Predictive Validity of the YLS/CMI as Administered in Nebraska Probation

October 4, 2017

University of Nebraska Law Psychology Program

Submitted By

Richard L. Wiener, PhD, MLS.

Alisha Caldwell Jimenez, J.D.

Taylor Petty, B.S.

Julie Wertheimer, B.S.

Table of Contents

Figures and Table ……...4

Introduction ……... 5

Introduction to the YLS/CMI.……..... 5

Statement of the Problem …….. 8

Reliability ……..... 9

Predictive Validity……..... 9

Statistical Significance and Effect Sizes ……11

Main Effects and Interactions……..12

Logistic Regression……..12

Summary of the YLS/CMI Studies in the Literature……...……13

Analysis Methods……...15

Description of the Sample……...…….15

Results……20

Results Section Overview……..20

Results Section: YLS/CMI Predictive Validity……22

YLS/CMI Total Scores versus YLS/CMI Risk Levels…….....…….22

Analysis of Success/Failure…….24

Pure Success…….25

Partial Success…….30

Semi-Success and Limited Success…….34

Results Section: Moderation of YLS/CMI – Predictive Validity……..38

The Need to Examine Moderation……...38

Youth Gender……...39

Youth Race and Ethnicity……....43

Results Section: YLS/CMI Domain Status – Predictive Validity…….47

Results Section: Establishing Cutoffs with ROC Analysis……...50

Executive Summary and Conclusions……...59

References……..65

Figures and Tables

Figure 1 16
Figure 2 16
Figure 3 17
Figure 4 18
Figure 5 19
Figure 6 23
Figure 7 24
Figure 8 26
Figure 9 30
Figure 10 31
Figure 11 34
Figure 12 35
Figure 13 38
Figure 14 42
Figure 15 42
Figure 16 46
Figure 17 46
Figure 18 47
Figure 19 53
Figure 20 55
Figure 21 59

Table 1 20
Table 2 28
Table 3 29
Table 4 32
Table 5 33
Table 6 36
Table 7 37
Table 8 40
Table 9 41
Table 10 44
Table 11 45
Table 12 48
Table 13 49
Table 14 54
Table 15 55
Table 16 56
Table 17 56
Table 18 58

Introduction

This report details the work that the University of Nebraska/Law Psychology Program (UNLLP) completed to date to analyze the validity of the Youth Level of Service/Case Management Inventory (YLS/CMI) as Nebraska Probation uses it to assess risk levels of youth in the juvenile justice system. While the work is ongoing with more questions to answer, this report explains several major and important findings about the validity of the YLS/CMI. The report consists of five major sections: 1) introduction to the YLS and a brief review of existing literature about the measure, 2) statement of problem and defining concepts, 3) methods section, 4) results sections (validity, moderation, domain status and area under the curve analyses) and 5) a final executive summary with conclusions. UNLLP in cooperation with the Nebraska Administrative Office of Probation completed this work with Title II funds administered through the Nebraska crime commission.

INTRODUCTION TO THE YLS/CMI

Hoge and Andrews (2002) developed the YLS/CMI as an assessment tool in accordance with Andrew and Bonta's Risk-Need-Responsivity (RNR) model (Andrews et al., 1990). The RNR model emphasizes the importance of tailoring offender treatment to the needs of adults and youth in the justice system (Andrews & Bonta, 2010; Andrews & Dowden, 2006; Parent, Guay, & Knight, 2012; Luther, Reichert, Holloway, Roth & Aalsma, 2011; Olver, Stockdale, & Wormith, 2014; Taylor-Gooby & Zinn, 2006). The RNR model is an empirically based rehabilitation approach to corrections used worldwide to treat a variety of offenders (Andrews & Bonta, 2003; Andrews, Bonta, & Wormith, 2006; Andrews & Dowden, 2006; Andrews, Zinger, Hoge, Bonta, Gendreau, & Cullen, 1990; Gendreau & Andrews, 1990; Ward, Melser, & Yates, 2007). According to RNR, the principles of assessing risk level through measuring criminogenic

needs and intervening through cognitive behavioral learning techniques are the most effective way to bring about desistance (Andrews & Bonta, 2010). The risk principle proposes that the level of treatment should match the level of risk so that high-risk offenders should receive stronger doses of intervention, while low risk offenders should receive minimal or no treatment. The need principle states that treatments should focus on criminogenic needs, which are the factors most predictive of decisions to engage in criminal activity. The responsivity principle also advises that correctional programs should match the characteristics of the offenders (e.g., learning style, motivation, intensity, etc.) with treatments that fit their individual strengths and weaknesses.

Several studies have provided evidence to support the RNR model as a generally effective means of reducing recidivism (Andrews et al., 1990) with general and special populations, such as violent offenders (Dowden & Andrews, 2000), women (Dowden & Andrews, 1999a), and juveniles (Dowden & Andrews, 1999b; Dowden & Andrews, 2003). The first component of the RNR model requires objective and unbiased assessment of offender risk. The NAOP adopted the RNR model to supervise and rehabilitate offenders serving probation sentences in 2006 and implemented the Level of Service/Case Management Inventory (LS/CMI) (Andrews, Bonta, & Wormith, 2004) as a risk assessment and need assessment tool for adults. The LSI measures adult offender (i.e., inmates, probationers and paroles) risk of recidivism, including an assessment of offenders' criminogenic needs resulting in service recommendations and level of required supervision recommendations. Each scale on the LS/CMI includes a series of binary items that together measure one of the "Big Four" predictors of criminal behavior (i.e., criminal history, anti-social attitudes, antisocial associates, and antisocial personality) or one of the remaining four scales that make up the "Central Eight" criminogenic factors (i.e.,

education/employment, family/marital status, leisure recreation, and substance abuse) (Andrews & Bonta, 2010).

The YLS/CMI is an adaptation of the LSI-R modified to measure risk and needs of youth in the juvenile justice system, which Hoge and Andrews developed in 2002 specifically for probation officers and mental health professionals to administer to the youth with whom they work. It is a 42-item standardized instrument administered in a semi-structured 60 to 90 minute interview with youth that results in individual scores (ranging from 3 to 9) and corresponding risk levels (i.e., low, moderate, and high) for each of eight criminogenic domains (prior and current offenses, family circumstances/parenting, education/employment, peer relations, substance abuse, leisure/recreation, personality/behavior, and attitudes/orientation). All the domain factors except the first are dynamic, that is they are variables that are associated with offending and are changeable overtime (Clarke, Peterson-Badali, & Skilling, 2017). Prior and current offenses is the only static factor, that is, a strong and direct predictor of criminal conduct, which cannot change through social or psychological interventions (Clarke, Peterson-Badali, & Skilling, 2017). The YLS/CMI produces risk scores and levels in each of the 8 criminogenic domains, which roughly correspond to the Central Eight, as well as, a final total score that test administrators may use to establish an overall level of risk.

In order to test the validity of a risk instrument, it is first necessary to address the definition of success and failure, which serves as the outcome variable in studies of validity. That is, researchers define risk as the risk that a youth will fail at some component of the juvenile justice system and then use an assessment tool to predict outcomes labelling validity as the ability of the instrument to predict the outcome. One definition of failure features outcomes for specific components of the justice system such as failure in probation (Childs, Ryals, Frick,

Lawing, Phillippi, & Deprato, 2013) due to a new offense or due to technical violations (Schmidt, Sinclair & Thomasdóttir, 2015). However, more studies define failure across multiple components of the juvenile justice system and rely on the likelihood of reoffending or returning to the system as their outcome measures. Thus, many existing assessments of the YLS/CMI validity use scores and levels on the instrument to predict the likelihood of some form of reoffending or returning to the system over a specific period of time. Recent studies have examined the YLS/CMI as a predictor of future violent offenses (Schmidt, Campbell, & Houlding, 2011), sexual offenses (Schmidt, Sinclair & Thomasdóttir, 2015); property offenses (Garcia-Gomis, Villanueva & Jara, 2016) and changes in risk over time (Viljoen, Shaffer, Gray & Douglas, 2017). However, the most common definition of risk in the empirical literature is general reoffending of any type. In 2014, Olver, Stockdale and Wormith conducted a meta-analysis of the YLS/CMI that included 36 separate studies that used general recidivism as the outcome factor. The current research combined success/failure in probation and return to probation as its primary outcome factors, using scores and levels on the YLS/CMI to predict that outcome. The extent to which the YLS/CMI was able to accurately forecast this type of failure was our measure of its predictive validity.

STATEMENT OF THE PROBLEM

The problems for the current project were to measure the predictive validity of the YLS/CMI as it is used in Nebraska probation, determine if the instrument performs differently across gender and race classifications in the state, determine the strongest domains in the predictive validity equation, and estimate the alternative cutoff scores for levels of risk as probation officers use the YLS/CMI in Nebraska. The methodology for answering these questions appears in the next section of the report. This section of the report first defines several

of the technical research concepts that appear in the methods and results sections along with a summary of the status of the YLS/CMI validity studies in the literature to date.

Reliability. All instruments involve unpredictable random errors that move scores up or down in erratic directions. These inconsistent fluctuations in direction and magnitude result from individual differences in interviewers' and interviewee's emotions, attitudes, and cognitive understanding of the evaluation system. In other words, different youth respond to the same interview questions on the YLS/CMI in different ways that result in some unpredictability in their answers. Random error is impossible to eliminate but is possible to reduce in magnitude and the resulting instrument's reliability is strong to the extent that random error is minimal. Researchers measure reliability using internal consistency statistics (e.g., Cronbach's alpha) that document the extent to which items within a specific domain agree with each other across a sample of youth or by using interrater agreement statistics (e.g., percent agreement, Kappa Coefficient and Intraclass Correlation Coefficient) to document that different interviewer/raters score the same youth in the same way across a number of factors. Higher scores show lack of random error and greater reliability. An analogy brings clarity to the concept of reliability: An archer who shoots all his arrows in a tight circle even if they are not on the bullseye is reliable, that is, the archer displays little random error.

Predictive Validity. Systematic errors are deviations in measurement that are consistent in direction and magnitude. This type of error may result from youth responding to irrelevant components of interview questions in the same way regardless of the other relevant components. Systematic error results from asking the wrong questions or asking the right questions in a way that does not measure the underlying domain under investigation. In theory, it is possible to eliminate systematic error but in reality no instruments are free of all systematic error. Validity is

present to the extent to which systematic error is minimal allowing an instrument to accurately measure and predict the construct (here success/failure) that it is theorized to predict. Returning to the archer analogy, if the shooter's arrows wind up tightly distributed on the target surrounding the bull's-eye, then the archer is both reliable and valid.

That is, higher levels of risk on the YLS/CMI should be associated with higher rates of failure and lower risk with lower rates of failure. Researchers measure validity with tests of the "effect size" or strength of the relationship between the instrument and the outcome. In other words, the effect size for the YLS/CMI is the strength of the relationship between the YLS scores and the dichotomous outcome of success or failure. This report uses the point-biserial correlation coefficient, r, as an indicator of effect size. It ranges from -1.00 to 1.00 with the sign measuring the direction of the relationship (i.e., positive numbers indicate that increases in risk level predict increases in the likelihood of failure or returning to probation in the future) and the absolute magnitude (0 to 1.00) indicating the strength of the relationship. Values of *r* that are less than .10 are small effect sizes, values between .10 and .30 are moderate effect sizes, and those between .30 and .50 are large effect sizes. (Note: effect sizes rarely exceed $r = .50$ in most areas of research – for example, the r value typically cited for the strength of association between smoking cigarettes and lung cancer is $r = .40$).

Another measure of effect size that this report uses is the AUC statistic or area under the ROC (Reviewer Operating Characteristic) curve which plots the false negative rate of a hit (here -- predicting a success when a failure occurs) on the x-axis of a graph and the true positive rate of a hit (here -- predicting a failure when a failure actually occurs). Researchers use AUC statistics to help select cutoff values for an instrument's scores (e.g., determining cutoff levels to use for YLS/Scores to frame risk levels). A conceptual definition of an AUC is that it is the

probability of an outcome (here - failure) for an individual who is selected at random from the outcome group predicted to fail. If a risk instrument has an AUC of .50 then an individual selected at random from the group predicted to recidivate has a 50% chance of recidivating and a 50% chance of not recidivating – therefore, the prediction is of no value. However, if an instrument has an AUC value greater than .50, then it has value because it can predict the outcome greater than at a chance level. Values of AUC that are between .50 and .56 are small effect sizes, values between .56 and .67 are moderate effect sizes, and those between .67 and .79 are large effect sizes.

This report does not include reliability analyses because to do so would require multiple YLS/CMI administrations on the same youths by different officers. Instead we analyze predictive validity with *r* values and AUC statistics. It is important to note that one way to improve validity is to improve reliability because random error or noise in measurement limits the instrument's validity. There are ways to test and improve reliability and UNLLP can assist juvenile probation to accomplish this if it is a priority.

