The Nonlinear Dynamics of Criminal Behavior

Thomas K. Arnold

M.S. Criminal Justice

St. Cloud State University

Abstract

One of the more enduring puzzles in criminology is the combination of both stability and change in criminal behavior over the life course. This study examines how the criminal offender's risk of recidivism changes over time and provides guidance for criminology, forensic psychology, and criminal justice. A multidisciplinary approach leads to improvements in the methods of dynamic risk assessment, and provides an analysis of the stability and dynamism present in offender risk levels, both between and within individuals. Methods are demonstrated for improving measurement accuracy in test-retest environments, and creating measurement criteria to determine the change in risk score needed to predict a change in recidivism. Results include comparisons of between group and within group changes in risk levels over time, indications of how the rank order levels of risk change over time, measurements of the temporal stability of risk, and an indication that changes in risk are typically nonlinear, rather than linear in nature. An analysis of the results is presented that suggests methods for improving the risk assessment process, precautions needed when measuring results from treatment programs, ways to improve desistance research, and future directions for theory.

The Nonlinear Dynamics of Criminal Behavior

There are many unanswered questions in criminology related to the temporal stability of the criminal offender's risk of committing new crimes. Gottfredson and Hirschi (1990) suggest that the level of risk is stable over the life course, only declining due to the effects of age. Sampson and Laub (1993) suggest that risk levels are more dynamic, and can change over time if offenders encounter the right local life circumstances (Horney, et. al., 1995). Nagin and Paternoster (2000) examined the question of whether the level of criminal activity is due to stable individual differences in the propensity for criminal behavior, or to changing life circumstances of the offender, and conclude that both matter. They remark that there is much we don't understand about the relative contributions of between individual and within individual variability in the level of offending over the life course, and more study is needed. This article addresses two main issues relevant to understanding the stability of criminal offender risk over time, the measurement and interpretation of change, and the dynamics of risk.

These issues have important implications for corrections practice, risk prediction research, and theory building in criminology. It may not be apparent to the reader unfamiliar with these concepts, but the discussion and analyses done in this study provide much needed guidance in areas that are not often addressed in criminal offender research. That isn't to say that research and theory building with regard to offender risk hasn't been done. There has been an extended focus on the dynamics of short-term risk in sexual and violent recidivism studies (Douglas, & Skeem, 2005), and general recidivism (Zamble, & Quinsey, 1997). There is also a considerable amount of research that has been done on the long-term dynamics of risk (Ezell, & Cohen, 2005; Laub, & Sampson, 2003). There has been much less emphasis placed on the dynamics of risk and offending in the intermediate periods covering several years. For exceptions see Farrington et. al. (1986), Horney, et. al. (1995), and Piquero, et. al. (2002). This study adds to the knowledgebase on the dynamics of risk in intermediate time periods.

This in itself would appear to provide a sufficient reason for this study, but this study was actually not designed to serve the research community, it was designed to serve the corrections practitioner. Longitudinal research efforts of experimental quality with criminal offenders are extremely limited due to the high cost of the data collection, but there is a vast army of corrections practitioners doing longitudinal offender research every day, and for the most part, scientific analyses of the results of their efforts are not being done. When scientific analyses are done, the results are typically not published. This army of researchers consists of the corrections practitioners, who are collecting huge amounts of longitudinal data on offenders.

Corrections officials need to know the principles of longitudinal research, and what to expect in terms of changes in risk. It might be suggested that practitioners are not doing research, but I will argue that each time a corrections practitioner reassesses an offender, there is an effort to do an N=1 study to see if there has been any progress in reducing the risk of offending since the last assessment was done.

There are two areas the practitioner needs to be concerned with: dynamic measurement and the level of change to expect. Any time a second assessment is done, the element of time enters into the measurement process, adding multiple levels of complication. Much more care must be taken in dynamic assessment than in static assessment, and these issues have generally been ignored. It is important to know how much change is occurring when setting measurement frequency. It makes no sense to spend 1-2 hours reassessing someone who isn't changing. On the flip side, it also might not pay to continually reassess if changes in risk aren't stable over time. Discussions of time and the level change expected are missing from the practice literature.

Guidance in the area of measurement and the level of change to expect in any given time period is important, since corrections departments work with the largest population of offenders under supervision. At the end of 2006, over five million offenders, or 1.9% of the U.S. population, were under supervision in community corrections (Glaze, & Bonczar, 2007). This number is over twice the number of offenders that were held in prisons and jails. One of the key functions of community corrections workers is assessing the offender's likelihood of recidivism. The field of corrections has taken a much more dynamic approach to risk than traditional criminologists, and most corrections departments use some form of structured dynamic risk assessment instrument for offender classification (Hubbard, et. al., 2001).

Structured dynamic risk assessment instruments measure a number of static risk indicators such as criminal history, as well as a number of time varying risk indicators such as unemployment and drug use. A risk score is produced that is generally associated in a linear fashion with the criminal offender's likelihood of recidivism. This study uses a dynamic risk measure called the Level of Service Inventory-Revised (LSI-R; Andrews and Bonta, 1995) to assess changes in the risk of recidivism. The LSI-R is one of the more accurate risk prediction instruments available (Gendreau, et. al., 1996; Gendreau, et. al., 2002), and is used by about 15% of corrections departments in the U.S. (Hubbard, et. al.).

The LSI-R is a revised version of the Level of Supervision Inventory (LSI; Andrews, 1982), developed in Canada. The LSI-R, which will hereafter be referred to as the LSI, uses a structured interview and criminal records search to answer 54 yes/no questions related to both static and dynamic risk factors. The yes answers are added together to produce a risk score. According to its authors, the LSI is able to measure changes in risk that occur over time (Andrews, & Bonta, 2006: p. 290; Andrews & Bonta, et. al., 1990:p. 32; Bonta, 2002:p. 368).

Andrews and Bonta recommend periodic reassessment with the LSI and corrections practitioners have been following their advice. The corrections officials who generated the data for this study were reassessing with the LSI at seven-month intervals (S.D.=2-3 months). The National Corrections Institute sample plan (Kreamer, 2004) recommends reassessment with the LSI every 12 months. The Center for Substance Abuse Treatment recommends reassessment, and provides a sample schedule from Orange County, CA for 6-month evaluations with the LSI (Peters, & Wexler, 2005; p. 40). A one-year reassessment period with the LSI is a policy mandate in Minnesota (State of MN, 2006). The recommended time periods do not appear to be guided by research that indicates the level of change in risk to expect in these periods.

The following discussion and analysis will provide an overview of issues related to risk assessment on multiple occasions. This will be followed by a general overview of risk assessment, with an emphasis on the structure and history of the LSI. Then, there will be a series of discussions of the various analyses done in this study. Briefly, this will include a discussion of some earlier studies done on the dynamic properties of the LSI and the temporal stability of risk, an overview of issues related to improving reassessment accuracy, methods that can be used to assess prediction accuracy of risk instruments, the level of change in risk that might be expected based upon the results of previous studies and theoretical predictions, and a discussion of methods used to analyze individual growth curves.

The following discussions and analyses are not limited to issues with the LSI. The principles of other dynamic risk assessment instruments are essentially the same as the LSI principles. These analyses could provide methods that would open up a vast storehouse of data for researchers trying to understand how offender risk changes over time. The results of the analyses done in this study provide a unique look at the dynamics of offender risk.

Longitudinal Research with Risk Scores

Most research in criminology has examined cross sectional differences between individuals in the likelihood of criminal behavior with respect to one or more variables of interest. More recently, an increased emphasis has been placed on longitudinal research that measures changes in the level of offending within individuals over time. In within-individual studies of change, individual offenders are followed for some period of time that must include at least two time points. The present study does something relatively new to the risk assessment field by using a ratio level variable as a measure of the criminal offender's risk of recidivism both between individuals and within individuals over several time points.

By using a ratio level measure of risk, represented by a risk score, much more interesting questions can be asked than when using a simple dichotomous variable such as criminal/non-criminal to describe behavior. Adding the element of time to the analyses creates a number of interesting variables, and as more time points are added, the amount of information that can be gained increases. With two time points, the mean risk level and the direction, magnitude, and rate of change can be analyzed. Outcome can be assessed at time 1 and time 2. Using three time points, the trajectory of change can be determined. Change can continue in the same direction at the same or a different rate, either accelerating or decelerating, change can stop, or change can occur in a new direction, with a new magnitude and rate. It is possible to study the range, mean score, and standard deviation of an individual's risk scores, and the trend in scores can be calculated. As more time points are added, a trajectory of change can be plotted, and the variation of trends in the changes in risk over time can be determined. The total amount of change that occurs over time can be assessed, both within and between individuals. This study will explore the effects of many of these changes in risk score over time.

On Time and Method

In their book, "On Time and Method", Kelly and McGrath (1988) provide one of the more insightful discussions of issues related to the study of human behavior over time. Many theorists and researchers are subject to a fundamental error in thinking because they tend to think that human behavior has inertia. To paraphrase their position with an analogy, an assumption of inertia leads to the conclusion that behavior changes obey Newton's laws of physics. The Newtonian perspective on change implies that change doesn't occur unless the person is acted upon by an outside force, the level and rate of change are proportional to the force applied, changes tend to occur in a single direction and have a uniform rate over time, and changes keep occurring until some sort of friction slows them down. Changes in behavior patterns can have a number of functional forms besides linear change, including all at once, maintained over time, not maintained over time, delayed, and cyclic nonlinear. Human behaviors often change in a nonlinear cyclic manner, and this fact is often ignored in research studies because researchers and theorists take a Newtonian approach to behavior change.

In addition to issues with trajectory there are several other factors that must be examined when doing longitudinal research. Kelly and McGrath list four reasons that an assessment score might change between assessments: 1) real change in the phenomena of interest, 2) fluctuation in the phenomena of interest over time, 3) systematic differences in the measurement process, and 4) unreliability of the measurement instrument. These four items are confounded with each other and each will occur at some level any time a measurement is done on two or more occasions (p. 58). The commentaries of Kelly and McGrath will be used as a guide in the discussions that follow. Many of the conclusions drawn from previous research appear to be related to not attending to one or more of these four items, and the adoption of a linear inertial paradigm.

The Dynamics of Risk

Kelly and McGrath repeatedly stress the need for understanding the time frame in which change occurs. A discussion by Brown (2002) suggests that there are four time frames that studies of the dynamics of offender risk often focus on, and these are related to the rates of change in risk factors, which include; static, slowly changing, intermediate changeability, and rapidly changing. Static risk factors don't change over the offender's entire life. Gender, and age of the onset of criminal behavior are static risk factors. Slowly changing risk factors, such as personality traits, change over long periods of time. Intermediate risk factors, such as alcohol or substance dependence, change over several months and years. Rapidly changing risk factors, such as intoxication level and mood, change over minutes, hours and days.

This study uses some static risk factors, such as gender and year of birth, as controls, but analysis of rapidly changing risk is not possible because the reassessment intervals are too long (M=7 mo., S.D. = 2-3 mo.). The data used in the following analyses provides information on the two remaining time frames, slow change, and intermediate change.

The long-term trajectory of risk. The trajectory of risk over the life course is fairly well understood. Hirschi and Gottfredson (1983) demonstrated that the population level age-crime distribution follows a distinctive pattern, referred to as the age-crime curve, which shows that crime rates rise for children and adolescents, peak in late adolescence and early adulthood, and then begin to decline over time, first rapidly, and then more slowly, until there are few offenders left in the 60 year old category. Group level modeling indicates that there are different groups of offenders, with various peak ages and levels of offending, but that most offenders follow some variation of the age-crime curve (Ezell, & Cohen, 2005).

The dynamics of risk are conceptualized as a slowly changing process by many of the dynamic theoretical orientations, including criminal career research (Blumstein et. al., 1986), developmental criminology (Loeber, & Le Blanc, 1990), and the general theory of crime (Gottfredson, & Hirschi, 1990). In the following discussion, these three orientations will be contrasted with the position taken by life course criminology (Sampson, & Laub, 1993), which has adopted a more dynamic and nonlinear view of offender behavior change.

Blumstein et. al. conceptualized criminal offending as a career. They found that the incidence of offending varied dramatically between persons, with some criminals, called career criminals, committing disproportionately more crimes than others. They suggested that researchers focus the majority of attention on career criminals, since non-career criminals commit fewer crimes. Career criminals were found to have a considerable stability in criminal offending over time. Estimates of residual career length indicated that a criminal offender committing a crime at any point had a mean residual career length of 7 to 10 years between the ages of 20 to 45, dropping to about 2 years at age 60 (p. 93).

Loeber and Le Blanc called for the study of criminal behavior from a developmental perspective, and introduced a framework for studying the developmental processes involved in the trajectory of criminal behavior, which include activation, aggravation, and desistance. The concepts of a developmental trajectory, acceleration, stabilization, and deceleration were used, which bring to mind an image of criminal behavior as a curvilinear inertial process.

Gottfredson and Hirschi (1990) developed one of the more popular explanations for the age-crime distribution of criminal offenses, suggesting that there are differences between individuals in the propensity for criminal conduct. These differences are related to a personality trait called low-self control, which is thought to develop in childhood and is essentially unchanging for the person's entire life, like a person's height (Hirschi, & Gottfredson, 2001: p. 91). They attribute the decline in criminal activity over the life course to the effects of socialization. Within individual variability in the numbers of crimes committed is seen as a function of self-control, socialization, and opportunities to commit crimes. From a time interval perspective, levels of self-control are a life-long property, socialization is slowly increasing and operates equally for everyone with age, and opportunity operates in shorter term time periods. They posit that trajectories of risk vary in height between individuals and the rank orders of risk do not vary as people age. An image of life course risk trajectories stacked like Russian dolls seems to fit their conceptualization. This is a particularly linear inertial perspective.

Sampson and Laub (1993) have taken a more nonlinear perspective. They build on the life course dynamics framework developed by Elder (1985), and propose a life course model for the study of criminal behavior. Elder suggested that three basic elements characterize life course dynamics; trajectories, transitions, and turning points. Trajectories are seen to provide a framework for "linking states across successive years" (p. 31). Life states consist life circumstances such as employment, marriage, earnings, or health status. Trajectories can encompass several transitions occurring in shorter periods of time. Transitions are seen as "changes in state that are more or less abrupt" (p. 30), such as getting or losing a job. Some transitions, called "turning points" (p.35), can have a major impact on the direction of life paths. Sampson and Laub argued that the trajectory of criminal behavior is due to the age-graded development of informal social control over the life course. Sampson and Laub attributed the trajectory of risk over the life course to within individual differences in age-graded informal social control, and found that negative transitions such as being sent to prison or reform school, or positive transitions such as getting a job, joining the military, or entering into a stable relationship could create turning points in the life course, leading to an early desistance from crime. The concept of a transition is inherently nonlinear, since it indicates abrupt changes. A transition could fit the model of a change followed by ongoing stability.

These theoretical orientations each develop the concept of an inertial trajectory to some extent. All of these orientations appear to predict declining rather than increasing risk levels over time. None of these theories appears to predict risk levels that increase over time, or cyclical nonlinear fluctuations in risk level.

The Intermediate Interval Dynamics of Risk. There are few studies that have looked at risk levels over the intermediate term. Two studies are often cited that support a claim that offending rates change within the individual over time as risk factors change. Farrington et. al. (1986) examined the level of offending for 411 adolescent London males from the ages 14 to 18 and found that they offended less frequently when employed than unemployed. Horney et. al. (1995) interviewed 658 newly convicted male felons, asking questions regarding their local life circumstances and simultaneous levels of offending in the 25 to 30 months before their arrest. They found that the level of offending changed when offenders went through transitions in life states. Using drugs and living with a girlfriend were significantly related to increases in any type of crime. Starting work, heavy drinking, and drug use were significantly related to an increase in property crime. Living with a wife was significantly related to a decrease in the level of assault. The findings by Horney et. al. provide support for the contention that offending rates change within individuals. It is interesting to note that their findings indicate that starting work increases the risk for property crime. This is directly opposite to the result expected by Andrews and Bonta.

