory-Revised. Toronto: Multi-Health Systems, Inc.

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

Andrews, D.A., & Bonta, J. (2006). *The Psychology of Criminal Conduct* (4th ed.). Cincinnati, OH: Anderson Publishing Co.

Andrews, D.A., Bonta, J., & Wormith, J. (2006). The Recent Past and Near Future of Risk and/or Need Assessment. *Crime & Delinquency*, 52(1), 7-27.

Andrews, D.A., & Robinson, D. (1984). The Level of Supervision Inventory: Second Report. Toronto: Ontario Ministry of Correctional Services.

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

Andrews, D.A., Bonta, J., & Wormith, J.S. (2004) *Level of Service/Case Management Inventory: LS/CMI Manual*. Toronto: Multi-Health Systems, Inc.

Andrews, D.A., Zinger, I., Hoge, R.D., Bonta, J., Gendreau, P., & Cullen, F. (1990). Does Correctional Treatment Work? A Clinically Relevant and Psychologically Informed Meta-Analysis. *Criminology*, 28(3), 369-404.

Austin, J. (2004). The proper and improper use of risk assessment in corrections. Paper presented at the Association of Paroling Authorities International Savannah Conference 2004. Retrieved March, 25, 2007from http://www.apaintl.org/Pub-Conf2004-ProperUse.html.

Bachman, R., & Schutt, R.K. (2003). *The Practice of Research in Criminology and Criminal Justice* (2nd ed.). Thousand Oaks, CA: Pine Forge Press.

Baird, S., Heinz, R., & Bemus, B. (1979). Wisconsin case classification/staff deployment project: A two-year follow-up report. Madison, WI: Department of Health and Social Services. Bureau of Community Corrections.

Bandura, A. (1977). *Social Learning Theory*, Prentice-Hall, Englewood Cliffs, N.J.

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. (2000). Offender Assessment: General Issues and Considerations. *Compendium 2000 on Effective Correctional Programming*, Retrieved on March 28th, 2007 from http://www.csc-scc.gc.ca/text/rsrch/compendium/2000/chap_4_e.shtml.

Bonta, J. (2002). *Offender Risk Assessment: Guidelines For Selection and Use. Criminal Justice and Behavior*, 29(4), 355-379.

Bonta, J., & Motiuk, L.L. (1987). The Diversion of Incarcerated Offenders to Correctional Halfway Houses. *Journal of Research in Crime and Delinquency*, 24(4), 320-323.

Burgess, E. W. (1928). Factors determining success or failure on parole. In A. A. Bruce, E. W. Burgess, J. Landesco, and A.J. Harno (Eds.) *The workings of the indeterminate sentence law and the parole system in Illinois*. Springfield, IL: Illinois Committee on Indeterminate-Sentence Law and Parole.

Byrne, J. (2006). Introduction: Why Assessment "Matters" in an Evidence-Based Community Corrections System. Federal Probation, 70(2). Retrieved on March 28th, 2007 from http://www.uscourts.gov/fedprob/September_2006/intro.html.

Center for Substance Abuse Treatment. (2005). Substance Abuse Treatment for Adults in the Criminal Justice System. Treatment Improvement Protocol (TIP) Series 44. DHHS Publication No. (SMA) 05-4056. Rockville, MD: Substance Abuse and Mental Health Services Administration, 2005.

Chiappe, D., Brown, A., Dow, B., Koonz, J., Rodriguez, M., & McCulloch, K. (2004). Cheaters are looked at longer and remembered better than cooperators in social exchange situations. *Evolutionary Psychology*, 2, 108–120.

Cosmides, L. (1989). The logic of social exchange: Has natural selection shaped how humans reason? Studies with the Wason selection task. *Cognition*, 31, 187-276.

Coulson, G., Ilacqua, G., Nutbrown, V., Giulekas, D., & Cudjoe, F. (1996). Predictive utility of the LSI for incarcerated female offenders. *Criminal Justice and Behavior*, 23(3), 427-439.

Evans, E.C. (1969). Physiognomics in the Ancient World. *Transactions of the American Philosophical Society*, New Serial, 59(5), 1-101.

Gendreau, P., Goggin, C., & Little, T. (1996). Predicting Adult Offender Recidivism: What Works! User Report: 1996-07, Ottawa: Solicitor General of Canada.

Gendreau, P., Goggin, C., & Smith, P. (2002). Is the PCL-R Really the “Unparalleled” Measure of Offender Risk?: A Lesson in Knowledge Cumulation. *Criminal Justice and Behavior*, 29(4), 397-426.

Girard, L., & Wormith, S.J. (2004). The Predictive Validity of the Level of Service Inventory-Ontario Revision on General and Violent Recidivism Among Various Offender Groups. Criminal Justice and Behavior, 31(2), 150-181.

Glaze, L.E. & Bonczar, T.P. (2006). *Probation and Parole in the United States, 2005*. (Report No. NCJ-215091). Washington, D.C.: U.S. Department of Justice, Bureau of Justice Statistics.