Statistical Significance and Effect Sizes. An effect size may vary in magnitude based upon the laws of chance. That is, an effect size of $r = .20$, may vary across samples from a low of, say -.10 to a high of say, .40 as a result of chance or random factors in the data. In this example a value of $r = 0$, would meant that there is no relationship in the population. Statistical significance measures the probability of obtaining an effect size in a sample that is greater than 0 ***by chance alone.*** Tests of statistical significance (e.g., t, F, chi square, Wald) determine if the effect size is due to chance. The social sciences rule out chance when the probability of obtaining an effect size greater than 0 by chance alone is less than .05 or 5 out of 100 times ($p < .05$). If p <

.05 the effect size is statistically significant but if $p > .05$, it is more likely to have been due to chance.

Main Effects and Interactions. In this report, we are interested in two types of effect sizes, main effects and interaction effects. "Main effect" refer to the relationship between one predictor (here, the YLS/CMI score or level) and one outcome factor (here, failure in probation or return to probation). To the extent that YLS/CMI scores show significant and positive effect sizes with probation failures, the YLS/CMI demonstrates validity. The greater the effect size is, the more validity the instrument shows. "Interaction effects" refer to moderation or the extent to which the effect of one variable depends upon the level of a second variable. Thus, if there is an interaction between sex of the youth and the validity of the YLS/CMI, that means the effect size of the YLS/CMI on failure depends upon the sex of the youths. In other words, the ability of the YLS/CMI to predict failure is different for boys and girls. Interaction effect sizes can also be statistically significant or they can be due to chance. This report tests the interactions of the YLS/CMI with sex of the youth and minority status of the youth. Significant and large interactions between the YLS/CMI and sex of the youth or race of the youth are problematic because they suggest that the instrument could be biased against one sex (likely girls) or against minorities. It is optimal that there are no interactions between the YLS/CMI and either of these factors so the association, i.e., validity, of the YLS/CMI does not depend upon sex or minority status.

Logistic Regression. Two types of analyses make up this report, Logistic Regression and ROC/AUC analyses. Logistic regression predicts the outcome of a binary outcome variable (i.e., failure v. success) based on one or more predictor variables (e.g., Possible time youth was in the system, YLS/CMI scores, gender and race). Logistic regression calculates the optimal

weights for each predictor variable (Beta's, effect sizes for each predictor variable (Odds ratios – converted to *r*'s) and tests the statistical significance of the predictors with the Wald statistic and chi-square statistic. This report presents results from a number of logistic regressions to test the validity of the YLS/CMI. The report uses the ROC/AUC analyses (see above) to estimate cutoff points for the YLS/CMI scores.

Summary of the YLS/CMI studies in the literature. There are a large number of studies testing the validity of the YLS/CMI in the existing literature, most of which make use of these basic concepts. In 2014, Olver, Stockdale and Wormith conducted a meta-analysis of all the existing studies testing the validity of the LSI measures including the Adult LS/CMI and the YLS/CMI. A meta-analysis is a statistical analysis of groups of studies that include the same predictors and the same outcomes. A meta-analysis aggregates multiple studies and calculates the overall effect sizes across the studies, tests the statistical significance of the resulting effect sizes and examines moderators for the effect size. Olver et al., (2014) examined 128 studies conducted between 1981 and 2012 conducted in Canada (where the LSI scales were developed), the United States, Australia, the United Kingdom, Singapore, Germany, Japan, New Zealand, and Pakistan. There were 92 studies of the adult LSI scales and 36 studies of the youth YLS scales. The YLS overall effect size was moderate and statistically significant ($r = .25$, $p < .001$), meaning that the probability of finding this effect size by chance was less than 1 in 1000. In Canada, the effect size was $r =. 33$, in the United States, $r = .22$ and outside North America, $r = .28$, all statistically significant ($p < .001$). While there have been additional tests of the validity of the YLS/CMI since 2014, Olver et. al.'s study is the most recent meta-analysis. ***Effect sizes calculated in this report of the YLS/CMI in Nebraska probation that are at or above r = .22***

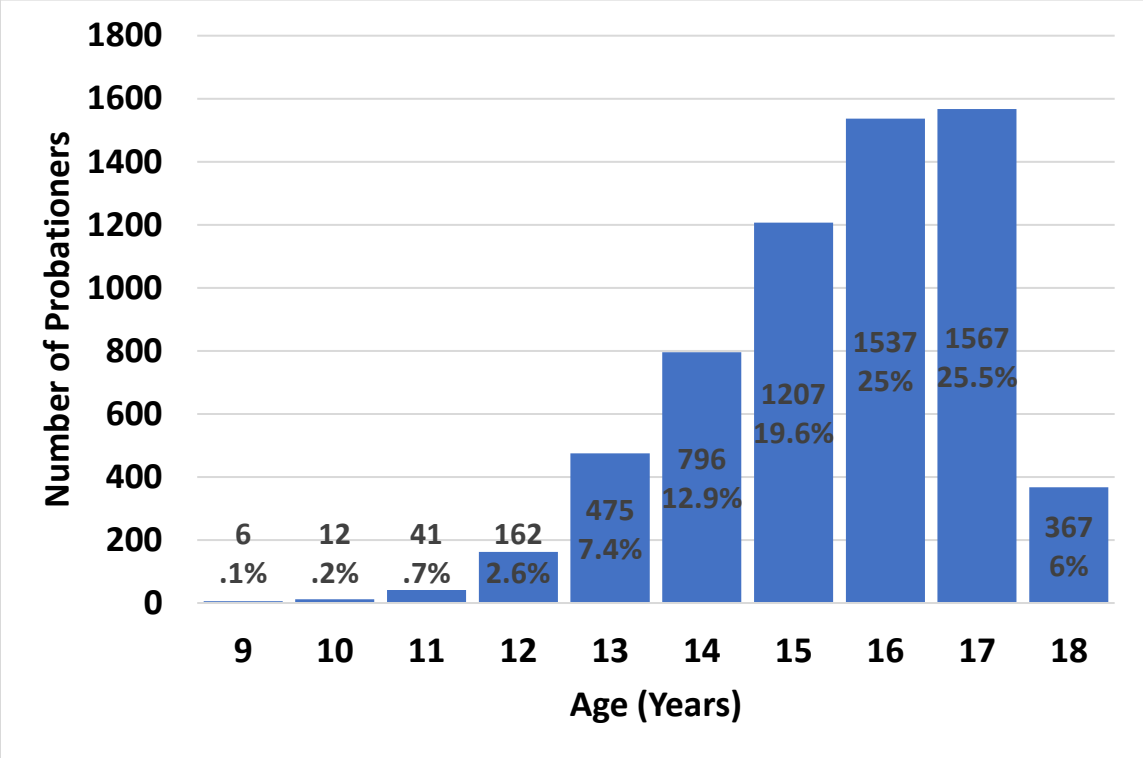
show that the validity of the YLS/CMI in predicting failures for Nebraska youth is comparable to those established in the rest of the United States.

UNLLP has conducted two validity studies of the LS/CMI as used in adult probation in Nebraska. The first relied on NPACS data and defined a failure as returning to probation or jail after termination of probation. The effect size for this analysis was $r = .22$, which was statistically significant ($p < .001$) and similar to the Olver et al.'s meta-analysis for the LSI scales in the United States ($r = .21$, $p < .001$). The second study combined the NPACS database with LS/CMI data and data from JUSTICE, which is the Nebraska court database that includes entries for each conviction in a Nebraska court. UNLLP turned these merged data bases into a recidivism rate using the Nebraska Supreme Court's definition: "As applied to adults, recidivism shall mean a final conviction of a Class I or II misdemeanor, a Class IV felony or above, or a Class W misdemeanor based on a violation of state law or an ordinance of any city or village enacted in conformance with state law, within 3 years of being successfully released." (Nebraska Supreme Court Administrative Operations, Article 10, §1-1001). This record is more accurate and more complete than the NPACS estimation of recidivism defined as returning to adult probation. Using LS/CMI raw scores to predict outcome, UNLLP found the effect size to be $r = .26$, higher than the effect size in the Olver et al. (2014) meta-analysis for the United States. The current analyses of the YLS/CMI use NPACS data to define outcomes as failure in probation and/or return to probation. At the current time, UNLPP is working on analyzing the JUSTICE recidivism data for youth in Nebraska Probation and will be able to connect the YLS/CMI scores to that measure of recidivism in the future.

ANALYSIS METHODS

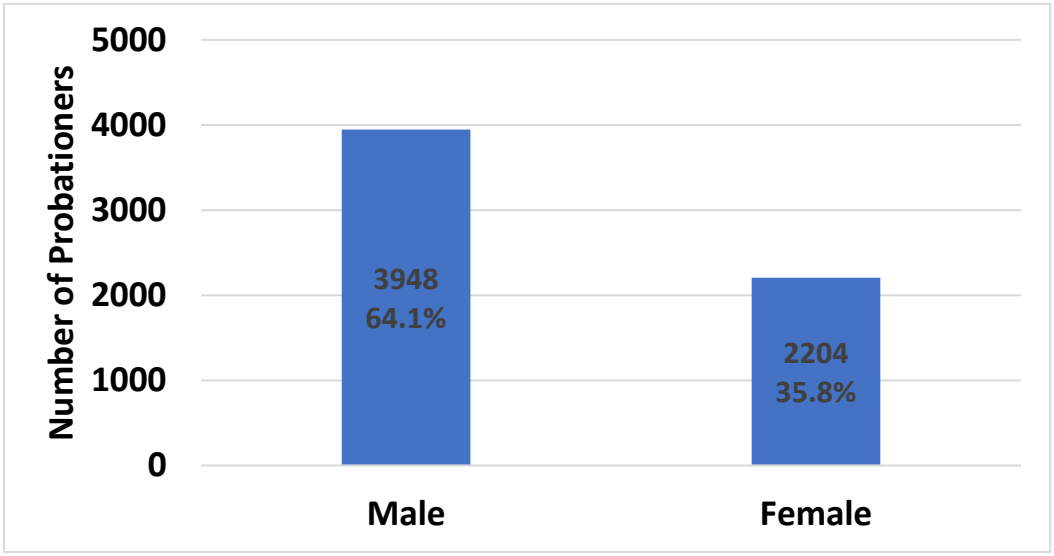
Description of the Sample. The dataset used for all the analyses in this report included records for 6,158 individual juvenile probationers downloaded from NPACS, each of whom had an index YLS/CMI assessment (i.e., the first one within our time frame) between May 24, 2007 and November 11, 2015. Figure 1 displays the age of the youth (in years) at the time of the first charge that resulted in the administration of the YLS/CMI and a probation sentence. The average age of the first YLS/CMI administration was 15.54 years with the youngest being 9 and the oldest 18. The median age was 16 and the mode was 17, showing that the most common age for a youth's first contact with probation was 16 or 17 years old. As shown in Figure 1 some of the children were eligible for additional youth probation charges and sentences for almost 9 years, while others may have been eligible for only a matter of days. To control for the variable eligibility rates we calculated a new variable, "Possible Time in the System" marked how long each child was eligible for inclusion in the juvenile probation system. All validity analyses controlled for how much time a youth was eligible for probation by including this control factor in the logistic regressions that follow.

Figure 1. Youth's Age at First YLS/CMI Assessment Date



Next, Figure 2 displays the gender of the youth in the sample showing that most of the children in the data set were boys, outnumbering the girls by a ratio of 1.79 to 1, that is, for every girl in the system there were 1.79 boys on probation.

Figure 2. Gender of Youth Included in the Sample



With regard to race, the majority of the final sample self-reported that they were White (European Caucasian) (53.5%), 22.4% were African American (Black), 3.0% were American Indian, <1% Asian/Pacific Islander and 20.1% identified as "Other" (see Figure 3). With regard to ethnicity, 19.9% of the probationers identified as being of "Hispanic Origin" (see Figure 4).