Dynamic Risk Assessment

Initial methods for assessing the criminal offender's risk of recidivism grew out of a failed attempt to discover whether it was possible to predict recidivism from offender histories. Warner (1923) had taken on the job of trying to analyze survey data collected on 680 prisoners released from the Massachusetts reformatory. Of the 680 prisoners, 80 were not released, 300 were released and recidivated, and 300 were released and didn't recidivate. Warner analyzed the survey data and advised that it had no value for predicting parole success. Hart (1923) reanalyzed the data by using correlation analysis and found that the data could be used to create a scale that would predict, with a fair degree of accuracy, which offenders would succeed on parole. Burgess (1928) went on to popularize the method Hart developed, and many different offender classification systems have been developed in the interim.

A problem with the early methods of risk prediction was the reliance on static factors, consisting of historical data that could only increase in magnitude. This meant that reductions in the risk of recidivism could not be measured. In order to rectify this deficiency, the Wisconsin Case Management Classification System (CMC; Baird et. al., 1979) was developed in the U.S., and the Level of Supervision Inventory (LSI; Andrews, 1982) was developed in Canada. Both instruments added dynamic risk factors that could change over time in either direction.

The accuracy of risk assessment instruments, or "predictive validity", is typically calculated using a point bi-serial correlation to determine the correlation rate between a set of risk scores and a dichotomous measure of rule violation during a fixed period of time after the assessment (usually one year). Typical predictive validity levels for various current generation risk assessment instruments fall between .30 and .40, with the LSI tending to be one of the more accurate risk prediction instruments available (Andrews et. al. 2006).

The Level of Service Inventory-Revised

Andrews and Bonta (2006) suggest that dynamic risk factors such as unemployment and drug use are criminogenic "needs". They make a convincing argument that criminogenic needs are related to the risk of recidivism, and that treatment efforts designed to reduce criminogenic needs will result in a lower rate of recidivism. Their argument is that effective offender treatment should follow the Risk, Needs, and Responsivity (RNR) principles. The Risk principle states that most treatment efforts should be focused on higher risk offenders, and less effort should be placed on treating lower risk offenders (Andrews, & Dowden, 2006). The Needs principle states that treatment is most effective when the focus is placed on treating criminogenic needs. The Responsivity principle states that offender treatment should be tailored to the individual offender's learning style and motivation level. One of the primary requirements of the RNR model is accurate dynamic risk assessment, and Andrews and Bonta have developed the LSI to provide risk assessment and treatment classification.

The LSI manual states the total LSI score is purported to measure the "propensity for rule violation" (Andrews, & Bonta, 1995: p. 37). The LSI can be broken down into ten sub-scales which include Criminal History (10 pts.), Education and Employment (10 pts.), Financial (2 pts.), Family/Marital (4 pts.), Accommodations (3 pts.), Leisure and Recreation (2 pts.), Companions (5 pts.), Alcohol and Drugs (9 pts.), Emotional and Personal (5 pts.), and Attitude and Orientation (4 pts.). The criminal history subscale consists of historical (static) factors, and the rest of the subscales are partly static and partly dynamic measures of criminogenic needs.

Andrews and Bonta have promoted the use of the LSI as a dynamic risk prediction instrument and claim that the LSI is accurate enough to measure changes in offender risk. This claim is supported by analyses that were done using the LSI in test-retest studies (Andrews, &

Robinson, 1984; Motiuk, 1991, Motiuk, et. al., 1990; Raynor, 2007, Raynor, et. al., 2000). These test-retest studies demonstrated that when LSI scores are collected on two occasions, the second LSI scores are better at predicting recidivism in the period after the second assessment than the first. In other words, the predictive validity improves with reassessment. The conclusion drawn from these studies was that their first premise (risk changes over time) must be true, and this is the reason retest scores are more accurate in the test-retest analyses.

In an attempt to replicate the earlier research on test-retest validity, Arnold (2007) found evidence that challenged the conclusions drawn by Andrews and Bonta. Arnold used the previous methods with one important difference; he analyzed four sets of LSI scores instead of two. Using four sets of scores allowed tests for improvements in predictive validity from LSI 1 to LSI 2, LSI 2 to LSI 3, and LSI 3 to LSI 4. The accuracy of LSI 3 was also compared with the accuracy of LSI 1 and LSI 2, and LSI 4 was compared with LSI 1, LSI 2, and LSI 3. Analysis of test-retests showed that LSI 2, LSI 3, and LSI 4 all had higher predictive validities than LSI 1 when predictive validity was compared after each assessment. Contrary to the prediction of dynamic predictive validity, substantial improvements were not shown for LSI 3 over LSI 2, and significant improvements in predictive validity were not shown for LSI 4 over LSI 2 or LSI 3.

Given Arnold's results, it appears that one or both of the original premises of Andrews and Bonta, risk changes, or the LSI can measure changes in risk, might be questionable for any analysis after the second assessment. This study explores questions regarding the temporal stability of criminal propensity and issues related to the measurement of offender risk. A series of analysis demonstrates that several factors may have been responsible for the findings by Arnold. The implications of these results for practice, research, and theory will be discussed.

The Dynamic Predictive Validity of the LSI

The purpose of this section will be to discuss the problems that were found regarding the lack of improvements in predictive validity for LSI scores after LSI 2 that were found by Arnold. The underlying premises behind the test will be explored, as well as possible new ways to test for temporal stability in risk scores.

The ability of a risk assessment instrument to measure changes in risk level is assessed by measuring the "dynamic predictive validity" (Motiuk, et. al., 1990) of the risk instrument. Dynamic predictive validity occurs when two sets of risk scores are created at different points in time, generally 6 or more months apart, and the predictive validity of both sets of scores are analyzed in the time period after the second assessment. If the second assessment scores have a higher predictive validity than the first assessment scores, the risk instrument is thought to have dynamic predictive validity.

Several tests of the LSI, using LSI scores collected at two points in time, have indicated that the LSI has dynamic predictive validity (Andrews, & Robinson, 1984; Motiuk, 1991, Motiuk, et. al., 1990; Raynor, 2007, Raynor, et. al., 2000). Each of these studies, using a variety of methods, has indicated that LSI 2 scores are better predictors of recidivism than LSI 1 scores.

Arnold (2007) tried to replicate the previous dynamic predictive validity tests, extending the pair wise analysis from two LSI assessments to four LSI assessments. Five different methods were used to calculate the differences in predictive validity between scores, direct comparison of LSI scores with violation rates using a cross-tabs table (Andrews, 1982), correlation rates, calculating the improvement in scoring over chance, regression analysis, and comparison by splitting the file to avoid regression to the mean effects (Raynor, 2007). Dynamic predictive validity was found for all subsequent LSI assessments when compared with LSI 1, but not for any subsequent LSI assessments when compared with LSI 2 or other subsequent assessments. An examination of the implicit assumptions of the dynamic predictive validity thesis is needed to find the reasons for the discordant results.

The dynamic predictive validity test is essentially designed to assess the temporal stability of risk. The logic behind this test is that if risk changes over time, and one is measuring risk with a dynamic risk assessment instrument, the second score should be a better predictor of risk after the second assessment. There are several implicit assumptions to this test, which can be broken into risk related assumptions and measurement assumptions.

The four assumptions flow from Kelly and McGrath's reasons a score might change, and are 1) risk is temporally unstable and the changes in risk are great enough to make a difference in recidivism rates, and 2) the changes in risk are linear and inertial in nature. The measurement assumptions are, 3) the error rates in the two sets of LSI scores are similar, and 4) the LSI is accurate enough to measure any changes in risk that occur.

1) It is assumed that the offender's risk of recidivism changes. There is evidence for and against the temporal instability of risk. Arnold found evidence for temporal instability of risk in changes in LSI scores from LSI 1 to LSI 2 that ranged from -21 to 27 with a mean change of –1.43 (S.D. = 6.8) (p. 43). The standard deviation of 6.8 would appear to indicate a significant variability in the LSI scores between assessments. Evidence to challenge the instability thesis can be found in the success of the LSI as a risk predictor. How can a risk prediction instrument be so accurate at predicting risk if the risk levels are changing substantially after the assessment?

2) The dynamic predictive validity thesis assumes either a linear change in risk over time, or that the changes in risk are temporally stable once they occur. The logic of the dynamic predictive validity thesis is that if an offender has a risk score of 10 on LSI 1, and a risk score of

20 on LSI 2, the offender should offend at a rate comparable to other offenders with a risk score of 20 after LSI 2. This will only occur if the offender's risk trajectory is linear or stable enough not to change back during the measurement period. In the linear inertia assumption, the offender's risk would keep moving in the same direction. If it continued in the same direction and moved to 25 for instance, LSI 2 would still be a better predictor. In the linear temporal stability assumption, it is assumed that the risk level won't drop back below 15 after LSI 2, because if that happened, LSI 1 would become the better predictor of recidivism.

3) The error rates must be similar. Classic test theory posits that any test score (S) can be thought of as a sum of the true score (T) and an error score (e), shown as $S = T + e$ (Crocker, & Algina, 1986). High predictive accuracy depends on a low error score. If two sets of assessments are given, $S_1 = T_1 + e_1$, and $S_2 = T_2 + e_2$. Improvements in prediction can occur under two conditions: 1: when T_2 is significantly different from T_1 and the error rates are similar, or 2: if T_1 is similar to T_2 but e_2 is significantly lower than e_1. Both cases produce the same result, and it can be difficult to determine which case is the norm.

4) It is assumed that the LSI is accurate enough to measure change. The LSI produces a rank ordering of risk levels that results in a fairly linear relationship between risk scores and arrest rates. There is a fair amount of overlap in the distributions of arrests by LSI score however. One cannot predict with any degree of certainty that a person with a score of 26 will be any less likely to recidivate than a person with a score of 27. It is fairly clear from the many studies done with the LSI that a person with a score of 10 is substantially less likely to recidivate than a person with a score of 30. The exact accuracy of the LSI is not really clear at this time. A discussion of analyses that will be done to test these four assumptions will follow. Taken together, these analyses reveal the underlying form of the dynamics of risk.

Testing for Dynamic Predictive Validity and the Temporal Stability of Risk

The premise of Andrews and Bonta is that risk levels change over time because the dynamic predictive validity test shows that the LSI accuracy improves on the second LSI. The problem with this method of assessing the temporal stability of risk is that it is subject to confounding due to possible differences in measurement errors between the assessments. A more accurate method of testing for temporal stability is needed that doesn't require a retest score. It is possible to avoid the retest if the first LSI score is used to predict recidivism at different periods in time. For instance, the LSI assessments in this study were taken at seven-month intervals. If the LSI scores are used to predict recidivism in the year after each subsequent reassessment, the accuracy could be checked at 0-12 months, 7-19 months, 14-26 months, and 21-33 months. If there were linear changes in risk levels, the LSI scores should become poorer at prediction (lower correlation rates with violation levels) over time.

If the LSI scores become less accurate over time, it would mean that the offender risk levels are changing in a linear fashion. If the LSI scores don't degrade with time, there could be two possible reasons, risk levels aren't changing, or the changes are nonlinear in nature.

The conclusion that risk levels aren't changing might be questioned based on the results by Arnold provided in the previous section that show a range in LSI score changes of -21 to 27 from LSI 1 to LSI 2. These results indicate that there appear to be rather substantial changes in LSI scores between assessments. If risk levels aren't changing, but risk scores are, either the changes in LSI scores don't reflect a change in risk and are simply due to measurement error, or the underlying risk levels are changing in a nonlinear manner and changing back to the original level or below in the one-year test period for violation.

Inter-rater and Intra-rater Reliability, and Regression to the Mean

This section deals with issues related to the practice of reassessment. Cook and Campbell (1979 p. 51-55) list several items that can lead to erroneous conclusions in a test-retest environment. Two of these items are changes in the measurement instrument and regression to the mean. Each time a corrections official reassesses an offender, the intent is to do a before and after experiment with one person to see if the offender has changed in the intervening interval. It will be demonstrated that certain precautions in measurement and interpretation must be taken, whenever doing reassessments, if valid conclusions are to be drawn.

The measurement instrument used in risk assessment is the corrections official, or rater. There are several changes in the rater that can occur between a test and a subsequent retest. The rater can completely change to another person, causing issues with inter-rater reliability. Since the test is based upon knowledge of the offender, the rater could learn more about the offender in the intervening interval, causing issues with intra rater reliability. The rater could learn how to score an LSI better, changing the scoring method (Flores, et. al., 2006), which could cause another intra-rater reliability issue.

Problems with inter-rater reliability occur when the rater is changed between tests. Iner-rater reliability issues are a well-known problem (Nunnally, & Bernstein, 1994: p. 212) and a test to see if a measurement has poor inter-rater reliability is to have two raters rate the same group of offenders and then check the correlation rate between the sets of scores. The inter-rater reliability of the LSI, listed in the LSI scoring manual (Andrews, & Bonta, 1995), is above r=.80, which is commendable for a rating instrument. Up to 5 point score changes were noted in the manual when raters change, which means that up to a 10% difference in score can occur with no actual change in risk if the raters are switched.

Issues with intra-rater reliability are also a problem. Flores, et. al. (2006) found that raters tend to have a significantly higher correlation between LSI scores and violation rates after three years of practice. If intervals between scoring are long, or the rater learns how to score more accurately in the intervening interval, score changes can occur due to a change in rating.

Another intra-rater issue is a possible change in scoring due to the learning curve a rater has with an offender. The first assessment is usually done after the offender is convicted and before sentencing. The rater has to take the offender's word for many items. On the second assessment, the rater has worked with the offender for several months and can do a better job of scoring. Andrews and Robinson (1984: p. 5), and Aubrey and Hough (1997: p. 23), suggested that raters might get to know the offender better between assessments. The learning curve effect should be less for subsequent LSI assessments because most learning occurs early in the process.

A way to test for effects of inter-rater and intra-rater changes would be to look at the mean score changes between assessments when the rater is held constant and when the rater is switched between assessments for the same group of offenders. This would make the offender group the control group for the procedure. The learning curve could be tested by checking the mean score change from LSI 1-2 and then comparing it with the mean score change from LSI 2-3, and LSI 3-4. If score changes are less between LSI 2-3 and LSI 3-4 than between LSI 1-2, then a learning curve process could be indicated, given the high increase in predictive validity from LSI 1-2.

Another issue with score interpretation is due to regression to the mean (Campbell, & Kenny, 1999). The regression to the mean between LSI 1 and LSI 2 was assessed in the Arnold study and is provided for reference. Regression to the mean lead to average changes of 0-4 points upward for low scoring offenders, and 0-4 points downward for high scoring offenders.

The Precision of Offender Risk Assessment Instruments

There does not appear to be an indication in the assessment literature that indicates the size of change in LSI score required in order for there to be a change in violation rate. To use the words of Bateson (1979: p. 99), the magnitudes of the “differences that make a difference” are not known. Problems with measurement accuracy have been encountered before in other disciplines such as psychology, where there are problems with interpretation of test-retest scores. Two possible problems are measurement errors and temporal instability in the item being assessed (Jacobson, & Truax, 1991; Jacobson, et. al., 1984). Jacobson and others have developed a statistical formula that can be used to determine the amount of change that is needed to make a clinically significant difference in outcome, given the inaccuracies that are inherent in the measurement process. Clinically significant change, in terms of LSI scores, would indicate a change in score large enough to predict a significant change in violation rates.