Glover, A.J., Nicholson, D.E., Hemmati, T., Bernfeld, G.A., & Qunsey, V.L. (2002). A comparison of predictors of general and violent recidivism among high-risk federal offenders. *Criminal Justice and Behavior*, 29(3), 235-249.

Glueck, S., & Glueck, E. (1950). *Unraveling Juvenile Delinquency*. Cambridge, MA: Harvard University Press.

Grove, W.M., & Meehl, P.E. (1996). Comparative efficiency of informal (subjective, impressionistic) and formal (mechanical, algorithmic) prediction procedures: The clinical-statistical controversy. *Psychology, Public Policy, and Law*, 2, 293-323.

Grove, W.M., Zald, D.H., Lebow, B.S. Snitz, B.E. & Nelson, C. (2000). Clinical Versus Mechanical Prediction: A Meta-Analysis. *Psychological Assessment*, 12(1), 19-30.

Hanley, J.A., & McNeil, B.J. (1982). The meaning and use of the area under a receiver operating characteristic (ROC) curve. *Radiology*, 143, 29-36.

Harrison, P.M., & Beck A.J. (2006). *Prisoners in 2005* (Report No. NCJ-215092). Washington, D.C.: U.S. Department of Justice, Bureau of Justice Statistics.

Hart, H. (1923). Predicting Parole Success. *Journal of Criminal Law and Criminology*, 14(3), 405-414.

Heatherton, T.F., & Weinberger, J.L. (Eds.) (1994). *Can personality change?* Washington, D.C.: American Psychological Association.

Heppner, P.P., Kivlighan Jr., D.M., & Wampold, B.E. (1999). *Research Design in Counseling*. Belmont, CA: Wadsworth Publishing Company.

Hollin, C. R. (2002). Risk-needs assessment and allocation to offender programmes. In J. McGuire (Ed.). *Offender Rehabilitation and Treatment: Effective Programmes and Policies to Reduce Re-offending*. Chichester: John Wiley and Sons.

Hollin, C.R. (2004). To treat or not to treat? An historical perspective. In C.R. Hollin (Ed.), Essential Handbook of Offender Assessment and Treatment (pp. 1-13). Chichester: Wiley.

Howe, M., & Lore, J. (2005). *Implementing Evidence-Based Practice in Community Corrections: Quality Assurance Manual*. Boston, MA: Crime and Justice Institute.

Jenson, T.D. (1998). Examining the predictive validity of the Level of Service Inventory-Revised (LSI-R). Unpublished Master's Thesis. Mankato, MN.

Kroner, D.G., Mills, J.F., & Reddon, J.R. (2005). A coffee can, factor analysis, and prediction of antisocial behavior: The structure of criminal risk. *International Journal of Law and Psychiatry*, 28, 360-374.

Kurtzberg, R.L., Mandell, W., Lewin, M., Lipton, D.S., & Scuster, M. (1978). Plastic surgery on offenders. In N. Johnston and L. Savitz (Eds.) *Justice and Corrections*, (pp.688-700). New York: Wiley.

Langan, P.A., & Levin, D.J. (2002). *Recidivism of prisoners released in 1994*. Washington, DC: Bureau of Justice Statistics, U.S. Department of Justice.

Latessa, E.J., & Allen, H.E. (2003). Corrections in the Community (3rd ed.). Cincinnati, OH: Anderson Publishing Company.

Lombroso, C. (2006). *Criminal Man*. (M. Gibson & N.H. Rafter, Trans.) Durham and London: Duke University Press. (Original work published 1876)

Lombroso, C., & Ferrerro, G. (2004). *Criminal Woman, the Prostitute, and the Normal Woman*. (N.H. Rafter & M. Gibson, Trans.) Durham and London: Duke University Press. (Original work published 1893)

Lowenkamp, C.T., & Latessa, E.J. (2001). Validating the Level of Service Inventory-Revised in Ohio's Community Based Correctional Facilities. Unpublished Manuscript. Cincinnati, Ohio: University of Cincinnati Division of Criminal Justice.

Lowenkamp, C.T., & Latessa, E.J. (2002). Evaluation of Ohio's Community Based Correctional Facilities and Halfway House Programs. Cincinnati, Ohio: University of Cincinnati Division of Criminal Justice Center for Criminal Justice Research.

Latessa, E.J. & Lowenkamp, C.T. (2005). The Role of Offender Risk Assessment Tools and How to Select Them. *For the Record*, 4th Quarter 2005, 18-20.

Martinson, R. (1974). What Works?: Questions and Answers About Prison Reform. *The Public Interest*, 35, 22-45.

Maruna, S. (2001). *Making good: How ex-convicts reform and rebuild their lives*. Washington, D.C.: American Psychological Association.

Meehl, P.E. (1954). *Clinical Versus Statistical Prediction: A Theoretical Analysis and a Review of the Evidence*. Minneapolis, MN: University of Minnesota Press.