Figure 3. Self-reported Race of the Youth Included in the Sample

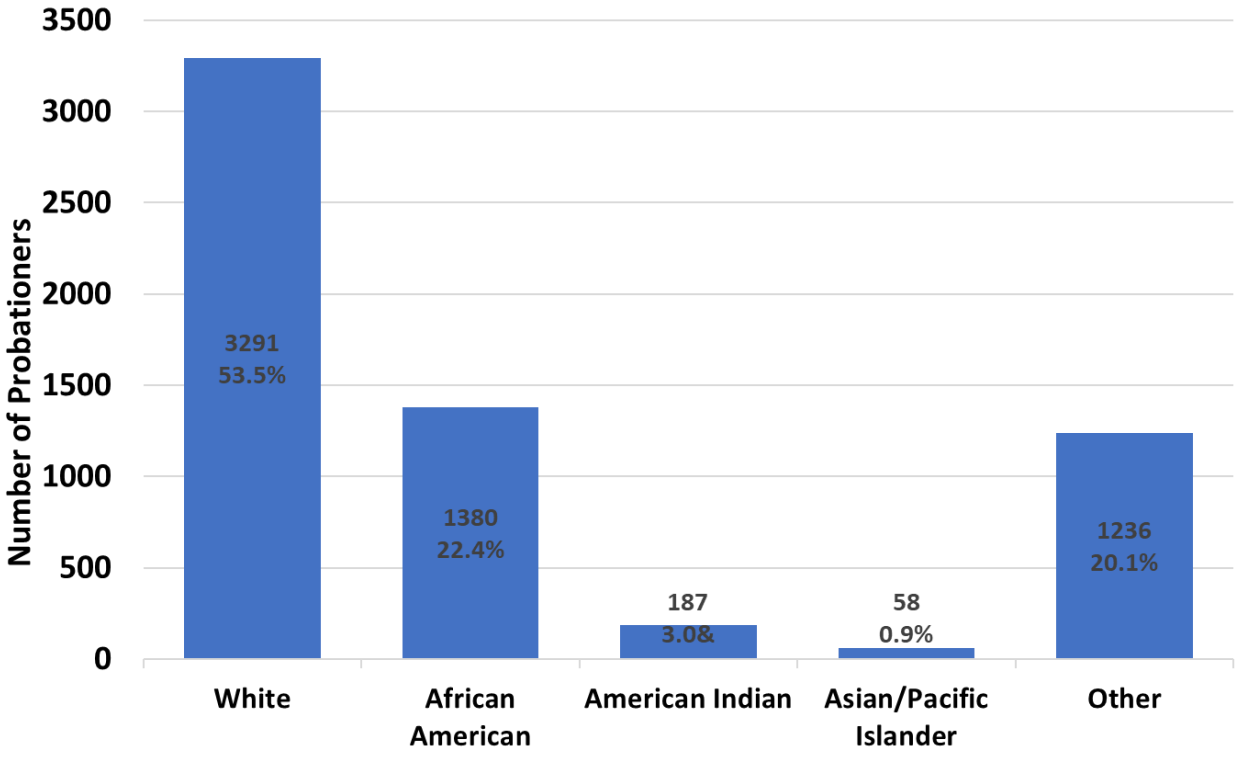
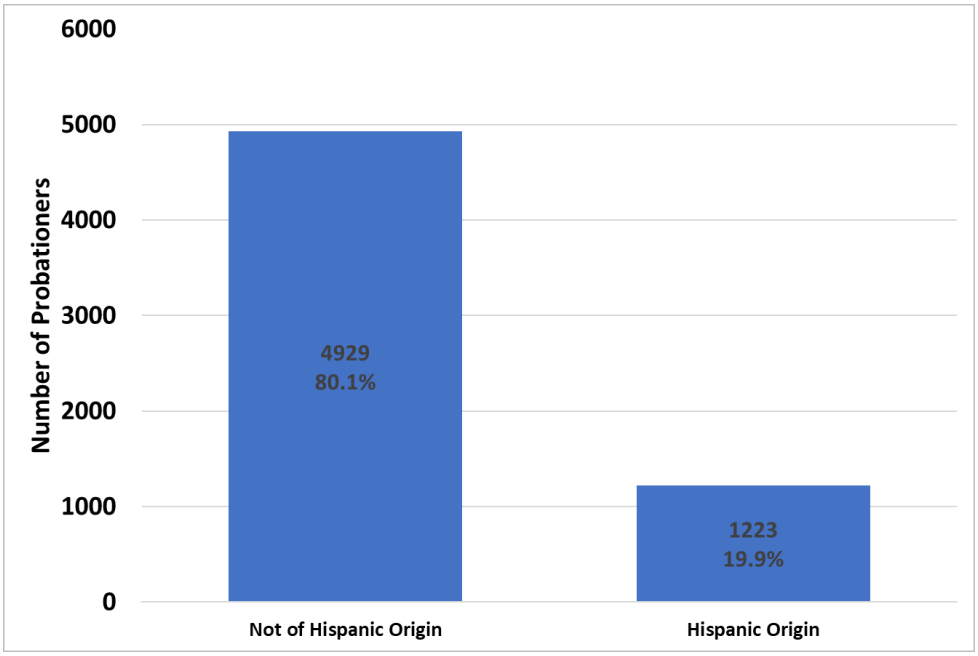
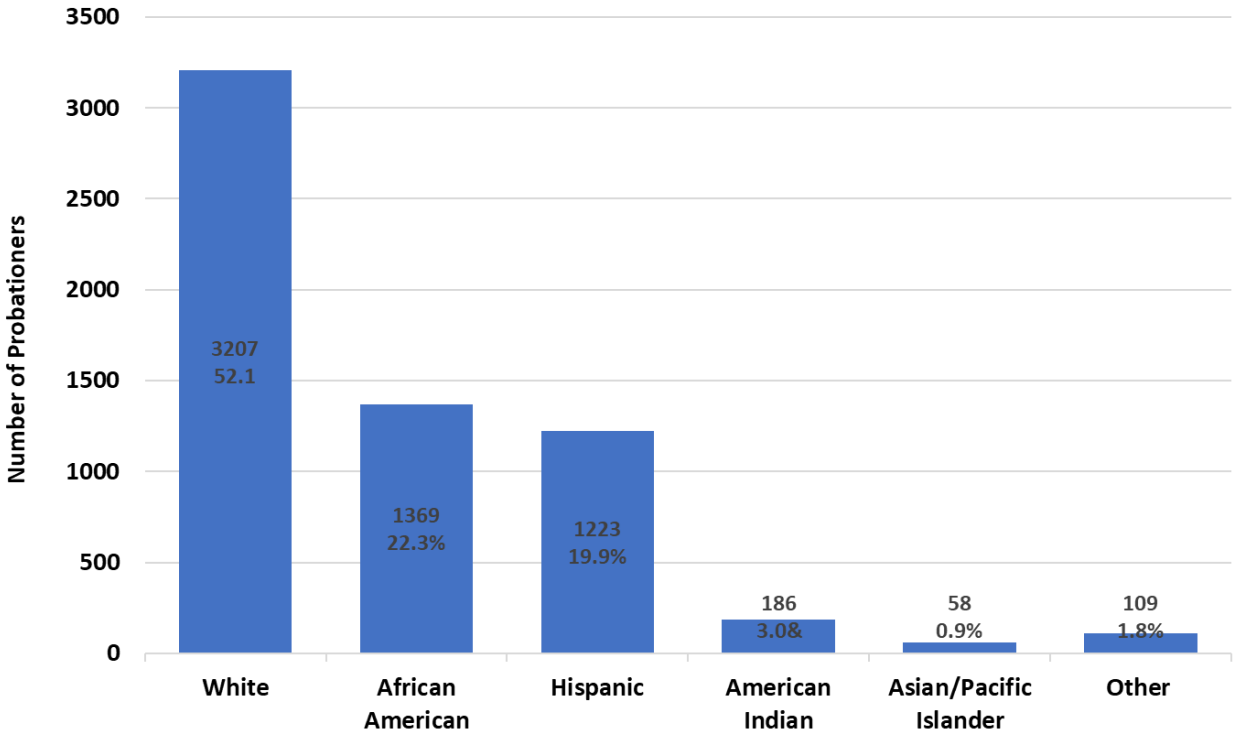


Figure 4. Self-reported Ethnicity of the Youth Included in the Sample



A contingency table assessing probationers' self-reported race crossed with self-reported ethnic categories revealed that the majority of those identifying as "other" in the race category were accounted for by the "Hispanic" ethnic category. After combining the ethnicity and race variables, the final racial-ethnic combined category was as follows: White, non-Hispanic (52.1%); African American, non-Hispanic (22.3%); Hispanic (19.9%); American Indian, non-Hispanic (3%); Asian/Pacific Islander, non-Hispanic (.9%); and Other (1.8%). Figure 5 displays the race/ethnicity breakdown of the full sample.

Figure 5. Self-reported Race and Ethnicity Breakdown of the Youth Included in the Sample – Minority Status



The youth in the sample committed a variety of charged offenses associated with their first YLS/CMI administration. Table 1 displays in order of occurrence the number and percent of the total probationers for each type of index offense associated with the first YLS/CMI score in the data file. The five most frequent offenses, in order, were Juvenile – 3B, Misdemeanor 1, Misdemeanor 2, Infraction, and Misdemeanor 3 accounting for 81.5% of the sample.

Table 1. Distribution of Charge Classification at First YLS/CMI Score

Classification of Charge	Number of Probationers	Percentage of Probationers
Juvenile – 3B	1772	28.8
Misdemeanor 1	1007	16.4
Misdemeanor 2	813	13.2
Infraction	730	11.9
Misdemeanor 3	689	11.2
Felony 3	289	4.8
Felony 4	275	4.5
City Ordinance	240	3.9
Felony 2	90	1.5
Misdemeanor W	51	.8
Felony 3A	31	.5
Misdemeanor 5	17	.3
Felony	8	.1
Felony 1D	4	.1
Misdemeanor 3A	3	0
Felony 1B	2	0
Felony 1C	2	0
Felony 2A	2	0
Misdemeanor	2	0
Misdemeanor 4	2	0

RESULTS

RESULTS SECTION OVERVIEW

To assess how well the YLS/CMI functions in the state of Nebraska, the UNLLP research team set out to answer four primary questions: (1) Does the YLS/CMI demonstrate predictive validity in the Nebraska juvenile probation population? That is, are probation officers administering the YLS/CMI in the most optimal manner so that the YLS/CMI validly discriminates between youths who are successful and those who fail in probation and return to probation? This validity question asks about the ***main effects*** of the YLS/CMI on probation outcome measures. (2) What factors moderate the validity of the YLS/CMI, that is, are there

some types of youth for whom the YLS/CMI is valid (predicts outcomes well) and others for which it is not valid (does not predict outcomes well)? This moderation question asks about the ***interaction effects*** of the YLS/CMI with youth gender and youth race/ethnicity. 3) What domains in the total YLS/CMI score best predicts probation outcomes? That is, do some of the YLS/CMI domains contribute more to the instrument's validity than others contribute? 4) What are the strongest scoring cut-offs for assigning the YLS/CMI risk levels in the Nebraska population? This analysis requires the use of AUC Reviewer Operating Characteristic Curves statistics and follow-up logistic regressions to determine any differences in predicting outcome as a function of variations in total YLS/CMI cutoffs.

For each of the success-failure outcome variables that produced a variable and testable distribution, logistic regression evaluated the extent to which the YLS/CMI total score and the YLS/CMI risk levels predicted likelihood of failure. In each analysis, we adjusted for the amount of time an individual was in the system with the inclusion of the "possible time in system" variable which controlled for how long the youth had from the time of the first charge until the youth's 19th birthday. UNLPP presents the results of the validity analyses for each outcome variable with tables in the text showing the results of the logistic regression analysis and a chart that displays the probability of failure for each YLS/CMI risk level. The text explains the tables and figures as they come up in the report.

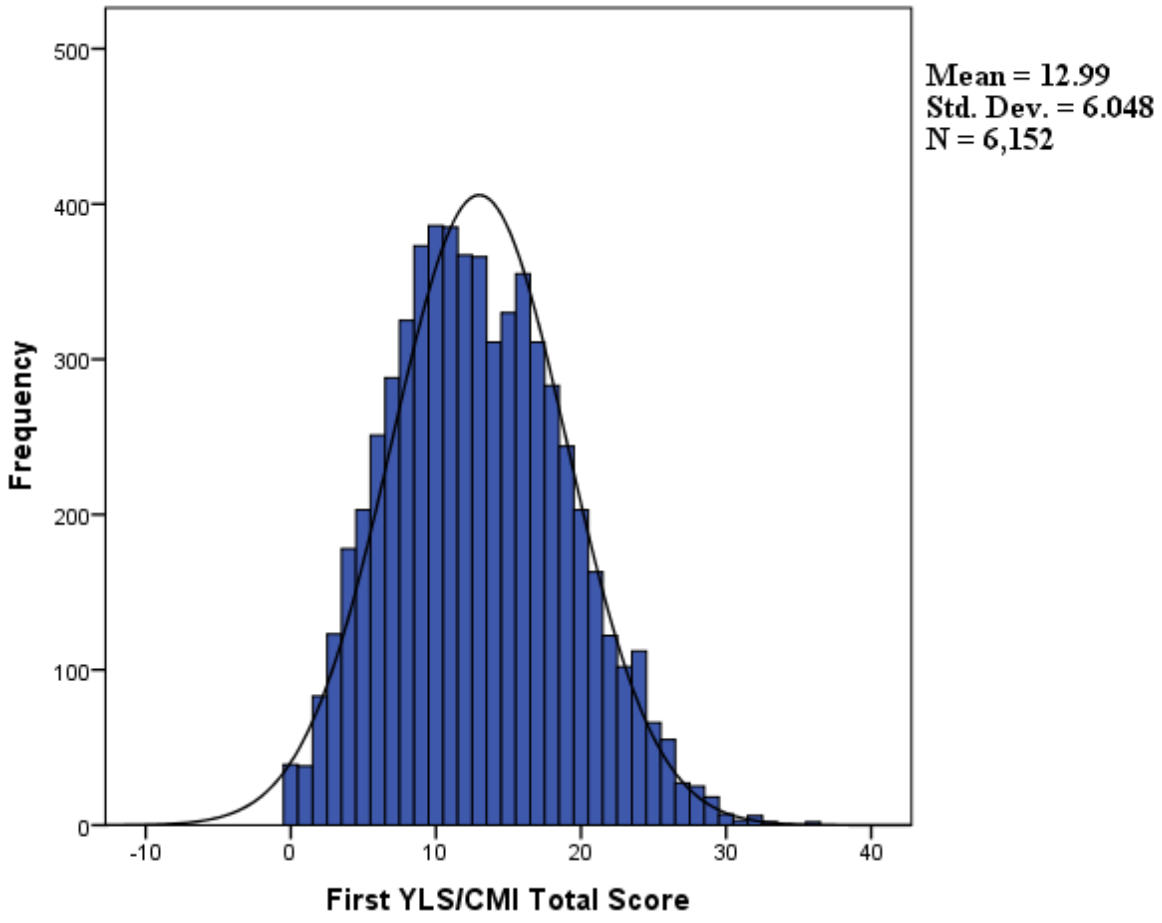
To test for moderation, analyses added the effects of youth gender and youth race/ethnicity into the logistic regressions – to ask whether the YLS/CMI risk levels predicted outcomes differently for boys, girls, Whites, Blacks, and Hispanics. Results from each analysis appear in tables and figures, accompanied by an explanation of the findings. Next, logistic regressions tested the contributions of each of the 8 criminogenic factors for predicting outcome

(success/failure). The text and tables present the overall results followed by a graph illustrating the findings and the contributions of each domain factor to the predictive validity. Finally, AUC ROC statistics offered a possible cutoff score change for the YLS/CMI risk levels followed by an additional logistic regression to test the impact of the alternative cutoff. Again the text displays the results in tables, figures and verbal explanations.