One way to determine the level of clinically significant change is to use a Reliable Change Index (RCI). The RCI provides a cutoff score that indicates the level of change in assessment score needed to produce a change in the behavior of interest. The formula for the RCI at a 95% probability level, taken from Smith, & Beaton (2008: 289), is

$$RCI = 1.96 * SQRT(2) * S.D._1 * SQRT(1- r_{1-2})$$

Where $S.D._1$ is the standard deviation of the measurement score on the first assessment, and r_{1-2} is the correlation between the first and second sets of assessment scores in a period of nor change. The calculation of the RCI will be done in as part of this analysis, as well as a regression test using dummy variables to determine the level of score change that is required to assure that there has been a significantly change in the actual risk of recidivism. This will provide a way to discern random changes in score from true temporal instability.

The Temporal (In)Stability of Offender Risk

The dynamics of offender risk have important implications for practice, research, and theory. Corrections departments are doing repeat assessments with almost no guidance as to their interpretation. As previously discussed, the amount of temporal instability in risk levels is not clear, there are issues with inter-rater reliability, intra-rater reliability, regression to the mean, and a lack of clear guidelines as to the level of change in risk score that is related to a significant change in the risk of recidivim.

In Arnold's study there was a rather constant rate of change in the LSI scores between assessments. In looking at the high level of change in the LSI risk scores, the question becomes, "How can LSI scores change that much over time and the LSI still provide excellent predictive ability?" One possible solution is provided by Glaser (1964), who suggested that offender follow a zig-zag pattern of offending. Lewin (1964) pointed out that successful change requires a three-step process: unfreezing, moving, and refreezing. What if the offenders never freeze at their new level of risk? The answer to this question requires an analysis of change over time.

The measurement of change is subject to a number of problems that go beyond simple measurement issues. There are issues with reliability of measurement when the researcher starts looking at change scores. At one time, because of these reliability issues, researchers Cronbach and Furby (1970) had suggested that perhaps we shouldn't measure change at all. This position was seen as overly restrictive and others have indicated that, with precautions, we can measure change effectively (Rogosa et. al., 1982; Singer, & Willett, 2003; Willett, 1989). Singer and Willett give explicit instructions for how to model individual growth curves to assess whether change is linear or non-linear. This study provides the very basic first steps in analyzing change, plotting risk trajectories and trying to fit linear trend lines to the individual data.

Study Overview

Several of the items just discussed will be examined as part of these analyses. A test will be performed to assess the dynamic predictive validity of the LSI and test for an underlying temporal instability in offender risk. The effects of changing raters, and having the same rater on multiple occasions, will be assessed to determine whether there are differences in the level of score change between assessments. The inter rater and intra-rater tests, combined with the results of the dynamic predictive validity and temporal stability tests, will provide some indication of whether the improvements in scoring from LSI 1 to LSI 2 are due to a change in risk levels or a learning curve effect.

In order to help practitioners set reassessment intervals, the amount of change needed to predict a significant change in recidivism rates will be determined. The level of change will be assessed over time and compared both between groups and within groups to determine the amount of change that occurs between assessments.

In order to determine the functional form of changes in risk, a set of analyses will be done to determine whether changes in risk level are linear or nonlinear when measured over several points in time. These analyses will include a test to see if changes that exceed the RCI index are maintained over time, and an initial analysis of individual risk trajectories as suggested by Singer and Willet.

The results of these analyses will be discussed in terms of their ability to show whether risk levels are stable or unstable over time. The implications of these results will be explored with an emphasis on possible future directions for theory and research.

METHODS

Sample Demographics

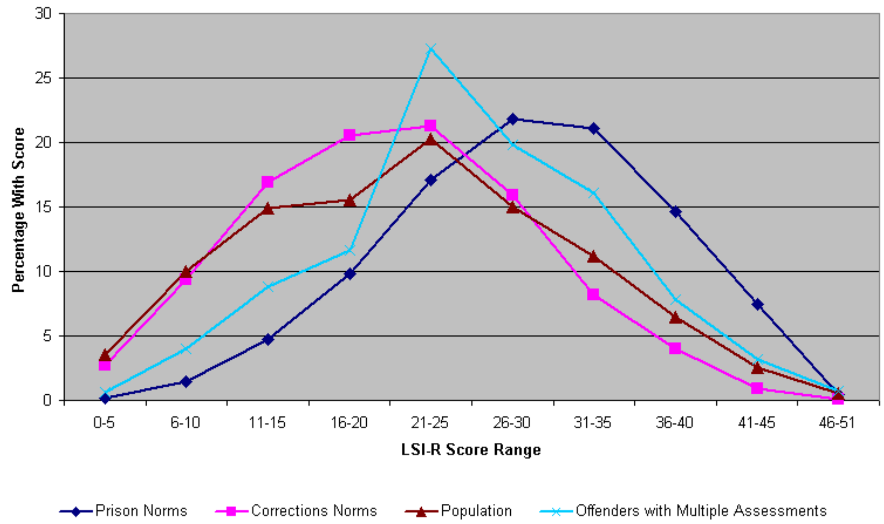
The LSI records used in the present study were collected by probation officers in a Midwestern county community corrections department from 2002 to 2006. LSI assessments were done on 4919 separate offenders, 1833 with a second assessment, 962 with a third assessment, etc., and 12 offenders with eight LSI assessments. The shrinking sample was of some concern, as Cook and Campbell (1979) had noted that selection bias might cause problems with the generalizability of results found in a biased sample. The samples were approximately 80% white and 80% male, with a mean age of about 34 years (S.D.=10), and a mean initial LSI score of approximately 25 (S.D.=8).

There were several known reasons for the shrinkage in the number of offenders with subsequent assessments. The records were collected over several years, and offenders who were assessed later in the collection cycle did not have as many reassessments. The county, adhering to the risk principle, generally only reassessed more serious offenders, leading to a lower number of low risk offenders. About half the low risk offenders who were assessed again had come back into the assessment stream because of a new violation. The other half were apparently given closer scrutiny for some other reason, perhaps because of violence or sexual offending. Another reason for a loss of offenders was a subsequent arrest or probation violation.

The risk score distributions of the offenders with one LSI, and those with multiple assessments, were compared with the LSI national norms score distributions for offenders in community corrections and prison (Andrews, & Bonta, 2003). (See Figure 1.) The national male and female norms were combined at an 80/20 ratio for male and female offenders to match the ratio of males and females in the current population of offenders. The total offender population's LSI distribution was similar to the score distributions of the national corrections norm sample, and the risk level of the sample of offenders with multiple LSI assessments was about midway between a normal corrections population and a normal prison population.

Figure 1

Comparison with National Corrections and Prison Risk Norms of Offender Population (N=4919) and the Sample of Offenders with Multiple Assessments (N=1833)



Design and Materials

Data Variables. The records used in this study were provided in a flat file format with one record for each LSI assessment. Each record had the names and birthdates of offenders, which were matched to two separate databases, a court services database that tracked probation violations that resulted in a prison commitment, and a BCA database containing arrest and conviction records. The initial LSI was done after arrest, and before conviction, which allowed the timing of arrests occurring before the LSI to be differentiated from those after the LSI. The BCA database also contained demographic data such as age, gender, and race. In total, 97% of the LSI-R records were matched to BCA records.

The independent variables used in this study were continuous (Age at first assessment, LSI-R score, Assessment Number, Days Between Assessments, and Score Change), and dichotomous (Male, White, Switched Raters). The dependent variable was a dichotomous variable called Any Violation (AV) that was a composite of arrest leading to conviction and/or probation violation leading to a prison commitment.

Procedure

The initial LSI database was exported to an SQL database and manipulated with Microsoft Visual Basic. Additional fields were added to the LSI-R SQL table as needed and populated with arrest and conviction data provided by the BCA, probation violation data provided by court services, or computations from other fields. AV was calculated by determining whether any further arrests or probation violations occurred in the year after the first conviction. Arnold (2007) discussed the process in more detail. It should be noted that a slightly different procedure was used to calculate the AV variable in this study. Arnold used the completion date of the LSI to calculate AV, and this study used the creation date, since some offenders were arrested after the LSI creation date, but before the completion date. One of the more useful data manipulations done was to flatten the data still further by putting all of the LSI scores and change scores in each record, which then could be selected by Assessment Number. After the data was manipulated it was exported to SPSS. The calculations done in this study were done in SPSS or Excel. The OLS calculations were done in accordance with the methods proposed by Singer and Willett (2003) for analyzing individual growth curves.

RESULTS AND DISCUSSION

The Dynamic Predictive Validity of the LSI and the Temporal Stability of Risk

The correlation rates were calculated between LSI scores and violation by one year after each LSI for LSI 1,2,3, & 4 (Mean interval = approx. 7 mo., S.D.= 3 mo.). The results are shown in Figure 1 (all $p < .001$). Reading across the rows, it is clear that subsequent LSI scores have consistently higher predictive validity coefficients than LSI 1. This is similar to the previous results found by others. Reading down the columns, it is clear that the LSI scores do not degrade with time, as would be predicted if the risk levels were changing over time. The LSI scores not only don't get worse, they appear to improve over time in some cases. At the very least, they are equivalent. This indicates that the underlying propensity for rule violation must have a high degree of temporal stability or, if risk levels are unstable, they are unstable in a nonlinear fashion.

Figure 1.

Correlation Between LSI Score and Violation by One Year after each LSI Assessment

	LSI 1	LSI 2	LSI 3	LSI 4
Sample 1(N=3216)				
After LSI 1	.275			
Sample 2 (N=1182)				
After LSI 1	.196			
After LSI 2	.245	.338		
Sample 3 (N=614)				
After LSI 1	.220			
After LSI 2	.250	.323		
After LSI 3	.227	.313	.344	
Sample 4 (N=287)				
After LSI 1	.225			
After LSI 2	.241	.289		
After LSI 3	.209	.279	.274	
After LSI 4	.218	.338	.353	.347

Inter-Rater and Intra-Rater Effects on Score Changes between Assessments

In order to observe the effects of changing raters, and the score changes due to the learning curve, the LSI scores of offenders with four LSI assessments (N=510) were grouped by 5-point intervals of score change, centered on 0, and plotted in figure 2A. The changes in LSI scores are markedly different for LSI 1-2 than from LSI 2-3, and LSI 3-4. This process was repeated for offenders who switched raters from LSI 1-2 and kept the same rater form LSI 2-3 and LSI 3-4 (N=178), and the plots were placed in Figure 2B. Again, it is clear that LSI 1-2 shows a marked difference in scoring. When the same rater is kept on all three LSI assessments (N=144), as shown in Figure 2C, it is clear that all three distributions are much closer in shape. When Figures 2B and 2C are compared, the difference in scoring due to a switch in raters is observed. This, from a visual standpoint indicates that switching raters has an effect on scoring. Figure 2D shows the distributions when raters are switched on all three LSI assessments. A subsection of offenders who had the same rater from LSI 3-4 (N=58) show less change in score than any of the other sets of score changes between LSI assessments.

The intra-rater differences in the magnitude of scoring from LSI 1-2 compared with LSI 2-3, and LSI 3-4 can be seen in Figure 2C. There is no shift in central tendency, indicating that rating is consistent within the same raters, but the dispersion in the changes in scores is greater from LSI 1-2, compared with LSI 2-3, and LSI 3-4, supporting the learning curve hypothesis.

The consistency in the amount of change in approximately equal seven-month intervals (S.D. = 3 mo.) between LSI 2-3, and LSI 3-4, is remarkable. This indicates that the temporal instability of offender risk is constant over time. This would have important implications for practice, as the group level of change appears to be predictable, and could be used to set reassessment guidelines.

Figure 2.

Percentages of Offenders for each 5 Point Change in LSI Score for LSI 1-2, LSI 2-3, and LSI 3-4

Figure 2A
All Offenders (N=510)

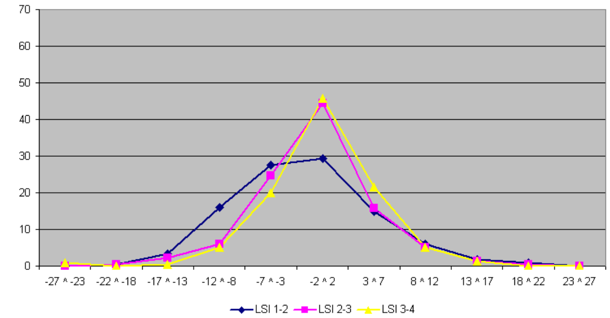


Figure 2B
Switched Raters LSI 1-2
Same Rater LSI 2-3, 3-4(N=178)

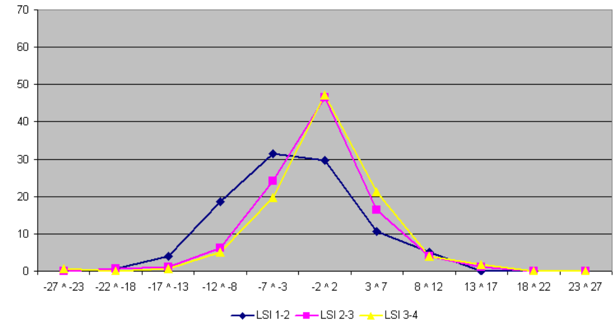


Figure 2C
Same Rater LSI 1-2, 2-3, 3-4 (N=144)

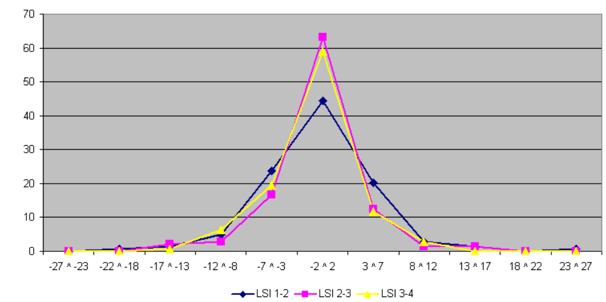
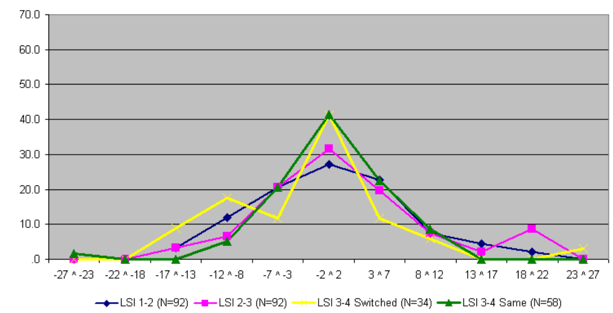


Figure 2D
Switched Raters LSI 1-2, 2-3 (N=92)
Switched Raters LSI 3-4 (N=34)
Same Rater LSI 3-4 (N=58)



A number of descriptive and analytical statistics were calculated for the groups of offenders shown in Figure 2, and placed in Table 2. The greater parts of these statistics are provided for reference. There are four sections, each with three columns and three sets of calculations. There are descriptive statistics for both the score changes, and the absolute values of the score changes. The absolute values of the score changes are provided because the almost perfect symmetry between increases and decreases in score changes renders the mean values of the score changes practically useless for determining the magnitude of change. A set of t tests was done between the columns to determine whether the means were the same.

In Column 2, note that when the rater is switched from LSI 1-2, both the mean score change, and mean absolute values of the score changes are significantly greater than when the rater is held constant. In the case of the raters held constant, there is no significant difference in scoring. Note that the correlation between LSI 1 and LSI 2, LSI 3, and LSI 4 is low when the rater is switched, but high between LSI 2, LSI 3, and LSI 4 when the rater is held constant. This suggests that some of the temporal instability of offender risk level for LSI 1-N is simply due to inter-rater instability.