Minnesota Bureau of Criminal Apprehension (2005). Minnesota Crime Information 2005. Minnesota Bureau of Criminal Apprehension, Criminal Justice Information Systems, Uniform Crime Report.

Moffitt, T.E. (1993). Adolescence-limited and life-course-persistent antisocial behavior: A developmental taxonomy. *Psychological Review*, 100, 674-701.

Motiuk, L.L. (1991). Antecedents and consequences of prison adjustment: A systematic assessment and reassessment approach. Unpublished doctoral dissertation, Carleton University, Ottawa, Ontario.

Motiuk, L.L., Bonta, J., & Andrews, D.A. (1990). Dynamic predictive criterion validity in offender assessment. A paper presented at the Canadian Psychological Association Annual Convention, Ottawa, June, 1990.

Mumola, C. (2000). *Incarcerated parents and their children* (Report No. NCJ-182335). Washington, D.C.: U.S. Department of Justice, Bureau of Justice Statistics.

National Institute of Corrections (2003). *Offender Assessment. Topics in Community Corrections, Annual Issue 2003.* Washington DC: U.S. Department of Justice.

National Institute of Corrections (2004). *Assessment Issues for Managers. Topics in Community Corrections, Annual Issue 2004.* Washington DC: U.S. Department of Justice.

O'Connor, T. (2006). *Anthropological Criminology.* Retreived on October 7th, 2006 from http://faculty.ncwc.edu/toconnor/301/301lect03.htm.

Panzarella, R. (2002). Theory and Practice of Probation on Bail in the Report of John Augustus. *Federal Probation,* 66(3), 38-43.

Parke, R., & Clarke-Stewart, K.A. (2002). Effects of parental incarceration on young children. Presented at: From Prisons to Home conference, January 30-31, 2002. Washington, D.C.: Urban Institute. Retrieved on March, 27th, 2007 from http://www.urban.org/url.cfm?id=410627

Peterson, W.W., Birdsall, T.G., & Fox, W.C. (1954). The theory of signal detectibility. *IEEE Transactions on Information Theory*, 4(4), 171-212.

Rafter, N. (2004). Earnest A. Hooton and the biological tradition in American society. *Criminology*, 42 (3), 735-771.

Raynor, Peter (In Press). Risk and need assessment in British probation: The contribution of LSI-R. *Psychology, Crime, & Law*, In Press.

Raynor, P., Kynch, J., Roberts, C., & Merrington, S. (2000) *Risk and Need Assessment in Probation Services: an Evaluation, Research Study 211*. London: Home Office.

Sampson, R.J., & Laub, J.H. (1993). *Crime in the Making. Pathways and Turning Points Through Life*. Cambridge, MA: Harvard University Press.

Sampson, R.J., & Laub, J.H. (2005). A life-course view of the development of crime. *Annals of the American Academy of Political and Social Science*, 602, 12-45.

Sellin, T. (1958). Pioneers in Criminology. XV. Enrico Ferri (1856-1929). *The Journal of Criminal Law, Criminology, and Police Science*, 58(5), 481-492.

Seto, M.C. (2005). Is more better? Combining actuarial risk scales to predict recidivism among adult sex offenders. *Psychological Assessment*, 17(2), 156-167.

Simourd, D.J. (2004). Use of dynamic risk/need assessment instruments among long-term incarcerated offenders. *Criminal Justice and Behavior,* 31(3), 306-323.

Simourd, D.J., & Malcolm, B.P. (1998). Reliability and validity of the Level of Service Inventory-Revised among federally incarcerated sex offenders. *Journal of Interpersonal Violence*, 13(2), 261-274.

Stephan, J.J. (2004). *State prison expenditures, 2004*. (Report No. NCJ-202949). Washington, D.C.: U.S. Department of Justice, Bureau of Justice Statistics.

The Carey Group (2005). *Evidence-Based Practices in the Community Corrections Division of Stearns County Human Services: Final report*, *September 15, 2005*.

Toby, J. (1957). Social Disorganization and Stake in Conformity. *Journal of Criminal Law, Criminology, and Police Science*, 48(1), 12-17.

Retrieved from "http://en.wikipedia.org/wiki/Social_control_theory"

Walsh, A. (2002). *Biosocial Criminology: Introduction and Integration*. Cincinnati: Anderson Publishing Co.

Warner, S.B. (1923). Factors determining parole from the Massachusetts Reformatory. *Journal of Criminal Law and Criminology*, 14, 172-207.

Watkins, M.W. (2000). An EXCEL program for calculating and graphing the Receiver Operating Characteristic (ROC) [Computer software]. State College, PA: Ed & Psych Associates.

Wright, K.N., Clear, T.R., & Dickson, P. (1984). Universal applicability of probation risk assessment instruments: A critique. *Criminology*, 22(1), 113-134.

Wolfgang, M.E. (1961). Pioneers in Criminology: Cesare Lobroso (1835-1909). The Journal of Criminal Law, Criminology, and Political Science, 52(4), 361-391.