RESULTS SECTION: YLS/CMI PREDICTIVE VALIDITY

YLS/CMI total scores versus YLS/CMI Risk levels. The total scores for the YLS/CMI for the full sample (first occurring score) produced an approximate normal curve (see Figure 6) with mean = 12.99, median = 13.00, mode = 10, standard deviation = 6.05, variance = 36.58 and the standard error of the mean equal to .077. The minimum total score in the sample was 0 (*n* = 39) and the maximum score was 36 (*n* = 2).

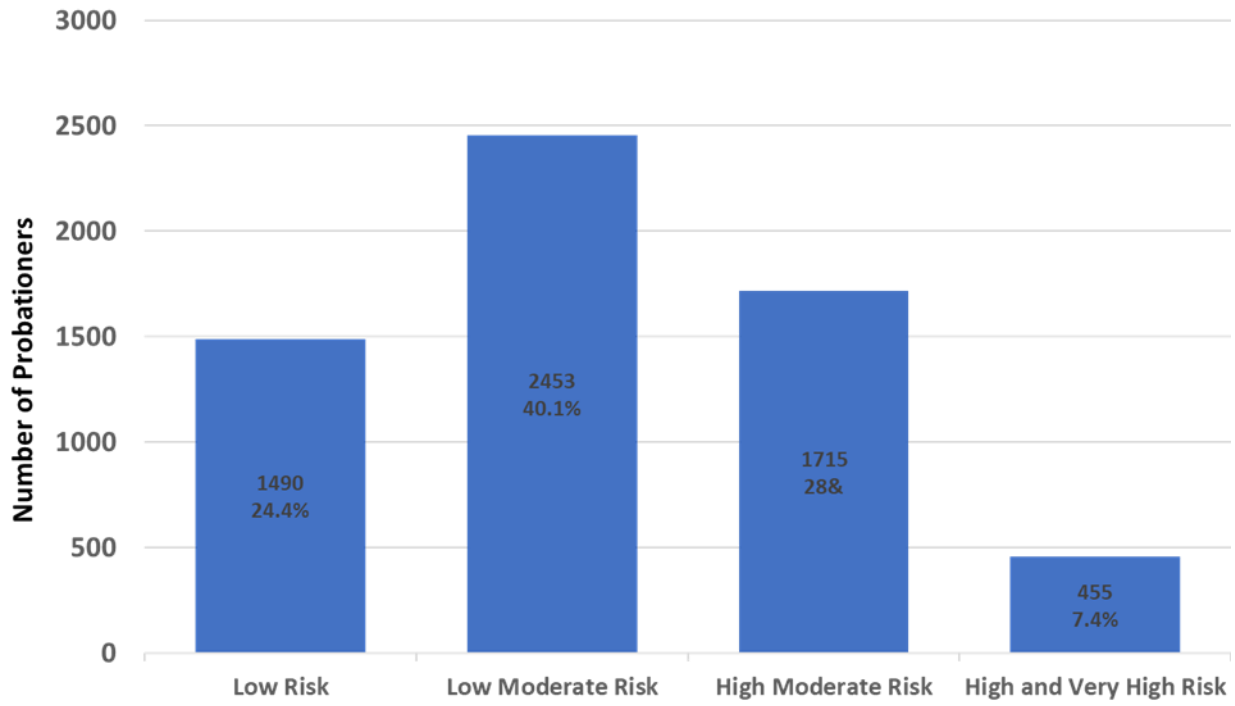
Figure 6. Distribution of YLS/CMI Scores for all the Youth Included in the Sample



Nebraska probation expanded the original three risk categories (i.e., low, medium and high) recommended for the total score (Hoge & Andrews, 2002) into five levels of risk, which UNLLP adopted for the analyses reported below: low (0 to 8), low moderate (9 to 15), high moderate (16 to 22) high (23 to 34) and very high (35 to 42). The number of youths in each category were: low, *n* = 1490; low moderate, *n* = 2453; high moderate, *n* = 1715, high = 453, very high, *n* = 2. The analyses that used total YLS/CMI risk level collapsed the high and very high group into one category because of the low sample size in the very high category. Figure 7 displays the breakdown of YLS/CMI risk levels for the analyses reported below. The validity

analyses that follow use both the raw score for total risk as well as the four category risk level. The number of youth in each category varied slightly from analysis to analysis because of missing data on some of the factors in each of the analyses.

Figure 7. YLS/CMI Total Risk Levels for all Youth Included in the Sample

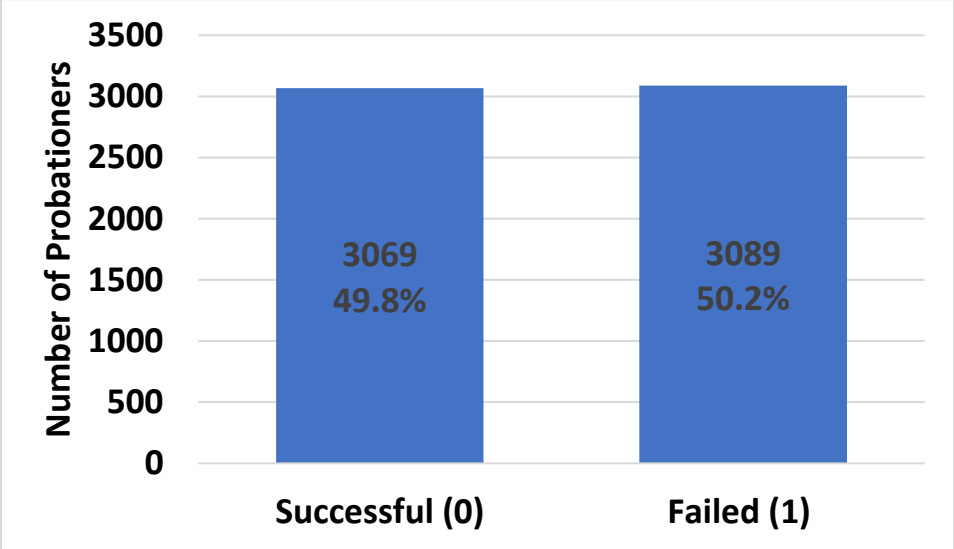


Analysis of Success/Failure. In order to measure the predictive validity of the YLS/CMI, UNLPP required outcome factors to use as criteria. There are many ways to calculate the outcome factor with the data in the NPACS file. The dataset included each youth's first arrest date and first disposition date, which was associated one or more cases, each with multiple charges, and each with a probation outcome (success vs. failure). The dataset also included subsequent juvenile cases, up to ten cases with as many as six offenses for each case. Each of these subsequent cases included an arrest date, a disposition date, and a discharge date. Making the task even more complicated case 1 through case 10 did not appear in the data set in the order

of occurrence, that is, case 3 could have occurred before case 2. UNLPP used all available information to calculate several different outcome measures.

Pure Success. The first definition, *pure success,* included any youth with a success at the first disposition and no further occurrences in the database. Such a youth would have successfully completed one probation and never returned to probation. UNLPP assigned a 0 to every youth who satisfied these requirements and a 1 to any youth that did not (failure). Failed youth with this definition of outcome (pure success) were those who either did not successfully complete their first probation and/or youth who returned to probation regardless of the first disposition. Figure 8 depicts the numbers of probationers classified as successful and not successful using this most conservative definition that combines probation outcomes with *probation* recidivism. Almost half the sample succeeded and the other half failed by this definition. (Note: this is not a pure recidivism measure because it does not include measures of new adjudications or adult convictions but instead includes any youth who returns to probation. UNLPP recently received compiled JUSTICE data for new adjudications and adult convictions and is in the process of analyzing these data, which will include a true recidivism measure.)

Figure 8. Number of Successes and Failures for Pure Success Defined in terms of Probation Outcomes and Probation Returns



Possible time in the System. The present data analyses were adjusted to account for the maximum potential time a youth could remain under the juvenile probation jurisdiction. In other words, we limited our assessment of youth outcomes to the time between the youth's index offense and the date of the individual's nineteenth birthday. UNLPP created a variable, "Possible time in the system", which accounted for the fact that some youth, those discharged from their first probation at a young age, as compared, to those discharged near 19 years of age could have a great deal more time to reoffend. The variable PTIS is the number of days the youth has left in the system from the time of his or her first discharge until he or she turns 19 years old.

The first logistic regression analysis predicted failure on this pure success outcome criterion, using two predictors: possible time in the system (PTIS) to control for the different ages at the time of the first YLS/CMI and the YLS/CMI total score. Table 2 displays the results of the analysis. The first column in this and all other logistic regression tables in the report lists the predictor, column two lists the regression weights for that predictor, column three the

standard error, column four/five list the test statistic (with significance levels) and the accompanying degrees of freedom, and finally column 6, the Odds Ratio of each predictor. The Odds Ratio is explained in more detail below. The superscripts attached to the Wald statistics in the table indicate if the effect is significant with "ns" indicating not significant and superscript asterisks indicating significance. The number of asterisks identifies the significance level, which can change from analysis to analysis, so that the Note at the bottom of the table explains precisely what the number of asterisks means for any given table. The Constant row shows the constant term in the final logistic regression analysis and is not important for the purpose of testing validity effects. All logistic regression tables in this report appear with the above defined features and characteristics.

Row two of Table 2 shows that YLS/CMI predicted failure significantly ($p < .001$) after controlling for PTIS in row one. PTIS is significant in this and all other logistic regression analysis in this report showing that the more possible time that youths could stay in the system, the more likely they are to fail probation and/or return to probation. PTIS is a control variable that will not be discussed further in this report. The summary at the bottom of the table shows that the logistic regression model was significant and explained approximately 10% of the outcome variability (*Nagelkerke* $R^2 = .105$). Most important the effect size for the YLS total score was $r = .29$, a moderate sized effect, demonstrating that with regard to failing at pure success, the YLS/CMI was a solid predictor. Furthermore, the odds ratio (1.106) shows that with each increase in a YLS/CMI point toward the total score, the odds of failure on the pure success measure increases about 1.1 times.

Table 2. Predictive Validity of the YLS/CMI Total Score for Pure Success Outcome–Predicting Failure (*N*=5782)

Predictor	Beta	S.E.	Wald	d.f.	O.R.
Possible Time in System	.0001	.000	7.846*	1	1.000
YLS Total Score	.101	.005	430.223**	1	1.106
Constant	-1.512	.080	358.578**	1	.220

Note: Model χ^2 (2) =478.527, $p < .001$; Nagelkerke $R^2 = .105$; $r = .29$; *$p < .01$. **$p<.001$.

The second logistic regression replaced the YLS Total Score with the YLS risk level. There are four risk levels (low, low moderate, high moderate and high/very high risk). Here the YLS/CMI risk level line measures the effect of the three YLS levels against the low risk level, with each subsequent line showing the effect of that level (low moderate, high moderate, and high/very high) compared to the lowest level. This comparison is used throughout the text in this report to represent the effect of the YLS/CMI risk level across all four levels in the logistic regression analysis. Table 3 shows that after controlling for PTIS, the four risk level factor significantly predicted failure on the pure success variable ($p < .001$). The full model is significant and most importantly, an *r* value can be estimated at $r = .26$ for the risk level factor. (Note: translation of logistic regression total scores into r values is more accepted than translation of dichotomous factors with multiple levels.) Furthermore the odds ratios show that the odds of a youth in the low moderate risk level to fail is 1.98 times more than one in the low risk level, one in the high moderate level is 3.81 times more likely, and the odds of youth to fail in the high/very high category is over five times more likely (5.19).