In Column 3, note that when the rater is kept the same between assessments the mean level of change is not significantly different between LSI 1-2, LSI 2-3, or LSI 3-4. This indicates that the same rater rates offenders consistently. Note that the mean absolute level of score change is significantly different between LSI 1-2 and LSI 2-3, but not significantly different between LSI 2-3, and LSI 3-4. This supports the learning curve hypothesis that the rater is getting to know the offender better from LSI 1-2. There is no differential learning effect from LSI 2-3, or from LSI 3-4. The correlation rates indicate that the LSI 1 scores have a lower correlation rate with LSI 3, and LSI 4, suggesting that temporal instability of risk in this case may be due to an interaction effect between the rater and the offender, rather than an actual instability of offender risk.

In Column 4, when score changes are compared for rater switched LSI 1-2 and rater switched LSI 2-3, no significant differences are seen between the mean levels of real and absolute change. When this is contrasted with Column 2, the differences between switching raters and keeping raters the same are seen. The relative differences between the means and standard deviations support the contention that switching raters and the rater learning curve both add to score changes.

Table 2.

Score Change Statistics and Absolute Score Change Statistics and Inter-correlation Rates for Offenders with Four or More LSI Assessments

Sample	**Column 1 All**			**Column 2 Switch 1-2, Same 2-4**			**Column 3 Same Rater 1-4**			**Column 4 Switch 1-2, 2-3, Split 3-4**			
LSI Test/Retest	**1-2**	**2-3**	**3-4**	**1-2**	**2-3**	**3-4**	**1-2**	**2-3**	**3-4**	**1-2**	**2-3**	**3-4**	**3-4**
Switched				**Yes**	**No**	**No**	**No**	**No**	**No**	**Yes**	**Yes**	**Yes**	**No**
N	510	510	510	178	178	178	144	144	144	92	92	34	58
Score Change between Assessments													
Mean	-.56	-.29	-.11	-2.43	-.63	-.08	-.01	-.40	-.75	.35	1.64	-1.56	-.34
S.D.	6.589	5.821	5.798	6.33	5.09	5.08	5.51	4.16	4.12	7.41	8.01	8.37	6.21
Median	0	0	0	-3	0	0	0	0	-0.5	0	0	-2	0
Mode	0	0	0	-5	0	1	0	0	0	-1	0	-2	1
Min.	-18	-19	-29	-18	-18	-25	-18	-14	-17	-14	-14	-15	-29
Max.	27	25	27	15	16	15	27	14	10	20	21	27	11
Range	45	44	56	33	34	40	45	28	27	34	35	42	40
t		-.708	.493		-2.962	1.021		.664	-.726		-.794		.794
Sig.		.479	.622		.003	.308		.507	.469		.428		.429
Absolute Value of Score Change Between Assessments													
Mean	5.09	3.99	4.12	5.42	3.58	3.60	3.88	2.73	2.80	5.80	5.97	6.26	4.21
S.D.	4.32	4.10	4.39	3.98	3.40	3.49	4.02	3.13	3.05	4.58	5.55	5.66	4.54
Median	4	3	3	5	3	3	3	2	2	5	4	4.5	3
Mode	1	1	1	5	1	1	1	0	1	3	1	2	1
Min.	0	0	0	0	0	0	0	0	0	0	0	1	0
Max.	27	25	29	18	18	25	27	14	17	20	21	27	29
t		4.187	.181		4.669	.031		2.713	.191		-.057		-1.912
Sig.		.000	.856		.000	.975		.007	.849		.954		.059
Correlations Between LSI Scores[a]													
	LSI1	**LSI2**	**LSI3**	**LSI1**	**LSI2**	**LSI3**	**LSI1**	**LSI2**	**LSI3**	**LSI1**	**LSI2**	**LSI3**	**LSI3**
LSI2	.684			.703			.803			.504			
LSI3	.582	.763		.569	.814		.785	.881		.398	.395		
LSI4	.572	.709	.772	.572	.784	.816	.745	.829	.880	.355	.305**	.451**	.676

a: All p<.001 unless noted, ** = p<.01

Determining the Clinically Significant Level of Change

In order to determine the level of change needed to produce a clinically significant change, the Reliable Change Index (RCI) for the LSI was calculated using the correlation rate between the LSI 1 and LSI 2 scores for the 0-90 day interval, as this would most likely be a time period in which no real change occurred. Note that one of the requirements of the RCI is that the correlation between assessments needed to be calculated in a period of zero change. The formula for the RCI at a 95% confidence interval is RCI = 1.96 * SQRT(2) * SQRT(1- $r_{1\text{-}2}$) * S.D.#1. The correlation between assessments at 90 days was $r_{1\text{-}2}$= .859 (N=20). Using the standard deviation of 8.443 that was found for the LSI 1 score distribution, the RCI is 8.8. This indicates that the percentages of offenders who changed more than 8 points could be assumed to have a clinically significant change in LSI score.

Determining the Magnitude of the Effect of Score Changes on Predictive Accuracy

Since the clinically significant change concept has not been tested with the LSI, and it is not clear what score change is needed to create a significant difference in predictive accuracy, a logistic regression model was created using violation at one year as the dependent variable, and age, race and gender, as well as the LSI 2 risk level as controls. A set of dummy variables, coded a 0 or 1 to indicate the level of changes in scores, were created to determine the level of score change needed to achieve a significant level of improvement in predictive accuracy. Offenders with three LSI assessments and at least a one-year follow-up period after LSI 3 (N=615) were used for this test. The results in Table 6 indicate that a change of 9 or more points is required for the LSI to detect a change in the likelihood of recidivism. This test provides a direct confirmation of the accuracy of the RCI, as a cutoff score of 8 or less is non-significant.

The dummy score change levels used in Table 3 cover a range of scores because the small number of offenders at each level would not allow analysis with smaller categories. Another logistic regression test, not shown, was done to bracket the level of score change needed from LSI 1-2, and LSI 2-3 to significantly predict violation. It was found that a score change of 4 points between LSI 1-2 improved the accuracy of prediction after LSI 2. This smaller level of change required to improve predictive accuracy is consistent with the improved accuracy of LSI 2 scores over LSI 1 scores. The score change needed from LSI 2-3 to significantly increase the predictive accuracy of LSI 2 in the 1-year period after LSI 3 was 9 points.

These three tests each independently confirm that a score change of 9 points or more is needed before a change in LSI score reflects a measurable change in risk. The generalizability of this result is not clear, but this does provide some initial guidance for corrections officials.

Table 3.

Logistic Regression Model for Comparison of Score Change to Predict Violation in One-Year Compared with Reference Interval of No Score Change (N=615)

	N	B	S.E.	Wald	df	Sig.	Exp(B)	95.0% C.I.for EXP(B)	
								Lower	Upper
Age		-.010	.011	.974	1	.324	.990	.969	1.010
White		-.422	.249	2.880	1	.090	.656	.403	1.068
Male		.241	.284	.719	1	.396	1.272	.729	2.218
LSI 2 Score		.118	.016	56.879	1	.000	1.125	1.091	1.160
Score Change LSI 2 to Final									
Changed 9 or more down	59	-1.561	.544	8.243	1	.004	.210	.072	.609
Changed 5 to 8 down	88	-.715	.472	2.298	1	.130	.489	.194	1.233
Changed 1 to 4 down	186	-.678	.431	2.480	1	.115	.508	.218	1.180
Reference – No Change	59								
Changed 1 to 4 up	127	.073	.436	.028	1	.868	1.075	.458	2.526
Changed 5 to 8 up	52	.656	.485	1.827	1	.176	1.927	.744	4.990
Changed 9 or more up	44	1.605	.496	10.493	1	.001	4.980	1.885	13.156
Constant		-3.490	.791	19.445	1	.000	.031		

Measuring Between Group Changes in Risk Level to Determine Optimal Reassessment Period

In order to provide corrections officials an indication of the level of change in risk level to expect over time, the offenders with 3 or more assessments and the same rater (N=615) were split by amount of time between LSI 2 and the final LSI. The first five groups included offenders split by 6 month intervals through 30 months, and the sixth group included all offenders with more than 30 months between LSI 2 and the final LSI. ANOVA analysis of ages and Chi-Square analysis of race and gender revealed that the demographic characteristics of these sub-samples of offenders were all similar. The seventh column is included for comparison and consists of all offenders whose rater changed from LSI 2 to the Final LSI.

The results, shown in Table 4, indicate that there is a fairly high degree of temporal stability, as indicated by the correlation rates between the two sets of LSI scores, until after 30 months. A separate calculation using a linear regression model (not shown) indicated that the mean score changes for offenders in the 19-24, and 25-30 month test-retest periods were significantly different from the mean score change of offenders who had shorter (0-6 month) test-retest intervals, but the mean score change for offenders in the 31-56 month group were not significantly different from the mean score change of the offenders in the 0-6 month group. The only time period with a significant absolute difference in the level of change was the 31-56 month time period. Aside from a slight jump in the temporal stability of risk for the offenders with 25-30 months between assessments, it appears that the amount of accumulated change in risk increases in a linear fashion over time. One item of interest is the fact that there would appear to never be a time when there was no change in risk level. In the 0-6 month period, almost 10% of offenders had a clinically significant change, and less than 50% could be considered to not have changed at all because of a change in score of only –2 to 2 points.

In terms of temporal instability, it appears that changes in risk level are the rule, but the accumulation of changes in risk level over time is slow. Interpretation is guarded as this is a between group comparison, and the groups could vary in unseen ways.

Table 4.

Statistics for Second and Final LSI Scores, and Score Changes between Second and Final LSI, Grouped by Time between Assessments for Offenders with Same Rater at Both Assessments. Statistics for Offenders with a Switch in Raters between Assessments are Included.

	Months Between Assessments						Total Same Rater	Switched Raters
	0-6 Mo.	7-12 Mo.	13-18 Mo.	19-24 Mo.	25-30 Mo.	31-56 Mo.		
N	115	222	163	52	31	32	615	347
LSI 2								
Mean	22.7	23.3	23.6	23.3	26.5	23.7	23.4	26.4
S.D.	7.3	8.4	9.2	7.6	9.2	8.2	8.4	7.9
Final LSI								
Mean	20.8	22.0	22.5	23.3	26.4	22.4	22.3	26.5
S.D.	7.9	8.9	9.3	8.0	9.5	8.6	8.8	8.2
Score Change LSI 2-Final								
Mean	-1.91	-1.32	-1.10	-.06	-.13	-1.34	-1.21	.06
S.D.	4.74	4.61	5.43	5.91	5.85	6.51	5.15	8.36
Minimum	-18	-17	-17	-16	-16	-15	-18	-24
Maximum	15	13	13	10	10	11	15	27
Absolute Score Change LSI 2-Final								
Mean	3.69	3.46	4.04	4.56	4.45	5.47	3.90	6.45
S.D.	3.53	3.31	3.78	3.70	3.70	3.65	3.57	5.31
Range	18	17	21	16	16	14	21	27
r (p<.001)	.758	.765	.814	.661	.788	.491	.755	.541
Percentages of Offenders in Each 5-Point Score Change Interval								
-27 thru -23								.3
-22 thru -18	.9		.6				.3	.9
-17 thru -13	.9	1.4	2.5	3.8	3.2	3.1	2.0	3.5
-12 thru -8	10.4	9.0	9.2	3.8	9.7	15.6	9.3	11.8
-7 thru -3	27.0	23.4	23.3	26.9	19.4	28.1	24.4	22.2
-2 thru 2	48.7	51.4	42.3	34.6	35.5	25.0	44.9	26.8
3 thru 7	9.6	12.2	17.8	17.3	29.0	18.8	14.8	16.1
8 thru 12	1.7	1.8	3.7	13.5	3.2	9.4	3.7	10.7
13 thru 17	.9	.9	.6				.7	3.5
18 thru 22								3.5
23 thru 27								.9
% Clinically Sig. (>= 9)	9.6	8.6	14.1	15.4	12.9	25.0	11.9	29.4

Measuring Within-Group Changes in Risk Level to Determine Optimal Reassessment Period

The previous analyses measured between group differences in risk level. It was necessary to use that method because the data that measured risk for a long enough period to show a substantial change was sparse. The following analysis will look at within-group differences in risk level over time, and thus provides its own control group.

The LSI scores distributions for LSI 1, 2, & 3 for all offenders with three assessments (N=965) shown in Figure 3A reveals a general pattern over time of a gradual decrease in risk level (Mean LSI (S.D.) = 25.9(8.5), 24.6(8.3), and 23.9(8.8).) A somewhat similar pattern of gradually decreasing risk can be observed in Figure 3B for LSI 2, 3, 4 and 5 for all offenders with 5 assessments (N=145). (Mean LSI (S.D.) = 22.9(8.1), 22.5(8.2), 22.1(8.0), and 22.4(8.7).) As can be seen from the mean LSI scores, the decrease in risk for the smaller sample was not completely linear.

Note that the change score distribution shown in figure 3C indicates an almost identical level of change in each approximately seven-month period. When all of the possible change score distributions for longer periods LSI 2-4, LSI 3-5, and LSI 3-5 are plotted with the change scores of LSI 2-3 as a reference, the intermediate periods LSI 2-4, and LSI 3-5 fall within the plots for LSI 2-4, and LSI 3-5. The plots for LSI 2-4 and LSI 3-5 are not identical, which could indicate possible problems with a stability of change hypothesis, or another possible interpretation is that the difference could be due to or a change in scoring method in the LSI 2-4 time period as indicated by the shift to the left for that distribution.

Figure 3E is provided to give some idea of the intra-individual level of change, and Figure 3F shows the ranges of score change over 5 assessments. This set of graphs indicates that there is a large amount of individual change, but a very small group level change in risk.

Figure 3A – LSI 1,2, & 3 Score Distributions for All Offenders with 3 Assessments. (N=965)

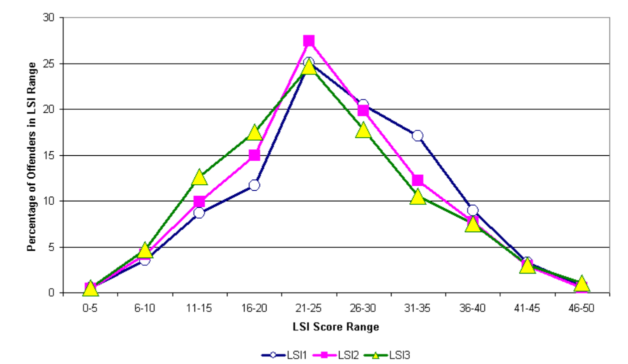


Figure 3B – LSI 2,3,4, & 5 Score Distributions for Offenders with 5 Assessments*. (N-145)

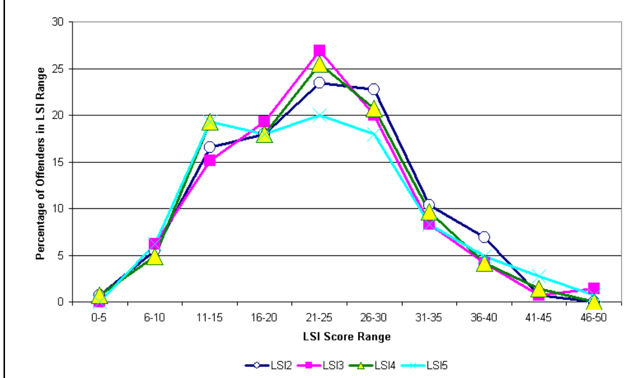


Figure 3C – Offenders in 3 Point LSI Score Change Groups, LSI 2-3, 3-4, 4-5. (N=145)

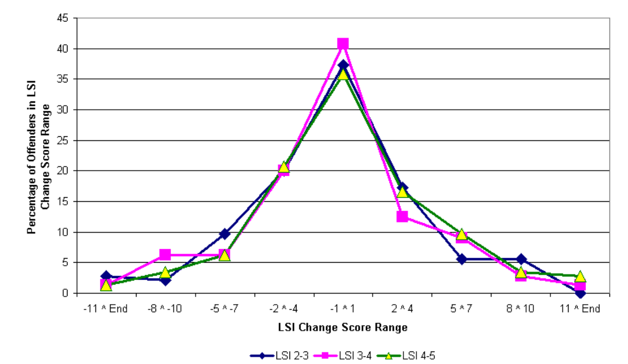


Figure 3D – Offenders in 3 Point LSI Score Change Groups, LSI 2-3, 2-4, 2-5. (N=145)

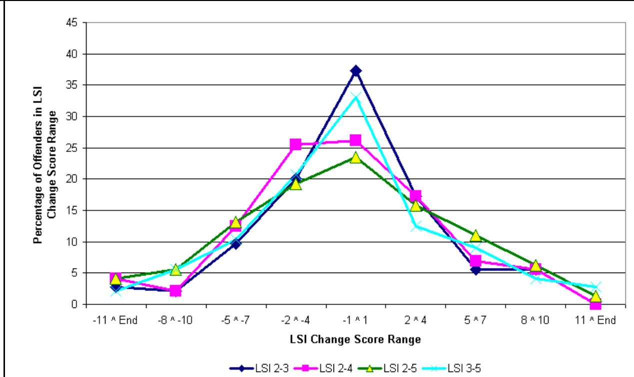


Figure 3E – Score Trajectories for Offenders with 5 Assessments* (N=145)

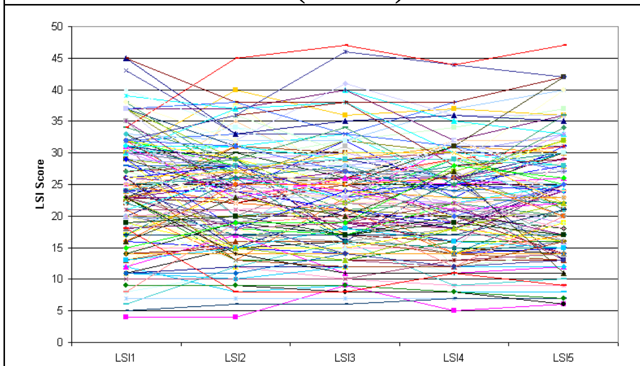
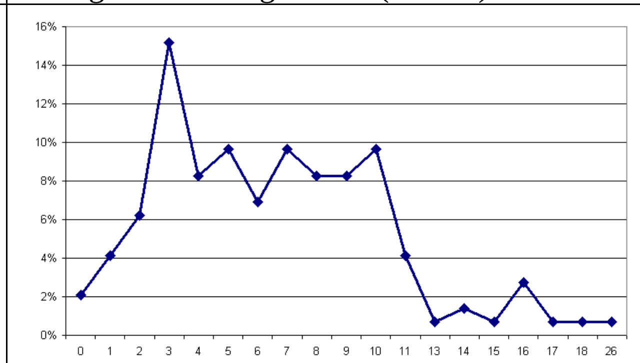


Figure 3F – Percentage of Offenders at Each Change Score Range Level (N=145)



* Same rater for all 5 assessments.

The change score statistics for the offenders with five LSI assessments were calculated and placed in Table 5. Note that the absolute level of score change is almost identical for equal time periods, and increases only slowly over time. This pattern is reflected in the percentages of offenders with a clinically significant change as well.

Table 5.

Statistics for Score Changes, Absolute Score Changes, Levels of Clinically Significant Change, Between LSI 2-5 for Offenders with 5 or more LSI Assessments and Same Rater (N=145)

	LSI 2-3	LSI 3-4	LSI 4-5	LSI 2-4	LSI 3-5	LSI 2-5
Mean Score Change						
Mean	-.37	-.41	.26	-.78	-.16	-.52
Median	0	0	0	-1	-1	-1
Mode	0	0	0	-3	-1	-3
Std. Deviation	4.5	4.5	4.6	4.8	5.6	5.7
Range	28	31	30	25	41	34
Minimum	-18	-17	-16	-16	-15	-16
Maximum	10	14	14	9	26	18
Absolute Score Change						
Mean	3.1	3.2	3.3	3.7	3.9	4.5
Median	2	2	3	3	3	4
Mode	0	1	1	3	1	1
Std. Deviation	3.2	3.2	3.2	3.1	3.9	3.6
Range	18	17	16	16	26	18
% >8	3.4	3.4	4.8	2.8	4.8	5.5
% <8	4.8	4.8	3.4	5.5	6.2	8.3
-8 < % > 8	8.3	8.3	8.3	8.3	11.0	13.8
Months between Assessments						
Mean	6.5	6.5	6.9	13.0	13.4	19.9
S.D.	1.7	1.8	3.1	2.6	3.8	4.2

Offender Change as Power Law Distribution

The mean score change for a rank ordered distribution with 5 offenders in each group was plotted for all offenders with five or more assessments and the same rater for each assessment. The plot is shown in Figure 4A. The plot is fairly linear with a small subset of the offenders on the right having a significantly higher rate of change. This is suggestive of a power law distribution. Power law distributions were first noted by Simon (1955) and are found in a number of natural phenomena Clauset et. al. (2007). The natural logs of the values plotted in Figure 6A were calculated and placed in Figure 6B to test the power law hypothesis. A linear trend line was plotted through the points y = 0.0673x + 0.866. The logged plot is somewhat linear suggesting that the rate of offender change could follow a power law distribution. This would need to be verified with much more data than is present in this sample. One problem with this analysis is that the LSI is not a completely linear instrument, so nonlinear jumps in scoring can occur that do not reflect jumps in risk.

Figure 6A: Plot of Offender Change Score Distribution – Each Point = Mean Score Change for 5 Offenders

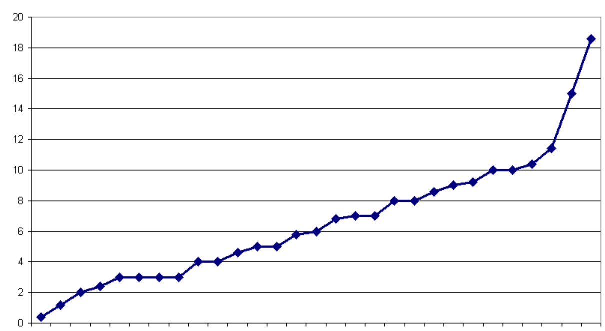
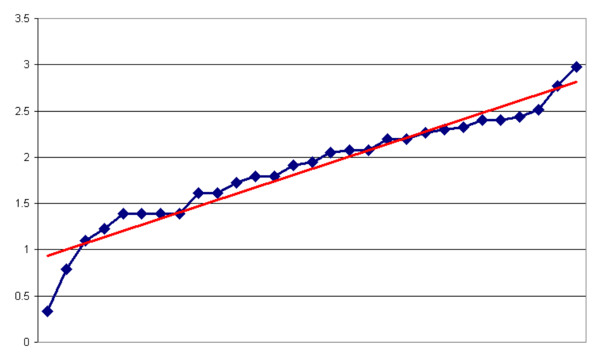


Figure 6A: Log Plot of Offender Change Score Distribution – Each Point = Log (Mean Score Change) for 5 Offenders (y = .0673x + .866)



The Intra-Individual Permanency of Score Change

To determine whether a clinically significant change was permanent, the mean score changes before and after a clinically significant positive or negative change in score were calculated. This was done for offenders with any rater from LSI 1-4 (N=511), and with the same rater from LSI 2-5 (N=145). The results are shown in Table 7. As can be seen in the table, a clinically significant increase in score was both preceded and followed by a mean decrease in score. In a similar fashion, a clinically significant decrease in score was preceded and followed by a mean increase in score. The net average change in either direction was much smaller than the amount needed to obtain a clinically significant change, suggesting that changes in score are generally non-linear and insignificant in the medium term (21 months).

Table 7.

Mean Score Change by Wave of Assessment for Offenders with a Clinically Significant Change in Score for Middle Wave of Assessments (Same Rater)

	> 8 Point Increase LSI 2-3 (N=31)		> 8 Point Decrease LSI 2-3 (N=30)	
	Mean	**S.D.**	**Mean**	**S.D.**
Total N=511				
Score Change LSI 1-2	-4.5	5.5	2.1	8.7
Score Change LSI 2-3	13.6	4.6	-11.9	2.5
Score Change LSI 3-4	-5.8	8.3	4.0	8.0
Net Change LSI 1-4	3.3	7.8	-5.7	10.9
	> 8 Point Increase LSI 3-4 (N=5)		**> 8 Point Decrease LSI 3-4 (N=7)**	
Total N=145				
Score Change LSI 2-3	-7.2	8.0	4.3	5.5
Score Change LSI 3-4	10.6	2.1	-10.7	2.9
Score Change LSI 4-5	-2.4	9.1	3.1	4.7
Net Change LSI 2-5	1.0	9.8	-3.3	8.8

The percentages of offenders who had a 9-point or greater score change, either up or down, from LSI 2-3 were calculated for all offenders with 5 assessments and the same rater. The percentages of offenders who had a 9-point change from LSI 2-3 and kept it from LSI 2-4 and also the percentage of offenders who had a 9-ponit change on LSI 2-3 and kept it from LSI 2-4 and LSI 2-5 were calculated and placed in Table 8.   No offender with a 9-point increase from LSI 2-3 stayed above 9 points on LSI 4. Of the 4.8% of offenders who had a 9-point decrease from LSI 2-3, less than 1/2, or 2.1% of the total maintained their decrease on LSI 4, and only about 1/4, or 1.4% of the total kept a 9-point decrease from LSI 2-3 all the way to LSI 5.

The total number of offenders with any 9-point change went from 8.3% on LSI 2-3, then dropped to 1/4 of that total or 2.1% on LSI 4, and then dropped a little more than 1/8, or 1.4% of the total number of offenders on LSI 5. These results indicate that very few offenders change more than 9-points, and the majority that do change 9 or more points change back on the next assessment. Only 1.4 % of the total number of offenders maintained a 9-point change for the entire 21 months.

Table 8.

Percent of Offenders with a Clinically Significant Change is Score from LSI 1-2
Clinically Significant Change Measured Again LSI 2-3-4 and LSI 2-3-4-5 (N=145)

	LSI 2-3	**LSI 2-3-4**	**LSI 2-3-4-5**
9-Point Change – Same Rater (N=145)			
% Up	3.4	0.0	0.0
% Down	4.8	2.1	1.4
% Total	8.3	2.1	1.4
Months	6.5	13.0	19.9

Determining Whether Intra-Individual Change is Linear

To determine whether the change score distributions were linear of non-linear, the methods for calculating growth curves suggested by Singer and Willett (2003) were used to calculate OLS regression equations for all offenders with 4 assessments and the same rater on all assessments (N=510). A t-test was used to determine whether the fitted OLS line slope was nonzero and a significantly linear match to the data. Two set of OLS regression equations were calculated. The first used data from LSI 1-4, and the second from LSI 2-4. Two calculations were run because the OLS regression assumes homoscedasticity in variances (Blalock, 1979; p.389) and it is clear that LSI 1 varies in a significant way from the other scores. There were 131 (25.6%) offenders with a significantly linear trajectory and nonzero slope when LSI 1 was used, and 135 (26.4%; 8.4% increasing; 18.0% decreasing) when LSI 1 was left out. This meant that in 73.6% of the cases, a case for non-linear non-zero change could not be made. From this analysis, it appears than linear changes in offender risk are the exception, rather than the rule.

The score distributions for all individual offenders with 8 data points and any rater are plotted in Figures 5.1-12. As can be seen by the plots, there is a significant heterogeneity to the distributions. Some show an almost completely linear trend, and others have significant fluctuations in score. Many of the plots appear to indicate that there is some periodicity to the fluctuations. Attempts to fit polynomial equations to the data with coefficients up to x^6 (not shown) failed to produce a match to the plots. From a temporal stability perspective, there does appear to be some amount of non-linear fluctuation in risk, but the fluctuation is bounded about a central tendency. The fact that some offender are showing linear trends to their risk scores, some increasing, and some decreasing, indicates that there are longer term shifts in risk that will lead to changes in the rank order of offender risk distributions over time.

Figure 5.1: Sig. Dec. 1-8, & 2-8

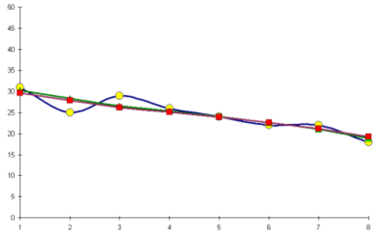


Figure 5.2: Sig. Dec. 1-8, & 2-8

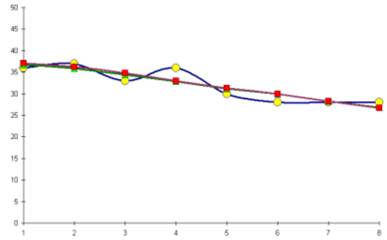


Figure 5.3: NS OLS Trend

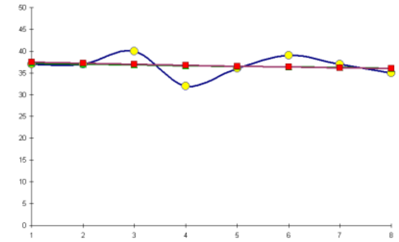


Figure 5.4: Sig. Dec. 2-8 Only

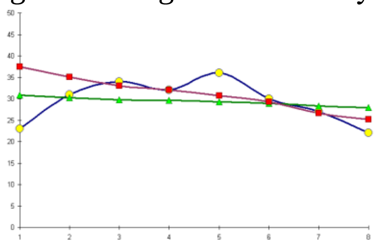


Figure 5.5: NS OLS Trend

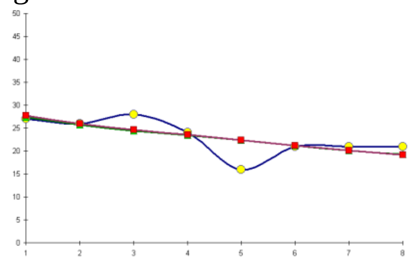


Figure 5.6: NS OLS Trend

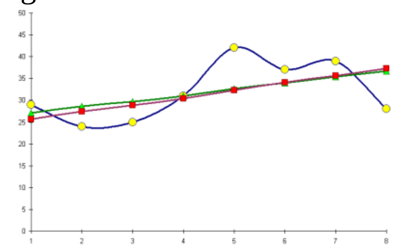


Figure 5.7: Sig. Inc. 2-8 Only

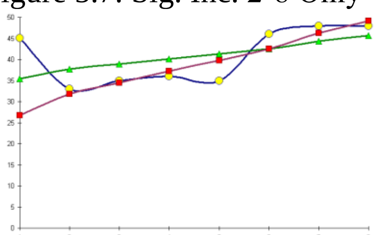


Figure 5.8: Sig. Dec. 1-8, & 2-8

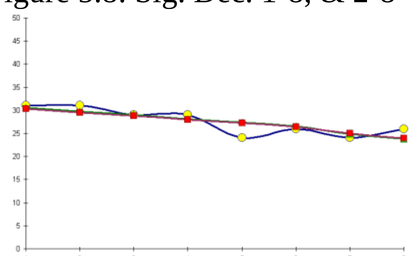


Figure 5.9: NS OLS Trend

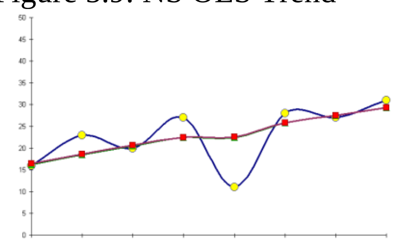


Figure 5.10: NS OLS Trend

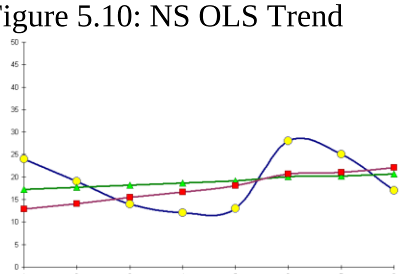


Figure 5.11: NS OLS Trend

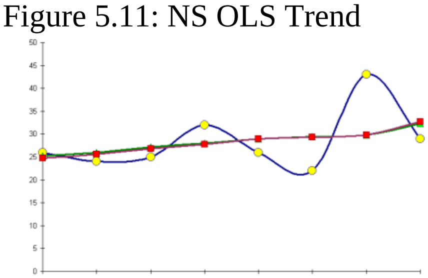
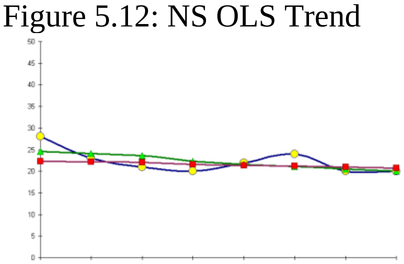


Figure 5.12: NS OLS Trend



LSI-R Scores OLS Trajectory 1-F OLS Trajectory 2-F

The offenders with one year after their final assessment were selected and the significance of the OLS trajectory as a linear increase or decrease was used to predict violation at one year, using the LSI 2 score, Age, Race, and Gender, and a rater switch as controls. Offenders who had a significant linear increase were no more likely to recidivate due to the increase, but offenders who had a significant linear decrease were less likely to recidivate at the p<.1 level. The same pattern was found when a 9-pont change up or down was used to predict violation. These models show that decreases in score predict recidivism, but increases don't.