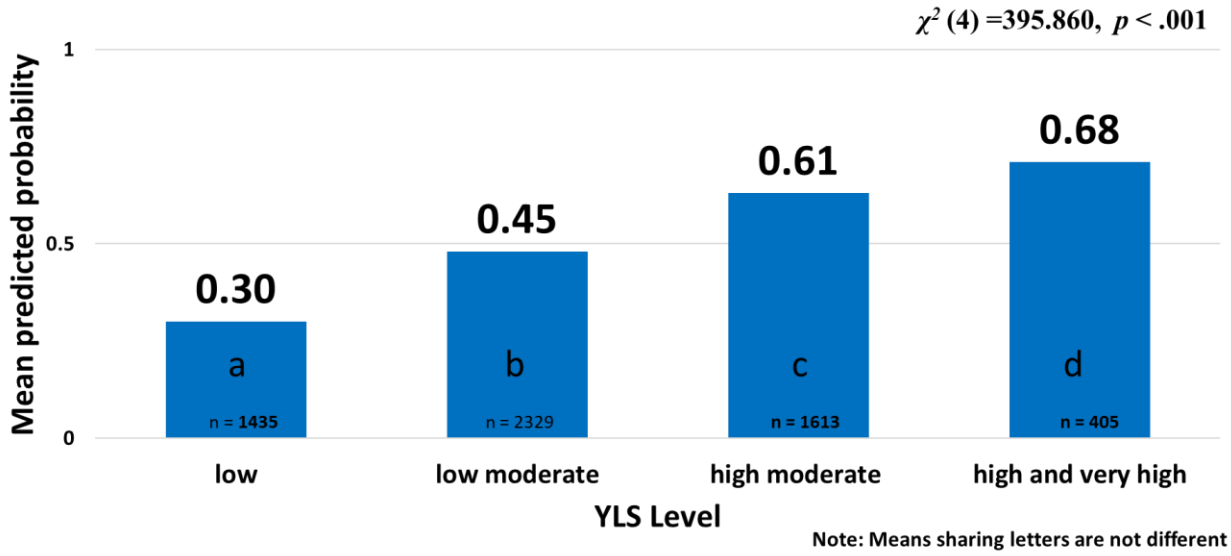
Table 3. Predictive Validity of the YLS/CMI Risk Level for Pure Success Outcome–Predicting Failure (N=5782)

Predictor	Beta	S.E.	Wald	d.f.	O.R.
Possible Time in System	.000	.000	5.572*	1	1.000
YLS/CMI Risk Level			369.802**	3	
Low Moderate vs. Low	.681	.071	90.933**	1	1.975
High Moderate vs. Low	1.338	.077	298.649**	1	3.810
High and Very High vs. Low	1.646	.122	183.215**	1	5.188
Constant	-.958	.069	191.379**	1	.384

Note: Model χ^2 (4) =395.860 $p < .001$; Nagelkerke $R^2 = .088$; $r = .26$; *$p < .05$, **$p < .001$

Figure 9 provides a graphic view of Table 3 showing the probability of failure on the pure success outcome for each level of the YLS/CMI risk level among Nebraska Youth in Probation. Each risk level produces a probability of failure that is significantly different from each other level such that youth in the low risk level have a probability of failure at 30% while youth in the high and very high level have a probability of failure at 68% (two-thirds of youth at the highest or very high level of risk failed on the pure success measure). Notice the step function of the graph demonstrating at increases in level of risk steadily increases likelihood of failure.

Figure 9: Mean Probability of Failure at Each YLS/CMI Risk Level for the Pure Success Outcome (N = 5782)



In summary, the YLS/CMI predicts failure on the pure success outcome at a moderate and significant level (r = .29, p < .001) after controlling for time in the system. This demonstrates moderate predictive validity for the tool. Furthermore as Nebraska computed risk level increases so does the probability of failure. This shows good evidence for the validity of the YLS/CMI as Nebraska Probation officers administer and interpret it.

Partial Success. The second definition, *partial success,* included any youth with a success at the first disposition and who did return to probation but with a success disposition. For example, if a youth was successful on their first disposition, but returned to probation and was successful on any and all his ensuing dispositions, this would be considered a partial success. Again, a youth received a zero if he or she satisfied this definition (success) and 1 if he or she did not (failure). Failed youth with this definition of outcome (partial success) were those who either did not successfully complete their first probation and/or those who did successfully complete the first probation, but returned to unsuccessfully complete at a later disposition.

Figure 10 depicts the number of probationers classified as successful and not successful using this more liberal definition that combines probation outcomes with probation recidivism. Two-thirds of the sample succeeded and one third failed by this more liberal definition.

Figure 10. Number of Successes and Failures for Partial Success Defined in terms of Probation Outcomes and Probation Returns

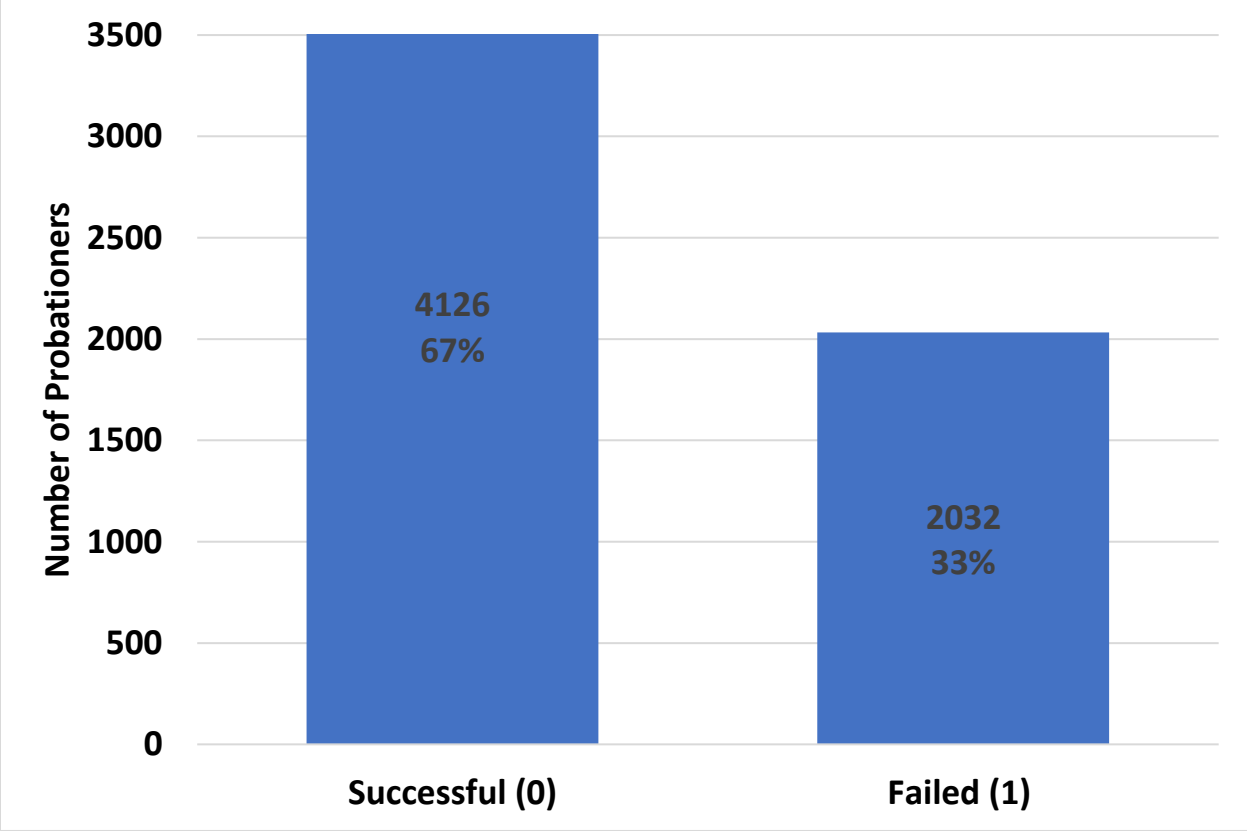


Table 4 shows the logistic regression predicting failure on partial success with YLS/CMI total scores and PTIS. There were two significant main effects one, for PTIS and most importantly a significant effect for YLS total scores with a slightly lower but still low moderate effect size (***r*** = .26). Again the model was significant and explained about 10 percent of the variability in the partial success outcome factor. Once again, the odds ratio shows that with each increase in a YLS/CMI point toward the total score the likelihood of failure on the success measure increases about 1.1 times.

Table 4. Predictive Validity of the YLS/CMI Total Risk Score for Partial Success Outcome – Predicting Failure (N=5782)

Predictor	Beta	S.E.	Wald	d.f.	Exp(B)
Possible Time in System	.000	.000	18.404*	1	1.000
YLS Total Score	.102	.005	382.238*	1	1.107
Constant	-2.105	.090	549.636*	1	.122

Note: Model χ^2 (2) =451.255, $p < .001$; Nagelkerke $R^2 = .106$; $r = .26$; *$p < .001$

Table 5 shows the results of the logistic regression that examined the predictive validity of the YLS risk level, predicting failure for the partial success outcome. Table 5 shows that after controlling for PTIS, the four risk level factor significantly predicted failure on the partial success variable ($p < .001$). The full model is significant and most importantly, the estimated $r = .25$ for the risk level factor. Furthermore, the odds ratios show that the odds of a youth to fail in the low moderate risk level is 1.98 times more than one in the low risk level, one in the high moderate level is 4.09 times more likely to fail, and a youth in the high/very high category is over si times (6.12) more likely to fail.

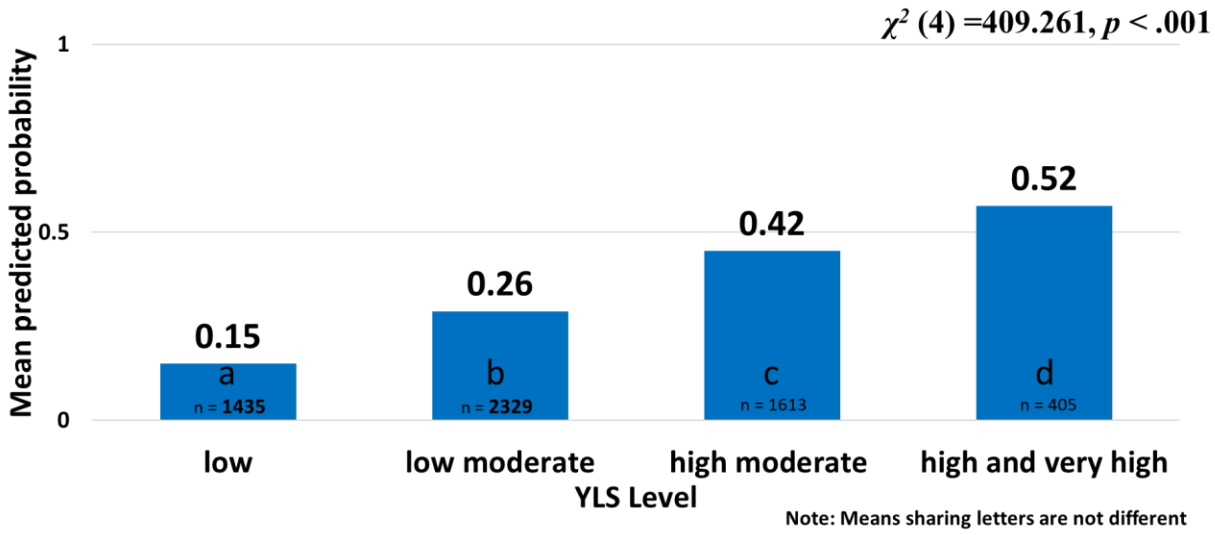
Table 5. Predictive Validity of the YLS/CMI Risk Level for Partial Success Outcome – Predicting Failure (N=5782)

Predictor	Beta	S.E.	Wald	d.f.	O.R.
Possible Time in System	.000	.000	21.665*	1	1.000
YLS/CMI Risk Level			348.610*	3	
Low Moderate vs. Low	.684	.089	59.484*	1	1.982
High Moderate vs. Low	1.409	.090	243.252*	1	4.092
High and Very High vs Low	1.811	.125	211.063*	1	6.118
Constant	-1.573	.084	346.970*	1	.207

Note: Model χ^2 (4) =409.261 $p < .001$; Nagelkerke $R^2 = .098$; $r = .25$; *$p < .001$

Figure 11 provides a graphic view of Table 5 showing the probability of failure on the partial success outcome for each category of the YLS/CMI risk levels. Again, each risk level produces a probability of failure that is significantly different from each other level such that youth in the low risk level have a probability of failure at 15% while youth in the high and very high level have a probability of failure at 52%. Again, notice the step function of the graph demonstrating that as increases in level of risk steadily increase so do increases likelihood of failure.

Figure 11: Mean Probability of Failure at Each YLS/CMI Risk Level for the Partial Success Outcome (N = 5782)



In summary, the YLS/CMI predicts failure on the partial success outcome at a low moderate and significant level (r = .26, p < .001) after controlling for time in the system. This again demonstrates moderate predictive validity for the tool. Furthermore as Nebraska computed risk level increases so does the probability of failure. Once again, this shows good evidence for the validity of the YLS/CMI as Nebraska Probation officers administer and interpret it.

Semi-Success and Limited Success. UNLLP considered several additional outcome measures from the NPACS data. First, for *semi-success,* a youth was successful if he or she succeeded on the first disposition but returned to probation and then failed on one or more later dispositions. There were only 116 youth that satisfied this definition of success, so that any further analyses would not have been meaningful. That is, it is unlikely that the YLS/CMI or any other measure could accurately predict less than 2% of the cases in the full sample. *Limited success* treated a youth who succeeded on any disposition as successful regardless if it was the

first case or a later case. There were, in fact, 4938 (80.2%) youth who fit this definition of successful outcomes with 1220 (19.8%) who did not succeed at any dispositions. UNLLP calculated several other types of outcomes but failed to find any in which number of failing cases exceeded 20%.

Figure 12 depicts the numbers of probationers classified as successful because they succeeded on at least one disposition (0) or failed (1) because they did not succeed on any dispositions. About 80.2% of the sample succeeded and 19.8% failed by this definition.

Figure 12. Number of Successes and Failures for "Limited Success" in Probation Outcomes and Probation Returns

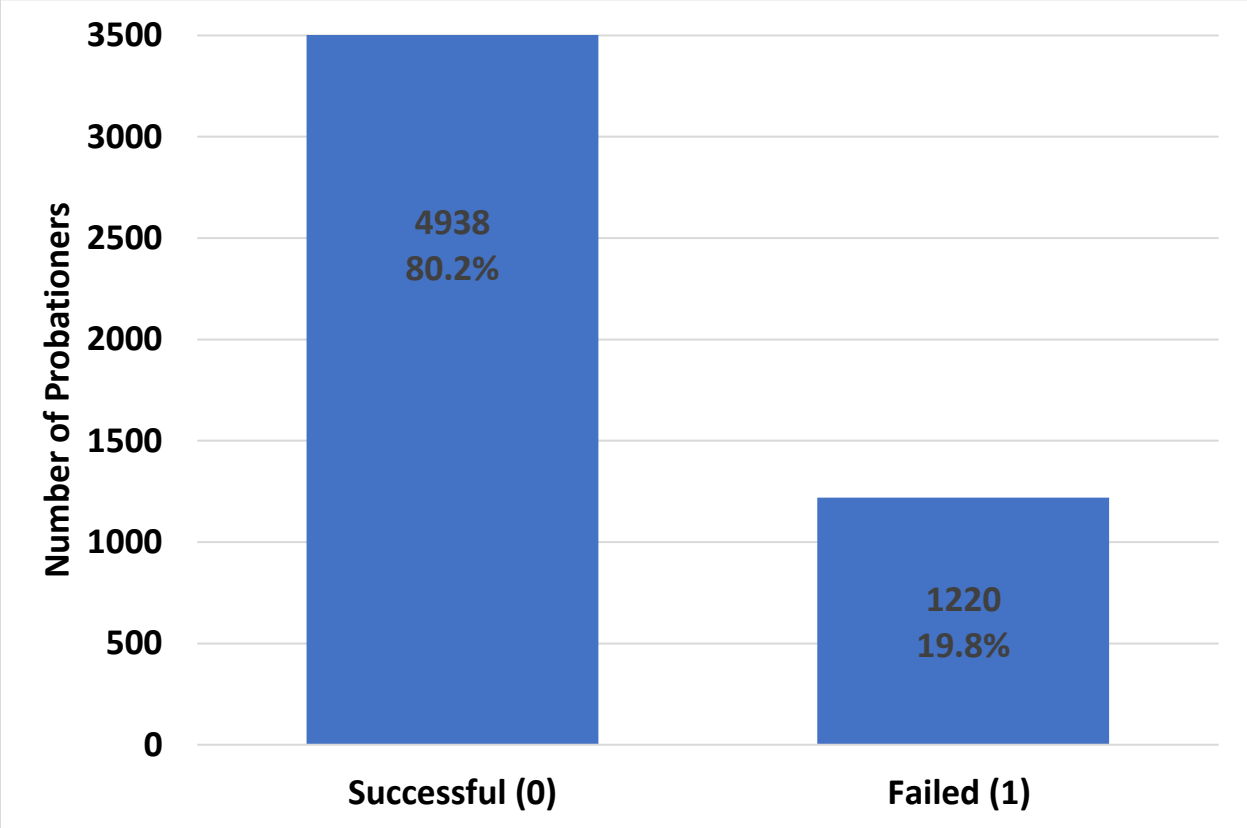


Table 6 shows the logistic regression predicting failure on limited success with YLS/CMI total scores and PTIS. There were two significant main effects, one for PTIS and one for YLS total scores with a small but still significant effect size (***r*** = .18). Again the model was

significant and explained about 10 percent of the variability in the partial success outcome factor. Furthermore, the odds ratio shows that with each increase in a YLS/CMI point toward the total score the odds of failure on the success measure increases about 1.1 times.

Table 6. Predictive Validity of the YLS/CMI Total Risk Score for Limited Success Outcome – Predicting Failure (N = 5782)

Predictor	Beta	S.E.	Wald	d.f.	O.R.
Possible Time in System	-.001	.000	169.044*	1	.999
YLS Total Score	.077	.006	188.318*	1	1.080
Constant	-1.797	.097	344.101*	1	.166

Note: Model χ^2 (2) =418.839, $p < .001$; Nagelkerke $R^2 = .108$; $r = .18$; *$p<.001$.

Table 7 shows the results of the logistic regression that examined the predictive validity of the YLS risk level, predicting failure on the limited success outcome. After controlling for PTIS, the four risk level factor significantly predicted failure on the limited success variable ($p <$ .001), that is, all dispositions were unsuccessful. The full model is significant, but the estimated at $r = .17$ for the risk level factor is a small effect size. The odds ratios show that the odds of failure for a youth in the low moderate risk level is 1.84 times one in the low risk level, the odds of one in the high moderate level to fail is 3.10 times more likely, and a youth in the high/very high category is just over 4 times (4.13) more likely to fail.

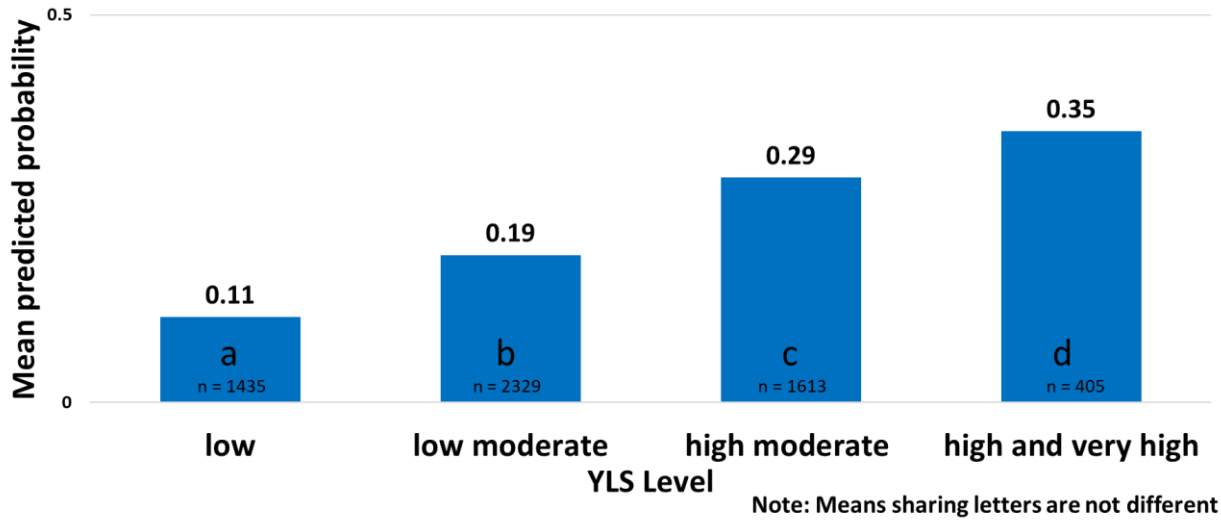
Table 7. Predictive Validity of the YLS/CMI Risk Level for Partial Success Outcome – Predicting Failure (N=5782)

Predictor	Beta	S.E.	Wald	d.f.	Exp(B)
Possible Time in System	-.001	.000	172.106*	1	.999
YLS/CMI Risk Level			168.938*	3	
Low Moderate vs. Low	.613	.101	37.146*	1	1.845
High Moderate vs. Low	1.132	.102	123.243*	1	3.101
High and Very High vs. Low	1.418	.136	108.533*	1	4.128
Constant	-1.460	.094	240.731*	1	.232

Note: Model χ^2 (4) =400.079 $p < .001$; Nagelkerke $R^2 = .104$; $r = .17$; *$p < .001$

Finally, Figure 13 provides a graphic view of Table 7 showing the probability of failure on the limited success outcome for each category of the YLS/CMI risk levels. Again, each risk level produces a probability of failure that is significantly different from each other level such that youth in the low risk level have a probability of failure at 11% while youth in the high and very high level have a probability of failure at 35%. Again, the graph demonstrates a step function showing steady increases in likelihood of failure associated with rising levels of risk.

Figure 13: Mean Probability of Failure at Each YLS/CMI Risk Level for the Limited Success Outcome (N = 5782)



In summary, the YLS/CMI predicts failure on the limited success outcome with a small but significant effect size (r = .18, p < .001) after controlling for time in the system. This again demonstrates predictive validity for the tool showing that as the Nebraska computed risk level increases so does the probability of failure. This analysis shows some evidence for the validity of the YLS/CMI as Nebraska Probation officers administer and interpret it to predict limited success but the YLS/CMI is a better predictor of pure success and partial success failures.

RESULTS SECTION: MODERATION OF YLS/CMI -- PREDICIVE VALIDITY

The need to examine moderation. Monahan and Skeem (2016) argue that we have understudied the instruments that measure criminal risk in youth and adults given the current surge of interest in using risk assessment in adjudication and most importantly in diverting low risk offenders from jail and prison into community corrections. They argue that the criminal justice system ought to carefully consider whether existing risk instruments may give rise to disparities in imprisonment, probation and parole because of the racial or economic limitations or

even bias in the instruments that are in use. Indeed, Olver et al.'s (2014) meta-analysis of the LSI family of instruments showed differences in recidivism effect sizes for the LS instruments across males (.30), females (.30), non-minorities (.32) and especially minorities (.23). The purpose of this section of the report is to examine the differences in predicting outcomes for the YLS/CMI in Nebraska according to gender and race.

In statistical terms, evidence of gender moderation – that the YLS functions differently for males versus females – is a significant interaction between the YLS risk (score or level) and gender of the youths. Similarly, evidence of race/ethnicity moderation – the YLS functions differently for members of minorities and non-minorities – is a significant interaction between YLS risk (score or level) and race/ethnicity of the youths. To test for moderation, UNLLP focused on the pure success outcome because it produced the highest validity effect sizes and added gender and race/ethnicity variables into the logistic regressions to test their interactions with the YLS/CMI. First, we consider gender and then race.

Youth Gender. Table 8 shows the logistic regression predicting failure on pure success with YLS/CMI Total Scores, PTIS, gender, and the interaction between gender and the YLS total score. There is a significant effect for YLS total score and one for gender of the youth. The sections above describe in detail the validity of the YLS total scores in predicting failure on pure success so it will not be repeated here. The significant main effect for gender shows that the probability of boys failing probation and/or returning to probation ($p = .50$) is significantly greater than it is for girls ($p = .43$). However, the interaction between gender and the YLS total score is not significant (Beta = .004, Wald = .168 – ns) showing that the YLS total scores predict failure for pure success similarly for boys and girls. Finally, the model was significant ($p < .001$)

and explained about 11 percent of the variability in the pure success outcome with these factors in the analysis but the effect size for the interaction was very small and non-significant, r = .005.

Table 8. Effects of Gender, YLS/CMI Total Risk Score and their Interaction on Pure Success Outcome – Predicting Failure (N = 5782)

Predictor	Beta	S.E.	Wald	d.f.	O.R.
Possible Time in System	.000	.000	8.086**	1	1.000
YLS Total Score	.099	.006	270.800**	1	1.104
Gender	-.331	.144	5.282*	1	.718
Gender * YLS Total Score	.004	.010	.168	1	1.004
Constant	-1.393	.094	218.501**	1	.248

Note: Model χ^2 (4) =502.114, p < .001; Nagelkerke R^2 = .110; r = .005; *p < .05. **p<.01.

Table 9 shows a similar logistic regression result as in Table 8 but replaces the YLS total scores with the YLS/CMI level factor. Once again, there are significant main effects for gender and YLS/CMI but not their interaction (Wald = .745 – ns) again failing to find a moderation effect for gender. The statistical model is significant.

Table 9. Effects of Gender, YLS/CMI Risk Level and their Interaction on Pure Success Outcome – Predicting Failure (N = 5782)

Predictor	Beta	S.E.	Wald	d.f.	Exp(B)
Possible Time in System	.000	.000	5.852*	1	1.000
YLS/CMI Risk Level			240.038**	3	
Low Moderate vs. Low	.660	.089	55.017**	1	1.935
High Moderate vs. Low	1.354	.097	195.331**	1	3.872
High and Very High vs. Low	1.600	.149	115.080**	1	4.951
Gender	-.286	.121	5.559*	1	.752
Gender * YLS/CMI Risk Level			.745^{ns}	3	
Gender * Low Mod vs. Low	.047	.149	.097^{ns}	1	1.048
Gender * High Mod vs. Low	-.057	.161	.125^{ns}	1	.944
Gender * High/V. High vs. Low	.103	.258	.161^{ns}	1	1.109
Constant	-.854	.082	109.602**	1	.426

Note: Model χ^2 (8) =419.928 $p < .001$; Nagelkerke $R^2 = .093$; $r = .01$; *$p < .05$, **$p < .001$

Figures 14 and 15 show graphic displays of the probability of failure on the pure success outcome for each category of the YLS/CMI risk levels – Figure 14 for boys and Figure 15 for girls. The graphs look very similar except that the probabilities of failure at each risk level of the YLS are higher for boys than for girls. Nonetheless, the shape and direction of the step function is very similar, in fact, the relationship between the YLS/CMI risk levels and the probability of failure for boys and for girls is not significantly different as seen in the Table 9 interaction, Gender * YLS Risk Level.