Table 9.

Logistic Regression Models Predicting Violation at One Year by Status of Score Change, Significant OLS Up and Down and Size of Score Change 9 or more Point Up or Down

	Model 1	**Model 2**	**Model 3**	**Model 4**
LSI 2	.105***	.097***	.105***	.130***
	(.028)	(.029)	(.028)	(.032)
Client Age	-.048*	-.046[a]	-.048*	-.049*
	(.024)	(.025)	(.024)	(.024)
White	-.556	-.607	-.531	-.404
	(.523)	(.533)	(.518)	(.532)
Male	.021	-.163	.008	-.181
	(.579)	(.597)	(.581)	(.582)
Switched raters 2-F	1.197**	1.140*	1.194*	1.206**
	(.471)	(.474)	(.477)	(.476)
Sig. OLS UP (N=11)	.364			
	(.919)			
Sig. OLS DOWN (N=34)		-1.790[a]		
		(1.073)		
Changed 9 Up (N=5)			-.178	
			(1.247)	
Changed 9 Down (N=13)				-2.437*
				(1.162)
Constant	-2.593[a]	-2.043	-2.557[a]	-3.012*
	(1.414)	(1.493)	(1.407)	(1.477)
-2 Log likelihood	135.206	131.048	135.336	128.823
Cox & Snell R Square	.175	.193	.174	.203
Nagelkerke R Square	.292	.322	.291	.338

a p<.10, * p<.05, ** p<.01, *** p< .001

CONCLUSIONS

Several tentative conclusions can be drawn from the preceding analyses, with any permanent conclusions being held in reserve until the results of these analyses are replicated. The tentative conclusions can be broken down into general conclusions, which have to do with measurement issues in a test-retest environment, and sample specific conclusions, which are limited to the non-random groups of offenders that were used in these analyses.

The general conclusions are that there is a learning curve for raters who use interview based assessment instruments, and that the first assessment is generally less accurate than subsequent assessments. There is a larger score change when raters are switched between assessments, and score changes will be difficult to interpret when the same rater is not used on both assessments. A general conclusion could also be drawn that there is some non-zero level of score change that cannot be assumed to be significant enough to make a difference in the outcome of interest, in this case, violation rate. The cutoff level, called a Reliable Change Index (RCI), can be determined on a sample-by-sample basis until a general cutoff score is found.

The sample specific conclusions are that offender risk levels are stable enough so that the LSI is not a better predictor of risk in the 21-33 month period than in the 0-12 month period. Changes in LSI score need to be 9-points or greater after LSI 2, when the same rater is used, in order for the change in LSI score to indicate a significant change in arrest rate. Only about 9.0% of offenders have a 9-point or greater change in score in a one-year period. This change is generally preceded and followed by a smaller change in score in the opposite direction, resulting in a less than 9-point change across three assessments. Of the approximately 9.0% of offenders who do have a 9-point change in score, none of the offenders who have a 9-point increase maintain the 9-point increase in score until the next assessment, and only 1/2 of the offenders with a 9-point decrease, or 1/4 of the 9% of offenders with a 9-point change maintain that change over the next seven months. This zig-zag pattern results in a net total of about 15% of offenders with a 9-point change in the 1-2 year time period and about 20% with a net change in the 2-3 year time period. At no time is there ever a period of no significant change.

When linear trends are plotted over periods of 2-4 years, 18.0% of offenders have a significant linear decrease in LSI scores over time, and 8.4% have a significant linear increase in scores. The rest of the offenders, totaling 73.6% had no significant linear change in score. A visual inspection indicated that the offender risk levels seemed to fluctuate in a nonlinear cyclic fashion over time. A plot of the total range of score changes indicated that there was a peak at the 3-point score change level, followed by an almost equal number of offenders (10% each) at the 4-10 point change levels, and a few offenders with very large levels of change. When the natural log of the offender distribution at each level of score change was plotted, it appeared to follow a power law distribution.

The rank order of the offender risk levels is maintained at a fairly high level of approximately r = .70-.80 until the 31-56 month time period, when it drops to r=.50. The percentage of offenders with a 9-point change in score jumps to 25% in this time period. These two measurements indicate that rank order changes in score do occur over longer periods of time. This result tends to be contrary to Gottfredson and Hirschi's (1990) stability of self-control thesis, and supportive, in principle, of Sampson and Laub's (1993) thesis that events can change the level of criminal propensity. There is not enough information to indicate the cause the rank order changes in risk level at this point.

Following a more in-depth review of each of the analyses done, the general implications from these results will be broken down into practice, research, and theoretical domains, and discussed in-depth.

The Dynamic Predictive Validity of the LSI and Temporal Stability of Risk

The test to determine whether offender risk levels are temporally unstable shows that the previous LSI scores did not become less accurate over time as would be suggested by a linear inertial model of change. This suggests that risk levels were either stable in the 21 months from LSI 1 to LSI 4, or the changes in risk were nonlinear in nature. The dynamic predictive accuracy of the LSI improved substantially from LSI 1 to LSI 2, indicating that a reassessment at 6 months might be a good idea if a more accurate risk score was desired. This analysis provided insufficient evidence, by itself, to indicate what a good reassessment period would be, but the results indicate that 7-21 months was too short of a time period, since the LSI 1 scores predicted recidivism equally as well over the 0-12 month period as the 21-33 month period.

Effects of Changing Raters and Learning Curve on Score Changes between Assessments

Both changing raters between assessments, and the effects of the learning curve from LSI 1 to LSI 2 have an effect on the level of score change between assessments. There is an approximately 2 point higher mean change in both score and absolute score when changing raters vs. keeping the same rater. This tends to skew results when trying to interpret changes in scores. There is about a 1-point higher absolute mean change in score from LSI 1-2 than from LSI 2-3, or LSI 3-4, when the rater is kept the same between assessments. Given the significant increase in predictive validity that occurs from LSI 1-2, it is probably be safe to assume that the increase in score change is due to a learning curve effect.

For day to day applications of practice, the change in score may not be enough to be concerned with, given the 9-point score change required to indicate a change in risk level, but for research applications dealing with changes in risk level over time, it appears that the lower accuracy for the first LSI would make it a poor candidate to draw conclusions from. These two results suggest that additional research is needed to determine whether studies examining changes in behavior should keep the same rater between assessments and throw out the first assessment due to problems with reduced accuracy.

It should be noted that many of the subsequent analyses done in this study would have been impossible if different raters were used or LSI 1 scores were included. Attempts had been made for 18 months to analyze changes in risk level using scores with different raters on each assessment and including the LSI 1 scores, and analyses kept producing inconsistent results. It was only when the rater was held constant and the LSI 1 scores were dropped that the true nature of change could be assessed.

The Precision of Offender Risk Assessment Instruments

To significantly increase predictive accuracy, a score change of 9-points or greater was needed. This result only pertained to the periods after the initial learning curve from LSI 1-2. The change in the LSI score needed to significantly predict a change in recidivism rates from LSI 1-2 was only 4 points. This could be due to the higher accuracy of the LSI 2 scores over the LSI 1 scores. Using classic test theory, any score can be thought of as the sum of a true score and some amount of error (Nunnally, & Bernstein, 1994; p. 224). A change in score from LSI 1 to LSI 2 is probably partly due to a reduction in the error coefficient, and it therefore takes less of a change in score to make a difference in the accuracy level.

The implications of the clinically significant change results for practice could be significant. Knowing the level of change required to predict a change in risk could provide guidance for corrections officials trying to interpret score changes. Small changes in score can be assumed to not predict a significant likelihood that risk has changed. These results could provide a more rational method for determining reassessment periods. It probably does not make sense to reassess someone who does not appear to be making significant changes in their lives. Some caution is needed before acting on these results however; as these results must be replicated before any policy changes are made.

Measuring Changes in Risk Level to Determine Optimal Reassessment Period

The results from the temporal stability study are mixed. On the one hand, it does not appear that there is ever a time where there are no changes in risk level. About 8-9% of offenders in any 7-month interval had a 9-point change in risk. This indicates that a small percentage of the offenders have highly labile risk trajectories. The total percentage with a clinically significant change did not appear to climb in a steady fashion, however, since the overall level of significant change only went up about 5% every 14 months. The rate of clinically significant change (9+ points) reached the 25% level at the 31-56 month period, but that period was so broad that it is difficult to determine how the rate increased over time.

When the mean risk levels were examined over three assessments for the offenders who had a 9-point change in LSI score between assessments, it appears that fluctuations before and after the change reduced the overall level of change considerably. Offenders who had a 9-point change in LSI score typically had smaller score changes in the opposite direction on the LSI assessments before and after the change, resulting in a mean shift in score that was not significantly different in risk from the beginning of the cycle.

These results indicate that reassessment with the LSI might simply capture short-term fluctuations in risk for some offenders. It appears that, for a subset of the offenders, keeping track of their risk level could be like trying to herd cats. There was a much larger group of offenders with scores that were much more stable. Most of those offenders never had a score change of over 9-points in the 21 months of observation. It doesn't appear that much would be gained from continually reassessing offenders who aren't changing, as the precision of the LSI is less than the change in risk. These results indicate that there is a considerable degree of inertia in the long term, but that there is also cyclic nonlinear change for a small percentage of offenders in the short to intermediate periods.

A number of the analyses indicate that the population level of changes in risk may be constant over time for any particular group of offenders. The levels of change appeared to be distributed in some sort of logarithmic fashion in the group of offenders, with a linear trend to the ranges of the change scores from 4-10 points, and a small subset of offenders exhibiting large changes in risk level. If the distribution of change levels is logarithmic, offender change may follow a power law. Per Bak (1996) investigated the properties of power law systems, and suggested that the power law phenomena are due to a property called self-organized criticality. In power law systems, there are many small events, and then, when conditions are right, a large event will occur. Many rather unpredictable phenomena, such as earthquakes and stock market crashes follow a power law distribution. If offender change is related to these classes of events, it could help explain some of the difficulties in predicting recidivism.

Determining Whether Intra-Individual Change is Linear

Some of the offenders (26.4%) were experiencing linear changes in risk over time. About twice as many offenders were found to be decreasing in risk (18.0%) as increasing (8.4%). The majority of offender's risk levels (73.6%) did not change in a significant linear fashion. From observing the risk trajectories and the OLS regression lines together, it appeared that there is a range of fluctuation in risk levels that appears to be centered on a more slowly changing average level of risk over time. The appearance of the change score plots indicates an almost universal cyclic sinusoidal trajectory to changes in risk that is similar to that predicted by Kelly and McGrath for some types of changes in human behavior. If offender risk levels follow a nonlinear trajectory, the null results from many of the offender treatment studies done to date would make more sense. Many studies use a single posttest period of a few months to a year to assess results. The sinusoidal patterns appeared to fluctuate through that time period, and so any results found in a one-year period would be ambiguous. Kelly and McGrath indicate that special care must be used when studying phenomena that fluctuate over time. As a thought experiment, they suggest the reader imagine what would happen if a treatment were done that set a nonlinear sinusoidal behavior pattern into motion. Researchers would get a positive result if behavior is assessed at the high peak, and a negative result behavior is assessed at the low peak. If individuals had different frequencies in their behavior cycles, the time of assessment could indicate a strong response, a negative response, or some intermediate level of response. Most research does not take the temporal properties of the behavior it is trying to assess into account as part of the study design, and therefore cannot measure change effectively. The results found in this analysis could have wide ranging implications for offender research. A much more complete analysis would be needed to determine the exact nature of these sinusoidal fluctuations.

Implications for Practice

From a practice standpoint, it appears that a second assessment with the LSI at about 6 months would be a good idea. The rater who is working with the offender should do the second, and all subsequent assessments if valid conclusions are to be drawn from any scores that are generated. If risk assessment is the goal, it does not appear that there is a significant benefit from reassessing with the LSI at any period under two, and possibly three years. Beyond the 2-3 year period, there would appear to be larger rank order differences in risk level and a reassessment would improve prediction. That isn't to say that some sort of short-term monitoring isn't needed. There was not a time period where there was no change in a small percentage of offenders. An exploratory analysis indicated that the areas that change quickly are prior offense, employment, drug and alcohol, living arrangements, and a few smaller areas. It may be advantageous to develop a ten-point checklist that could be used to detect those offenders who are changing significantly, and focus reassessment efforts on them. This would save the 1-2 hours of time that corrections officials are spending per reassessment.

When reassessments are done, practitioners can determine whether a significant change has occurred by calculating the RCI index as demonstrated in this article. This will allow practitioners to determine whether the change is significantly related to a change in risk. Practitioners should be aware that most changes in risk level are temporary in nature and are due to fluctuations in risk that occur naturally over time.

Implications for Research

Researchers should take the opposite approach than the one recommended for practitioners. A concerted effort should be made to determine the temporal stability for all of the individual items on the LSI as well as the total risk level at short periods of time, weekly or monthly, for a randomly selected sample of offenders over the space of several years. In this way, the period in which items change can be found, and less volatile items can be reassessed at longer time periods, and more volatile items can be assessed at shorter time periods.