Figure 14: Mean Probability of Failure at Each YLS/CMI Risk Level for the Pure Success Outcome for Boys (N = 3711)

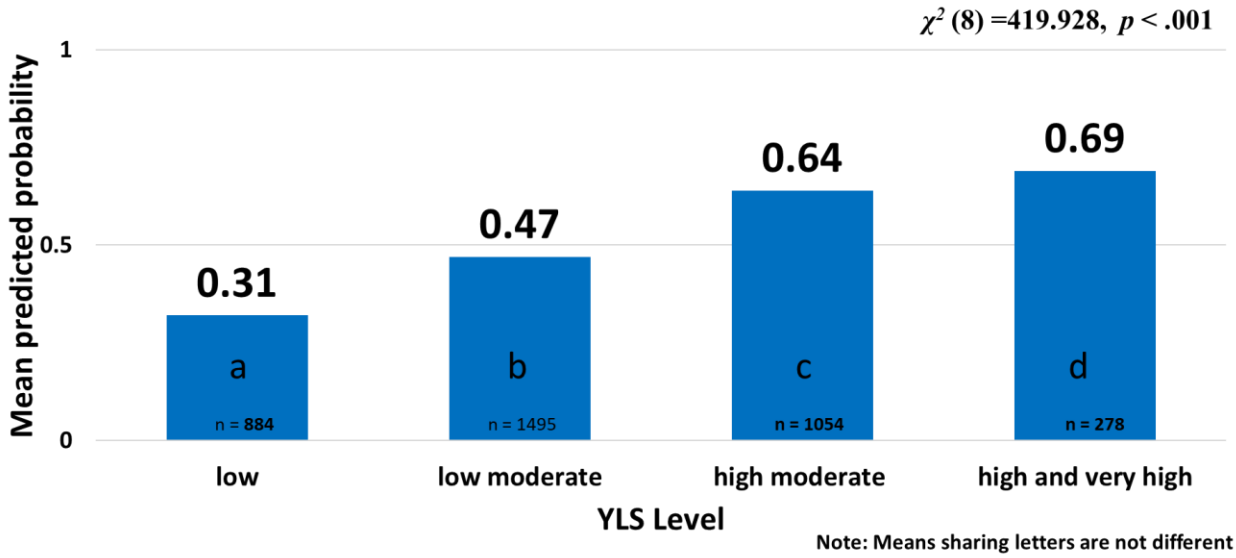
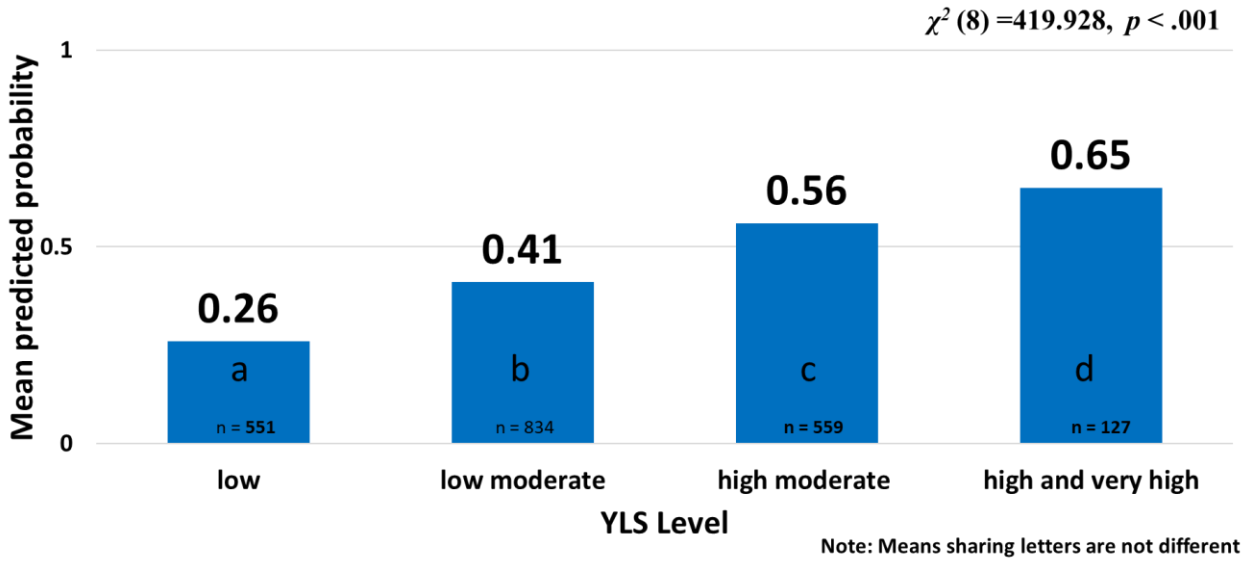


Figure 15: Mean Probability of Failure at Each YLS/CMI Risk Level for the Pure Success Outcome for Girls (N = 2071)



In summary, the YLS/CMI predicts failure on the pure success outcome similarly for boys and girls who are in the Nebraska probation system. The effect size is not statistically

different for boys versus girls, the probability functions are similar (except that boys are more likely to fail and/or return to probation), and in short, there is no evidence for gender moderation or disparate impact of the YLS due to sex of the youth.

Youth Race and Ethnicity. Table 10 shows the logistic regression predicting failure with regard to pure success with YLS/CMI Total Scores, PTIS, minority status (White, Black and Hispanic) and the interaction between minority status and the YLS total score. There is a significant main effect for YLS total score after controlling for PTIS but no other main effects or significant interactions were significant. Race/Ethnicity was coded so that the first line under minority status compares White youth to Black youth and the second line, White youth to Hispanic youth. Most importantly the interaction of minority status and the YLS total score is not significant (*Wald* = 1.294 – ns). The lack of a significant interaction between race/ethnicity and the YLS total score shows that the YLS total scores predict failure for pure success similarly for White, Black and Hispanic youth. Finally, the model was significant ($p < .001$) and explained about 11 percent of the variability in the pure success outcome with these factors in the analysis.

Table 10. Effects of Race/Ethnicity, YLS/CMI Risk Score and their Interaction on Pure Success Outcome – Predicting Failure (N = 5280)[1]

Predictor	Beta	S.E.	Wald	d.f.	O.R.
Possible Time in System	.000	.000	6.989**	1	1.000
YLS Total Score	.100	.007	225.692***	1	1.105
Minority Status			3.869^{ns}	2	
White v. Black	.336	.176	3.660^{ns}	1	1.400
White v. Hispanic	.013	.190	.005^{ns}	1	1.013
Minority Status * YLS Total Score			1.294^{ns}	2	
White v. Black * YLS Total Score	-.001	.012	.012^{ns}	1	.999
White v. Hispanic * YLS Total Score	.014	.014	1.136^{ns}	1	1.015
Constant	-1.617	.103	244.561***	1	.198

Note: Model χ^2 (6) =476.833, $p < .001$; Nagelkerke $R^2 = .113$; *$p < .05$. **$p<.01$, $p < .001$.

Table 11 shows a similar logistic regression result as in Table 10 but replaces the YLS total scores with the YLS/CMI level factor. As usual, there was is a significant main effect for YLS level after controlling for PTIS but now coding YLS as level rather than score produces some other significant effects, including an overall minority status effect. Again, minority status was coded so that the first line under minority status compares White youth to Black youth and the second line, White youth to Hispanic youth. The odds ratios show that the odds of a Black youth failing is 1.5 times the odds of a white youth, all things being equal and an Hispanic youth's odds of failing is 1.3 times greater than that of a white youth. The probability of failure for White youth was .44, for African Americans, .52, and for Hispanics, .49. Most importantly, the lack of a significant interaction between minority status and the YLS risk level shows that the

[1] The sample size for this analysis is smaller because those youth who were not White, African American or Hispanic were dropped. There were not enough youth in other race/ethnic categories to include in the analysis.

YLS risk level predicts failure for pure success similarly for White, Black and Hispanic youth (*Wald* = 4.492 – *ns*) . Finally, the model was significant (p < .001) and explained about 10 percent of the variability in the pure success outcome with these factors in the analysis.

Table 11. Effects of Minority Status, YLS/CMI Risk Level and their Interaction on Pure Success Outcome – Predicting Failure (N = 5280)

Predictor	Beta	S.E.	Wald	d.f.	O.R.
Possible Time in System	.000	.000	4.820*	1	1.000
YLS/CMI Risk Level			200.850***	3	
Low Moderate vs. Low	.756	.100	56.930***	1	2.130
High Moderate vs. Low	1.370	.107	163.602***	1	3.934
High and Very High vs. Low	1.691	.166	103.637***	1	5.427
Minority status			9.491**	2	
White vs. Black	.424	.148	7.802*	1	1.513
White vs. Hispanic	.317	.155	4.190*	1	1.372
Minority status * YLS/CMI Risk Level			4.493^{ns}	6	
White vs. Black * YLS * Low Moderate vs. Low	-.145	.182	.634^{ns}	1	.865
White vs. Black * YLS * High Moderate vs. Low	-.094	.196	.229^{ns}	1	.911
White vs. Black * YLS * High/V. High vs. Low	-.231	.317	.532^{ns}	1	.794
White vs. Hispanic * YLS * Low Moderate vs. Low	-.269	.191	1.988^{ns}	1	.764
White vs. Hispanic * YLS * High Moderate vs. Low	.031	.210	.022^{ns}	1	1.032
White vs. Hispanic * YLS * High/V. High vs. Low	.055	.336	.027^{ns}	1	1.056
Constant	-1.123	.090	155.943***	1	.325

Note: Model χ^2 (12) =404.632, $p < .001$; Nagelkerke $R^2 = .097$; $r = .26$; *$p < .05$, **$p < .01$, ***$p < .001$

Figures 16, 17 and 18 show graphic displays of the probability of failure on the pure success outcome for each category of the YLS/CMI risk levels – Figure 16 for White Youth, and Figure 17 Black Youth and Figure 18 for Hispanic Youth. The graphs look very similar except that the probabilities of failure at each risk level of the YLS are higher for Blacks and Hispanics that for Whites. Nonetheless, the shape and direction of the step functions are very similar, in fact, the relationship between the YLS/CMI risk levels and the probability of failure was not significantly different in any of the minority status conditions as seen in the Table 11 interaction Minority Status * YLS Risk Level.

Figure 16: Mean Probability of Failure at Each YLS/CMI Risk Level for the Pure Success Outcome for White Youth (N = 3028)

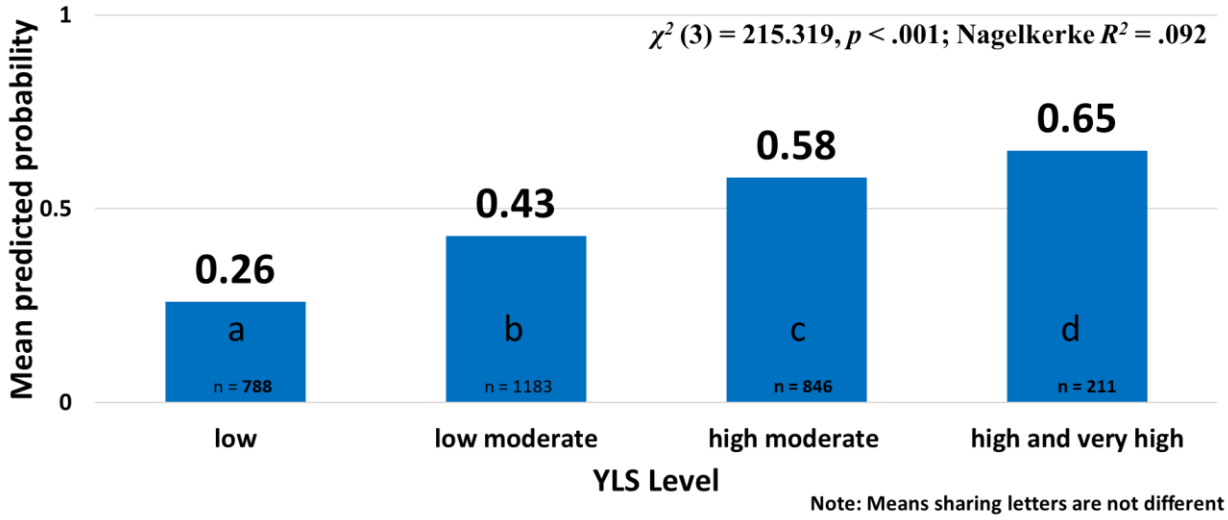


Figure 17: Mean Probability of Failure at Each YLS/CMI Risk Level for the Pure Success Outcome for African American Youth (N = 1280)