Researchers should try to determine why the risk levels tend to follow a cyclic nonlinear pattern. The implications of this pattern for research on treatments designed to reduce the overall level of risk would appear to be extremely significant. Kelly and McGrath indicate that special care must be used when studying phenomena that fluctuate over time. As a thought experiment, they suggest the reader imagine what would happen if a treatment were done that set a nonlinear sinusoidal behavior pattern into motion. Researchers would get a positive result if behavior is assessed at the high peak, and a negative result behavior is assessed at the low peak. If individuals had different frequencies in their behavior cycles, the time of assessment could indicate a strong response, a negative response, or some intermediate level of response. Most research does not take the temporal properties of the behavior it is trying to assess into account as part of the study design, and therefore cannot measure change effectively. The results found in this analysis could have wide ranging implications for offender research. A much more complete analysis would be needed to determine the exact nature of these sinusoidal fluctuations.

From a research perspective, this study opens up the possibility for using risk scores to study the process of desistance (Bushway, et. al., 2001). Maruna (2001) suggested that it is difficult in desistance research to study the absence of something (i.e. crime). If a risk level approach is taken in the study of desistance, longitudinal sets of risk scores could be analyzed to determine the factors that lead to increased socialization. Some of the risk trajectories followed a fluctuating pattern that Piquero (2004) and Burnett (2004) have suggested was involved in the process of desistance. The further analysis of sets of risk scores might lead to an explanation for these fluctuating patterns.

From a practical standpoint for researchers, risk scores should be readily available. Hubbard, et, al. (2001) report that 75% of corrections departments in the U.S. use structured dynamic risk assessment for case classification. If the precautions recommended by this study are followed and researchers only use records with the same rater on each assessment, throw out the first assessment, and determine the level of clinically significant change for each instrument and population, risk scores could possibly be used for research purposes. Structured risk assessment instruments include a number of theoretically interesting variables. By doing sequence analysis, causal relationships (Kadzin, et. al., 1997) between the individual items might be determined. Many offenders do not get multiple assessments, and so issues with censoring could be a problem but, as Lee Robins (1978) pointed out,

> "The more the populations studied differ, the wider the historical areas they span; the more the details of the methods vary, the more convincing becomes that replication. Thus 2 imperfect studies that agree are more persuasive than a single very elegant study. A truth so powerful that it outs despite the errors and ignorance of investigators is a sturdy truth indeed!" (pp. 611-612)

The answers to questions regarding the temporal stability of risk would appear to be extremely important for the field of corrections. One problem that continually plagues researchers in criminology is the combination of both temporal stability and temporal instability in offender behavior. This study indicates that a more complex approach may be needed than the standard static/dynamic conceptualization with an assumption of linear changes in risk level.

Implications for Research - Study Design and Statistical Issues

There are a number of issues related to study design and statistical analysis of offender risk data that this study uncovered. From a study design perspective, Kelly and McGrath indicate that even controlled experiments may not be sufficient to analyze behavior that is fluctuating in a cyclical fashion over time (p. 47). They suggest that study designs have to use a pre and post-test measurement period using several measurements carefully spaced in time if any causal relationships are to be established. A cursory review of the offender treatment literature would indicate that this is not being done.

An issue related to homoscedasticity in variances (Blalock, 1979; p.389) was discussed briefly in the section on OLS regression modeling of individual growth curves. Many statistical analysis methods such as OLS regression require that the data meet the requirements for homoscedasticity, which means that the variances must be equal for all of the measurements. Using different raters between assessments or using the first assessment scores will violate these requirements because the variances are greater for scores generated in the first LSI or when raters are changed than for the other scores.

A statistical issue not previously discussed arises from the fact that the overall score distribution is almost stationary and the individual scores fluctuate. The differences between the group level and individual level patterns lead to problems with ergodicity. Molenaar (2008) provides a more complete discussion of the issues involved. In essence, the statistical analyses that are used to study non-ergodic phenomena must be compatible with this type of distribution.

The other restriction that flows from non-ergodicity is that group level results cannot be used to predict individual behavior and individual behavior cannot be used to predict group level results. Adherence to the principles of statistics regarding non-ergodic systems may provide more accurate analysis of offender behavior.

Implications for Theory

In this study, there were changes in both the intermediate term levels of risk, and the changes in the long-term levels of risk. Short-term fluctuations of various magnitudes were found for most offenders. Some of this fluctuation may have been due to measurement error, and some fluctuation due to real changes in risk. There does appear to be a slow dynamic to the risk scores that creates a rank order change in risk level in the 31-56 month time interval, as demonstrated by the drop in correlation rate between LSI 2 and LSI Final to r=.491 shown in Table 4. This is consistent with Sampson and Laub's (1993) contention that long term differences in socialization can cause shifts in risk level over the life course, and inconsistent with the Gottfredson and Hirschi position that socialization affects everyone at the same level as they age. In fairness to Gottfredson and Hirschi, these offenders were of all different ages, and they have suggested that rank order differences in propensity are only maintained over the life course for offenders of the same age. There was insufficient data to control for age in this study.

This slow dynamic patterns of change are consistent with studies that show that rank order of levels of self-control, as measured by instruments designed to measure self-control, change over time (Hay, & Forrest, 2006; Arneklev, et. al., 1998; Turner & Piquero, 2002; Mitchell, & MacKenzie, 2006; Burt, et. al., 2006; Winfree, et. al., 2006). The time periods and rates of change found in those studies match the levels of changes in the rank order risk levels found using the LSI score.

The implications of the similarity in long-term risk patterns measured using measures of self-control and using LSI scores are not clear, since it isn't known whether the LSI and the measures used in the stability of self-control studies measure the same construct. The results from a study done by Kroner et. al. (2005) suggest that one measure of criminal risk level works about the same as any other measure of criminal risk level. This raises some question about the make-up of the LSI, the various levels of temporal stability, and the relationships between them. Can the various static and dynamic components of the LSI be split apart, or are all of the items measuring a single underlying factor, criminal propensity?

Adding to the confusion, there is a puzzling phenomenon in the short term that is found in the fluctuation patterns shown in Figure 5.1-12. Why would the intermediate dynamic risk factors fluctuate in sinusoidal rhythmic patterns over the space of several years? This would appear to bear some relationship to the discussion of zig-zag patterns of offending that has been going on in the offender literature for some years.

The stopping and starting of criminal offending was noted by Glaser (1964: p. 85), who characterized the process of moving from crime to non-crime and then back to crime as a *zig-zag path* that "almost all" offenders go through. Matza (1964) wrote about adolescents drifting in and out of criminal activity. More recently, Sampson and Laub (2003) studied offenders who had a zig-zag patterns of offending and found that that those offenders appeared to show a fluctuating pattern because of heavy alcohol use.

One theory that would explain zig-zag offending due to alcohol or drug use use is provided by Prochaska and others (Prochaska, & DiClemente, 1983; Prochaska, et. al., 1992;

Prochaska & Velicer, 1997). Their theory, called the Trans-Theoretical Model (TTM) of behavior change, has found wide applicability in addiction research, and explicitly predicts nonlinear change. Other explanations for nonlinear change are found in the sciences of cybernetics (Ashby, 1956), and complexity (Bird, 2003).

The TTM developed by Prochaska and others suggests that change is the result of a decision, and people don't change their behavior unless they want to. This would appear to describe many of the offender patterns. When people do decide to change, it is because the reasons for and against changing have changed. The decision making process is modeled by a decisional balance approach (Janis, 1959; Janis, & Mann, 1977; Velicer, et. al., 1985), consisting of the weighing of reasons for and against the addictive behavior. Prochaska, et. al. (1994) suggest that the process of addiction cessation often involves a series of flip-flops in the decisional balance that results in a non-linear path to addiction cessation.

It appears that addiction cessation might be a variant of the approach-avoidance problems proposed by Lewin (1935; Myers, & Salt, 2007: p. 10). Lewin proposed that a decisional conflict occurs when we have to choose between two desirable options (+/+), two undesirable options (-/-), or an option with both desirable and undesirable attributes (-/+). These conflict types have since been renamed, approach-approach (+/+), avoidance-avoidance (-/-), and approach-avoidance (-/+) conflicts. If the offender desired to move toward a crime free lifestyle, but found aspects of the crime free lifestyle somewhat unpleasant, there would be an approach-avoidance conflict in that direction. If the offender simultaneously found aspects of a criminal behavior both desirable and undesirable, there would be an approach-avoidance conflict in that direction as well, and the offender would be caught in a double approach-avoidance conflict (Criminal behavior -/+ Offender -/+ Non-criminal behavior). The closer the offender came to a crime free lifestyle, the more salient the unpleasant aspects would become, and the more desirable the criminal lifestyle would seem. As the offender moved back toward a criminal lifestyle, the unpleasant aspects of the criminal lifestyle would become more salient, leading to a desire too move in the opposite direction. This would result in a fluctuation between crime free and criminal lifestyles. It is worth noting that this same process is present in the yo-yo dieting syndrome that many people struggle with.

A second explanation for fluctuations can be found in cybernetics (Ashby, 1956). Stable fluctuations can be seen as the type of fluctuation that occurs in a home heating system. Whenever the temperature changes by a few degrees, the thermostat sends a signal to the heater or air conditioner to bring the house temperature back to the desired temperature. Eventually, the temperature reaches a state of homeostasis. Unstable fluctuations are due to poor control, because the thermostat has too wide a variability. It might help to ask, what would happen if a person with low self-control tried to make a controlled change in their lives?

A third explanation for nonlinear phenomena can be found in the science of complexity, which tries to explain nonlinear changes in systems. The offender in Figure 5.11 appeared to start out on a stable path and then moved to an unstable pattern. A discussion by Harvey and Reed (1994) on organizational change appears to list some reasons why change might begin,

> "Research into deterministic chaos and dissipative systems has revolutionized our thinking about the mechanics of evolutionary change. In the absence of significant perturbations, a dissipative system will usually follow a "normal" linear trajectory. Of course, there will be the usual boundary testing, but in the absence of any sustained increase in environmental energy, the system will return to its original point of reference. At some point, however, this stable regimen is disrupted, and, if the internal movement of the system is propitious, the system's stable behavior gives way to random fluctuations. Abandoning its original trajectory, the system destabilizes and exhibits a so-called "pitchfork bifurcation" pattern, one similar to the structures Feigenbaum (1983) has described. That is, once destabilized, the system begins to fluctuate between two or more new points. This oscillation continues until it abandons its original path and takes one of the alternative points as its path of development. But even when the move is made, the

> path taken is not automatically assured. There is still a chance that conditions may conspire to block the new evolutionary path, and force the system into an alternative trajectory, or, perhaps, even back to its original trajectory. ‘The longer the system has to reorganize itself around its newly established reference state, however, the less likely it is to reverse its developmental path.
> Over the course of a system’s history, then, a far-from-equilibrium system may repeatedly pass from order to chaos, and back to a new-found order. Because of the bifurcated pattern of system evolution, a retrospective history can be reconstructed in terms of the “choices” made by the system at each bifurcation point. As each new bifurcation closes off one set of alternatives, it opens up others. Consequently, the evolutionary history of a dissipative system can be depicted as a series of irreversible cascades-or “assisted bifurcations.” That is, each bifurcation determines in broad outline the evolutionary options the system will be offered at the next crisis point. But we can only discuss these options in retrospect, since no prior bifurcated choice can determine the path that will be taken at the next evolutionary choice point” (pp. 385-386).

Harvey and Reed cite Gemmill and Smith (1985), who describe the characteristics of system transformations in organizations, as a four-part process, disequilibrium, symmetry breaking, experimentation, and reformulation. This process would appear to provide an excellent description of the steps that an offender must go through to effect a change in behavior. The steps could be describes as follows,

- Disequilibrium: A state of no change requires that the system be in equilibrium. The system is maintained in a state of equilibrium due to efforts to avoid change. Disequilibrium, or a movement away from an equilibrium state, is required for change to occur.
- Symmetry Breaking: Symmetry breaking is seen as the result of a breakdown in existing functional relationships, interactions, or habits that have kept the system in equilibrium. Lewin (1947) described effective change as the unfreezing, changing, and refreezing of behavior at a new level. Symmetry breaking is the first step, the unfreezing of old behavior tendencies.

- Experimentation: In order for change to occur, new behaviors must be tried. This requires at least some openness to variation in the behavioral repertoire.
- Reformulation: Reformulation is the selection of new semi-permanent behaviors. The preferred behavior is selected from the various behaviors that have been tried in the experimentation process.

These discussions are not meant to be the definitive answer to these issues, but are proposed as possible directions for exploration. There would appear to be many avenues that future explorations of the dynamics of risk could take. How do the individual items on the LSI correspond to the static and slow dynamic, intermediate dynamic, and rapid dynamic risk levels? How do the sequences of changes in individual items affect the future trajectory of risk? What promotes change? What promotes stability?

Limitations of the Present Study

The data used in the analyses for this study was selected in several non-random ways and so generalization to other populations of offenders would be problematic. There were several steps to the non-random selection process. The corrections officials had censored the data initially by eliminating low risk offenders. Reductions made to get quality samples, such as eliminating test-retest pairs where the same rater was not used on each assessment and eliminating the first set of scores, censored the data even further in many nonrandom ways. No policy decisions should be made from this the results of this study. The sole purpose of this study is to guide future research efforts. Because the data collection interval was set at seven months, there could be much more fluctuation in risk levels than is indicated by these results. The patterns appear to show a gradual pattern to changes in risk, but it could be that scores change even more rapidly, and also with greater magnitude of the fluctuations, than these "point in time" assessments indicate. Further analysis would need to be done with random samples over shorter time periods to determine what the actual rate of fluctuation is.

Suggestions for Future Research

Further replication efforts of this group level study are needed with larger samples to determine whether the levels of changes in risk vary over time in the same way for each population of offenders. Replication efforts are also needed using shorter time periods to determine the frequency of fluctuations.

An analysis of individual items should be done using longitudinal panel research methods developed by sociologists for studying survey data. The data used in this study is much too sparse for such methods, and so more data would need to be collected. Exploratory studies could be conducted by culling records from much larger data sets using the rules provided by these analyses, which are to only use data from reassessments that keep the same rater, and throw out the first assessment. There are potentially millions of records in the U.S. alone, so datasets of several thousand offenders in community corrections settings could be acquired. The cost of these datasets would be negligible, since interviews are already conducted.

Going forward, datasets should be built with complete longitudinal sets of data for all offenders or, at the very least, for random samples of offenders. There may be related datasets already in place that provide longitudinal data on populations of offenders. Bridges (2005) indicates the Home Office in England had set a goal of assessing offenders every 16 weeks with the Offender Assessment System (OASys), a risk scoring system similar to the LSI. If that has progressed, the data needed is simply awaiting a way to analyze it.

References

Andrews, Don A. 1982. “The Level of Supervisory Inventory (LSI): The First Follow-up.” Toronto: Ontario Ministry of Correctional Services.

Andrews, Don A., and James Bonta. 1995. *LSI-R: The Level of Service Inventory-Revised User’s Manual*. Toronto: Multi-Health Systems, Inc.

Andrews, Don A., and James Bonta. 2003. *The LSI-R: Level of Service Inventory–Revised: U.S. Norms Manual Supplement.* Toronto: Multi-Health Systems, Inc.

Andrews, Don A. and James Bonta. 2006. *The Psychology of Criminal Conduct,* 4th ed. Newark, NJ: LexisNexis.

Andrews, Don A., James Bonta, and J. Stephen Wormith. 2006. “The recent past and near future of risk and/or need assessment.” *Crime & Delinquency*, *52*(1):7-27.

Andrews Don A., and Craig Dowden. 2006. “Risk Principle of Case Classification in Correctional Treatment: A Meta-analytic Investigation.” *International Journal of Offender Therapy and Comparative Criminology* 50(1):88-100.

Andrews, Don A., and David Robinson. 1984. “The Level of Supervision Inventory: Second Report.” Toronto: Ontario Ministry of Correctional Services.

Arnold, Thomas K. 2007. “Dynamic changes in the Level of Service Inventory-Revised (LSI-R) and the effects on prediction accuracy.” Unpublished Master’s Thesis, St. Cloud State University, St. Cloud, Minnesota.

Ashby, W. Ross. 1956. *An Introduction to Cybernetics*. London: Methuen & Co LTD.

Baird, S., Heinz, R., & B. Bemus. 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.

Bak, Per. 1997. *How Nature Works: the science of self-organized criticality*. New York: Copernicus.

Bateson, Gregory. 1979. *Mind and Nature: A Necessary Unity*. New York: E. P. Dutton.

Bird, Richard J. 2003. *Chaos and Life: Complexity and Order in Evolution and Thought*. New York: Columbia University Press.

Blalock, Hubert M. Jr. (1979). *Social Statistics*, 2nd Ed. Tokyo: McGraw-Hill Kogakusha, Ltd.

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

Blumstein, Alfred, Jacqueline Cohen, Jeffrey Roth, and Christy A. Visher. 1986. *Criminal Careers and Career Criminals*, Vol. 1. Washington, DC: National Academy Press.

Bridges, Andrew. 2005. "Realising the Potential. A short focused inspection on the Offender Assessment System (OASys)." London: HM Inspectorate of Probation.

Brown, Shelly L. 2002. "The dynamic prediction of criminal recidivism." Unpublished doctoral dissertation, Queen's University, Kingston, Ontario, Canada.

Burgess, E. W. 1928. "Factors determining success or failure on parole." In A. A. Bruce, E. W. Burgess, J. Landesco, & 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.

Burnett, Ros. 2004. "To Reoffend or Not to Reoffend? The Ambivalence of Convicted Property Offenders", in S. Maruna and R. Immarigeon (eds) *After Crime and Punishment: Pathways to Offender Reintegration*, pp. 152–80. Cullompton, Devon: Willan.

Burt, Callie Harbin, Ronald L. Simons, and Leslie G. Simons. 2006. "A longitudinal test of the effects of parenting and the stability of self-control: Negative evidence for the general theory of crime." *Criminology*, 44:353-96.

Bushway, Shawn D., Alex R. Piquero, Lisa M. Briody, Elizabeth Cauffman, and Paul Mazerolle. 2001. "An empirical framework for studying desistance as a process." *Criminology*, 39(2):491-513.

Campbell, Donald T., and David A. Kenny. 1999. *A Primer on Regression Artifacts*. New York: Guilford Press.

Crocker, Linda, & Algina, James. 1986. *Introduction to Classical & Modern Test Theory*. Belmont, CA: Wadsworth Group/Thompson Learning.

Cronbach, Lee J., Lita Furby. 1970. "How should we measure "change" - or should we?" Psychological Bulletin, 74(1):68-80.

Clauset, Aaron, Cosma Rohilla Shalizi, and M.E.J. Newman. 2007. "Power-law distributions in empirical data." Donloaded on November 11, 2008 from http://arxiv.org/abs/0706.1062.

Cook, Thomas D., and Donald T. Campbell. 1979. *Quasi-experimentation: Design and Analysis Issues for Field Settings*. Boston, MA: Houghton Mifflin.

Douglas, Kevin S., and Jennifer L. Skeem. 2005. "Violence risk assessment: Getting specific about being dynamic." *Psychology, Public Policy, and Law*, 11(3):347–383.

Elder, Glen H., Jr. (1985). "Perspectives on the Life Course." pp. 23-49 In Elder, Glen H., Jr., ed. *Life course dynamics: Trajectories and transitions, 1968-1980*. Ithaca: Cornell University Press.

Ezell, Michael E., and Lawrence E. Cohen. 2005. *Desisting from Crime: Continuity and Change in Long-Term Crime Patterns of Serious Chronic Offenders*. Oxford: Oxford University Press.

Farrington, David P., Bernard Gallagher, Lynda Morley, Raymond J. St. Ledger, and Donald J. West. 1986. "Unemployment, School Leaving, and Crime." *British Journal of Criminology,* 26:335-356.

Feigenbaum, Mitchell. 1983. "Universal behavior in nonlinear systems." *Physica,* 7D(1-3):16-39.

Flores, Anthony W., Christopher T. Lowenkamp, Alexander M. Holsinger, and Edward J. Latessa. 2006. "Predicting outcome with the Level of Service Inventory-Revised: The importance of implementation integrity." *Journal of Criminal Justice*, 34:523-529.

Gemmill, Gary, and Smith, Charles. 1985. "A Dissipative Structure Model of Organization Transformation." *Human Relations,* 38:751-756.

Gendreau, Paul, Claire Goggin, and Paula Smith. 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.

Gendreau, Paul, Tracy Little, and Claire Goggin. 1996. "A Meta-analysis of the Predictors of Adult Offender Recidivism: What Works!" *Criminology*, 34(4):575-607.

Glaser, Daniel. 1964. *The Effectiveness of a Prison and Parole System*. Indianapolis: Bobbs-Merrill.

Glaze, Lauren E., and Thomas P. Bonczar. 2007. "Probation and Parole in the United States, 2006, Publication No. NCJ 220218." Washington, DC: U.S. Department of Justice, Bureau of Justice Statistics.

Gottfredson, Michael R., and Travis Hirschi. 1990. *A General Theory of Crime*. Palo Alto, CA: Stanford University Press.

Hart, Hornell. 1923. “Predicting parole success.” *Journal of Criminal Law and Criminology*, 14(3), 405-414.

Harvey, David L. and Michael H. Reed. 1994. “The evolution of dissipative social systems.” *Journal of Social and Evolutionary Systems*, 17:371–411.

Hay, Carter, and Walter Forrest. 2006. “The Development of Self-Control: Examining Self-Control Theory's Stability Thesis.” *Criminology*, 44(4):739-774.

Hirschi, Travis, and Michael R. Gottfredson. 1983. “Age and the Explanation of Crime.” *The American Journal of Sociology*, 89(3):552-584.

Hirschi, Travis, and Michael R. Gottfredson. 2001. “Self-control theory.” In *Explaining Criminals and Crime*, Raymond Paternoster and Ronet Bachman, eds. Los Angeles, CA: Roxbury.

Horney, Julie, D., Wayne Osgood, and Ineke Haen Marshall. 1995. “Criminal Careers in the Short-Term: Intra-Individual Variability in Crime and its Relation to Local Life Circumstances.” *American Sociological Review,* 60(5):655-673.

Hubbard, Dana Jones, Lawrence F. Travis, III, and Edward J. Latessa. 2001. “Case Classification in Community Corrections: A National Survey of the State of the Art. Final report of grant no. 98-IJ-CX-0008 awarded by the National Institute of Justice.” Cincinnati: University of Cincinnati Center for Criminal Justice Research.

Jacobson, Neil S., William C. Follette, and Dirk Revenstorf. 1984. “Psychotherapy Outcome Research: Methods for Reporting Variability and Evaluating Clinical Significance.” *Behavior Therapy,* 15: 336- 352.

Jacobson, Neil S., and Paula Truax. 1991. “Clinical Significance: A Statistical Approach to Defining Meaningful Change in Psychotherapy Research.” *Journal of Consulting and Clinical Psychology*, 59: 12–19.

Janis, Irving L. 1959. “Decisional Conflicts: A Theoretical Analysis.” *The Journal of Conflict Resolution*, 3(1):6-27.

Janis, Irving L. and Leon Mann. 1977. *Decision Making*. New York: Free Press.

Kazdin, Alan E., Helena Chmura Kraemer, Ronald C. Kessler, David J. Kupfer, and David R. Offord. 1997. “Contributions of Risk-factor Research to Developmental Psychopathology.” *Clinical Psychology Review*, 17(4):375-406.

Kelly, Janice R., and Joseph E. McGrath. 1988. *On Time and Method: Applied Social Research Methods Series*, Vol. 13. Newbury Park, CA: Sage Publications.

Kreamer, Sally. 2004. “Quality Assurance and Training in Offender Assessment.” *Assessment Issues for Managers. Topics in Community Corrections, Annual Issue 2004*. pp. 13-19. Washington DC: U.S. Department of Justice.

Kroner, Daryl G., Jeremy F. Mills, and John R. Reddon. 2005. “A coffee can, factor analysis, and prediction of anti-social behavior: The structure of criminal risk.” *International Journal of Law and Psychiatry*, 28: 360-374.

Langan, Patrick A. and David J, Levin. 2002. “Recidivism of Prisoners Released in 1994”. Bureau of Justice Statistics, Special Report (NCJ, 193427). Washington, DC: U.S. Department of Justice, Office of Justice Programs, National Institute of Justice.

Lewin, Kurt. 1935. “Chapter 4: The Psychological Situations of Reward and Punishment.” *A Dynamic Theory of Personality: Selected Papers*. New York: McGraw-Hill.

Lewin, Kurt. 1964. "Group dynamics and social change." In Amitai Ezioni & Eva Etzioni, E. Eds., *Social change*. New York: Basic Books, Inc.

Loeber, Rolf, and Marc Le Blanc. 1990. "Toward a developmental criminology." In Michael Tonry and Norval Morris, Eds. *Crime and Justice: A Review of Research,* Vol. 12:375-473. Chicago and London: University of Chicago Press.

Martinson, Robert. 1974. "What Works? -- Questions and Answers About Prison Reform." *The Public Interest*, 10:22-54.

Maruna, Shadd. 2001. *Making Good: How Ex-Convicts Reform and Rebuild Their Lives*. Washington, DC: American Psychological Association Books.

Matza, David. 1964. *Delinquency and drift*. New York: Wiley.

Minnesota Department of Corrections. 2006. "Classification and Assessment." Division Directive: 203.016, Issue Date: 3/7/06. Downloaded on August 25th, 2007 from http://www.doc.state.mn.us/DocPolicy2/Document/203.016.htm.

Mitchell, Ojmarrh, and Doris Layton MacKenzie. 2006. "The Stability and Resiliency of Self-Control in a Sample of Incarcerated Offenders." *Crime & Delinquency*, 52(3):432-449.

Molenaar, P.C.M. (2008). On the implications of the classical ergodic theorems: Analysis of developmental processes has to focus on intra-individual variation. Developmental Psychobiology, 50, 60-69.

Motiuk, Lawrence.L. 1991. "Antecedents and Consequences of Prison Adjustment: A Systematic Assessment and Reassessment Approach." Unpublished doctoral dissertation, Carleton University, Ottawa, Ontario.

Motiuk, Lawrence L., James Bonta, and Don A. Andrews. 1990. “Dynamic Predictive Criterion Validity in Offender Assessment.” A paper presented at the Canadian Psychological Association Annual Convention, Ottawa, June, 1990.

Myers, Peter L, and Norman R. Salt. 2007. *Becoming An Addictions Counselor; A Comprehensive Text*. 2nd ed. Sudbury, MA: Jones and Bartlett Publishers.

Nagin, Daniel, and Raymond Paternoster. 2000. “Population Heterogeneity and State Dependence: State of the Evidence and Directions for Future Research.” *Journal of Quantitative Criminology*, (16)2:117-144.

Nunnally, Jum C., and Ira H. Bernstein. 1994. *Psychometric Theory,* 3rd ed. New York: McGraw Hill.

Peters, Roger H., and Harry K. Wexler. “2005. Substance Abuse Treatment For Adults in the Criminal Justice System: A Treatment Improvement Protocol TIP 44. DHHS Publication No. (SMA) 05-4056.” Rockville, MD: Substance Abuse and Mental Health Services Administration.

Piquero, Alex R., John M. McDonald, and Karen F. Parker. 2002. “Race, Local Life Circumstances, and Criminal Activity.” *Social Science Quarterly,* 83(3): 654–70.

Piquero, Alex R. 2004. “Somewhere Between Persistence and Desistence: The Intermittency of Criminal Careers.” In S. Maruna & R. Immarigeon, Eds., *After crime and punishment: Pathways to offender reintegration,* pp. 102-125. Cullompton, Devon, UK: Willan.

Prochaska, James O., and Carlo C. DiClemente. 1983. “Stages and processes of self-change of smoking: Toward an integrative model of change.” *Journal of Consulting and Clinical Psychology,* 51(3): 390-395.

Prochaska, J. O., Carlo C. DiClemente, and Norcross, J. C. (1992). In search of how people change: Applications to addictive behavior. American Psychologist, 47, 1102-1114.

Prochaska, James O., Wayne F. Velicer, Joseph S. Rossi, Michael G. Goldstein, Bess H. Marcus, William Rakowski, Christine Fiore, Lisa L. Harlow, Colleen A. Redding, Dena Rosenbloom, and Susan R. Rossi. 1994. "Stages of change and decisional balance for 12 problem behaviors." Health Psychology, 13(1):39-46.

Raffaelli, Marcela, Lisa J. Crockett, and Yuh-Ling Shen. 2005. "Developmental stability and change in self-regulation from childhood to adolescence." *The Journal of Genetic Psychology,* 166:54–75.

Raynor, Peter. 2007. "Risk and Need Assessment in British Probation: The Contribution of the LSI-R." *Psychology, Crime, & Law*, 13(2):125-138.

Raynor, Peter, Jocelyn Kynch, Colin Roberts, and Simon Merrington. 2000. "Risk and Need Assessment in Probation Services: an Evaluation. Research Study 211." London: Home Office.

Robins, Lee N. 1978. "Sturdy childhood predictors of adult antisocial behaviour: replications from longitudinal studies." *Psychological Medicine*, 8:611-622.

Rogosa, David, David Brandt, & Michele Zimowski. 1982. "A growth curve approach to measurement of change." *Psychological Bulletin*, 92(3):726-748.

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

Sampson, Robert J., & John H. Laub. 2003. "Life Course Desisters? Trajectories of Crime Among Delinquent Boys Followed to Age 70." *Criminology*, 41(3):555–592.

Simon, Herbert A. 1955. "On a Class of Skew Distribution Functions". *Biometrika,* 42(3/4):425–440.

Singer, Judith D., and John B. Willett. 2003. *Applied Longitudinal Data Analysis: Modeling Change and Event Occurrence.* New York: Oxford University Press.

Smith, Peter, & Dorcas Beaton. 2008. "Measuring Change in Psychosocial Working Conditions: Methodological Issues to Consider When Data are Collected at Baseline and One Follow-up Time Point." *Occupational Environmental Medicine*, 65: 288-296.

Sutherland, Edwin H. 1947. *Principles of Criminology*. 4th ed. Chicago: University of Chicago Press.

Turner, Michael G., and Alex R. Piquero. 2002. "The stability of self-control." *Journal of Criminal Justice,* 30:457–471.

Velicer, Wayne F., Carlo C. DiClemente, James O. Prochaska, and Nancy Brandenberg. 1985. "A decisional balance measure for assessing and predicting smoking status." *Journal of Personality and Social Psychology*, 48(5):1279-1289.

Warner, Sam B. 1923. "Factors Determining Parole from the Massachusetts Reformatory: The Report of the Director of Committee on Criminal Records and Statistics." *Journal of Criminal Law and Criminology,* 14(2):172-207.

Willett, John B. (1989). "Some Results On Reliability For The Longitudinal Measurement Of Change: Implications For The Design Of Studies Of Individual Growth." *Educational and Psychological Measurement*, 49:587-602.

Winfree, L. Thomas Jr., Terrance J. Taylor, Ni He, and Finn-Aage Esbensen. 2006. "Self-Control and Variability Over Time: Multivariate Results Using a 5-Year, Multisite Panel of Youths." *Crime & Delinquency*, 52(2):253-286.

Zamble, Edward, and Vernon L. Quinsey. 1997. *The Criminal Recidivism Process*. New York: Cambridge University Press.