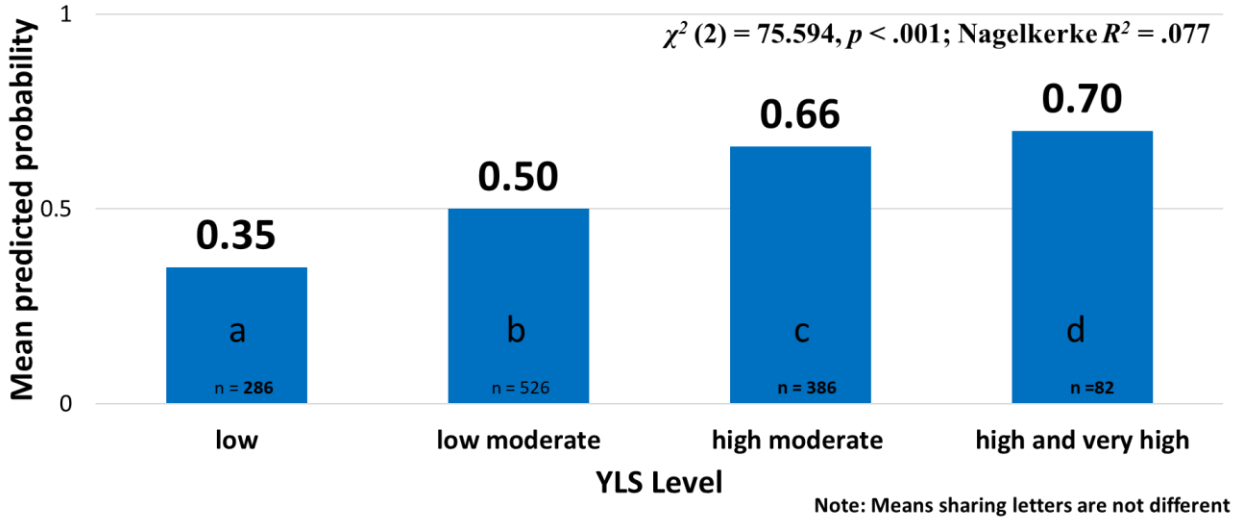
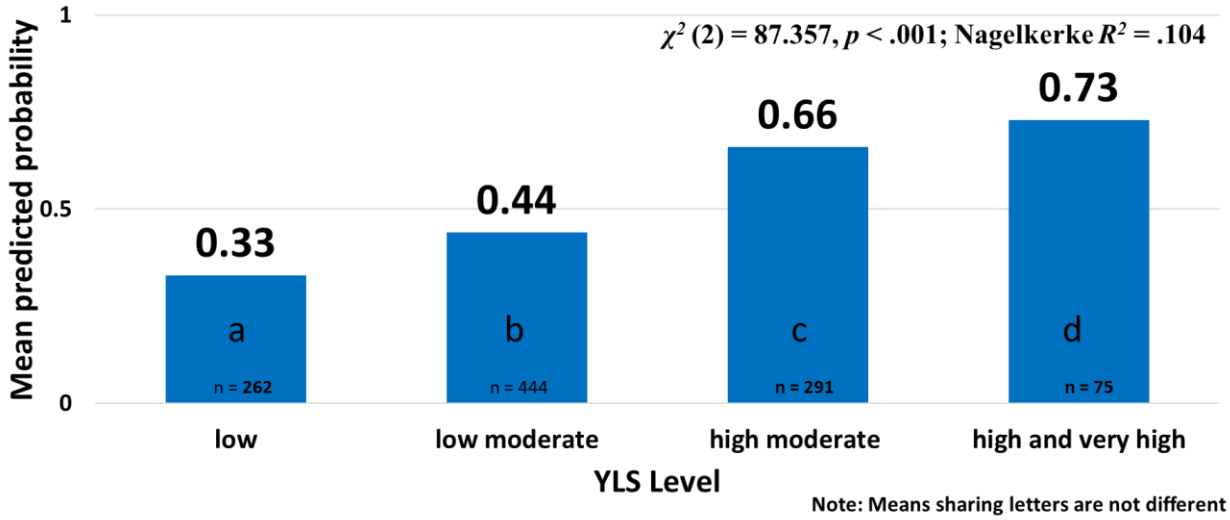


Figure 18: Mean Probability of Failure at Each YLS/CMI Risk Level for the Pure Success Outcome for Hispanic Youth (N = 1072)



In summary, the YLS/CMI predicts failure on the pure success outcome similarly for White, African American and Hispanic youth in the Nebraska probation system. The effect size is not statistically different for minority status, the probability functions are similar (except that White youth are less likely to fail probation and/or return to probation), and in short, there is no evidence for minority moderation or disparate impact of the YLS due to race or ethnicity of the youth.

RESULTS SECTION: YLS/CMI DOMAIN STATUS - PREDICIVE VALIDITY

The next step in the YLS/CMI analysis was to determine which YLS/CMI domains were the strongest predictors of failure with regard to pure success. A series of new logistic regressions were necessary to determine which of the LS/CMI risk level domains contributed the most to the outcomes. First, a logistic regression with all eight criminogenic domains (Attitudes and Orientation, Education and Employment, Family Circumstances and Parenting, Leisure and Recreation, Personality and Behaviors, Prior and Current Offenses, Peer Relations, and

Substance Abuse) after possible time in the system allowed a competitive comparison of the contributions of each domain to total risk of failure and/or return to the system. Table 12 displays the results showing that after controlling for PTIS, all domains except for Attitudes and Orientation (AO) and Leisure and Recreation (LR) contributed significant main effects to the model, Model χ^2 (9) = 584.904; $p < .001$; Nagelkerke $R^2 = .128$.

Table 12. Effects of Minority Status, YLS/CMI Risk Level and their Interaction on Pure Success Outcome – Predicting Failure (N = 5818)

Predictor	Beta	S.E.	Wald	d.f.	O.R.
Possible Time in System	.000	.000	12.006**	1	1.000
Attitudes and Orientation (AO)	.045	.032	1.986^{ns}	1	1.047
Education and Employment (EE)	.177	.021	71.807***	1	1.194
Family Circumstances and Parenting (FCP)	.077	.022	12.536***	1	1.080
Leisure and Recreation (LR)	.052	.032	2.702^{ns}	1	1.053
Personality/Behavior (PB)	.044	.020	5.023*	1	1.045
Prior and Current Offenses (PCO)	.464	.043	118.978***	1	1.591
Peer Relations (PR)	.144	.026	31.226***	1	1.155
Substance Abuse (SA)	.073	.022	11.179**	1	1.076
Constant	-1.582	.086	338.789***	1	.206

Note: Model χ^2 (9) = 584.904; $p < .001$; Nagelkerke $R^2 = .128$; ns = not significant; *$p < .05$, **p < .01, *** p < .001

The next logistic regression analysis included only the six criminogenic factors that showed significant main effects to the outcome when all the domains were included (i.e., (Education and Employment, Family Circumstances and Parenting, Personality and Behaviors, Prior and Current Offenses, Peer Relations, and Substance Abuse). This analysis dropped out the weakest predictors, those with non-significant main effects after controlling for the other domains. Table 13 displays the results of this analysis showing Substance Abuse (SA) was not

significant and dropped out of the model, Model χ^2 (6) = 593.468; $p < .001$; Nagelkerke R^2 = .123; r = .297. Note that the percent of variability in the outcome explained with all eight criminogenic factors was *Nagelkerke* R^2 =.128, (12.8%) while with only five domains value dropped minimally *Nagelkerke* R^2 =*.124*, (12.4%). This suggests that only five domains are needed to optimally predict failure in probation and/or return to probation and the other three add little to that prediction. In fact, an additional analysis not shown here tested the difference between two models, one with all eight factors and one with only the five in Table 13 and found that the difference that the two models contributed to failure outcomes was not significant. Table 13 orders the domains according to their odds ratios which is a measure of the size of their contributions to predicting risk of failure. One implication of this analysis is that officers would do well to address criminogenic needs in the order of: Education and Employment, Peer Relations, Family Circumstances and Parenting, and finally Personality and Behaviors. Prior and Current Offenses does not appear in priority list even though it is the strongest predictor because it is static factor not amenable to treatment, while the others are dynamic or changeable factors.

Table 13. Effects of Minority Status, YLS/CMI Risk Level and their Interaction on Pure Success Outcome – Predicting Failure (N = 5782)

Predictor	Beta	S.E.	Wald	d.f.	O.R.
Possible Time in System	.000	.000	6.874*	1	1.000
Prior and Current Offenses (PCO)	.466	.042	121.026**	1	1.594
Education and Employment (EE)	.190	.020	86.900**	1	1.209
Peer Relations (PR)	.183	.024	57.912**	1	1.200
Family Circumstances and Parenting (FCP)	.050	.019	19.123**	1	1.051
Personality/Behaviors (PB)	.092	.021	6.740*	1	1.097
Constant	-1.473	.080	358.844	1	.229

Note: Model χ^2 (6) = 567.274; $p < .001$; Nagelkerke R^2 = .124; r = .297, *$p < .01$, **$p < .001$

In summary, the YLS/CMI requires the five following domains, in order of their importance to optimally predict failure on the pure success outcome: Prior and Current Offenses, Education and Employment, Peer Relations, Family Circumstances and Parenting, and finally Personality and Behaviors. The other three domains (Attitudes and Orientation, Leisure and Recreation, and Substance Abuse) do not make a significant contribution to the validity of the prediction to the risk of failure. The first of these domains, Prior and Current Offenses, is a static factor while the other four are dynamic factors that are amenable to change. One interpretation of this result is that that officers should prioritize Education and Employment, Peer Relations, Family Circumstances and Parenting, and Personality and Behaviors in their work with the youth in Nebraska.

RESULTS SECTION: ESTABLISHING CUTOFFS WITH ROC ANALYSES

The final section of this report examines the existing cutoff scores that Nebraska probation uses to classify youths into 5 levels of risk on the YLS/CMI (low, low moderate, high moderate, high and very high) and explores an alternative cutoff systems to the one currently in use. Two points of qualification must accompany these analyses. First, the outcome factor that we use for this work was the factor we termed, "pure success" or more accurately failure at pure success. The reader will remember that youth who have failed on this measure either had an unsuccessful outcome on their first case or returned to probation at a later date or both. While return to probation is a type of recidivism, it is not recidivism in the typical sense, which normally refers to new juvenile adjudications or adult convictions for some specific type of crime. UNLLP expects that pure success and recidivism are closely related but they are not identical. UNLLP is in the process of analyzing recidivism data gleaned from the JUSTICE

database in Nebraska which will allow for a more typical definition of recidivism for future analyses of cutoff scores.

Second, setting cutoff scores is an inductive rather than a deductive process, which depends upon one's willingness to accept error balanced against the probability of accurate identification of finding a result. In this case because we are treating failure as a hit – failure is the goal of prediction for higher YLS scores, error is the probability of concluding that a youth is going to succeed after which time, he or she actually does fail. The probability of accurate identification refers to the probability of concluding a youth is going to fail and he or she actually does. The first outcome is called a "false negative" (FN), a type II error or a miss. The second outcome is referred to as a "true positive" (TP) or a "hit". The statistical reality is that cutoffs increase the probability of a true positive at the cost of also increasing the probability of a false negative or a miss so that establishing cutoffs requires decision makers to determine how much they are willing to risk misses in order to better predict hits.

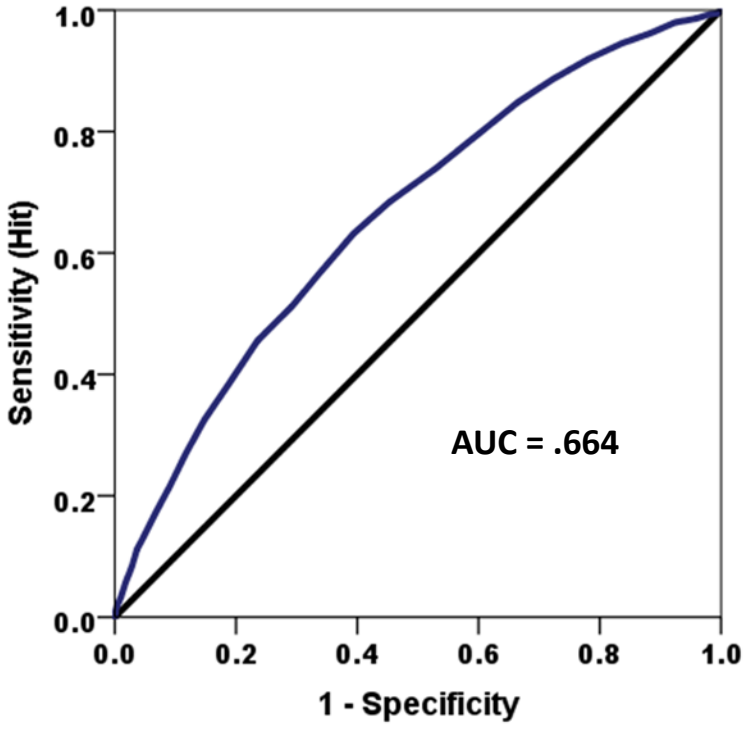
Recall the earlier discussion of the AUC statistic or area under the ROC (reviewer operating characteristic) curve, which plots the false negatives rate of misses (here predicting a success when a failure occurs) on the x-axis of a graph and the true positive rate of a hit (here predicting a failure when a failure actually occurs) on the Y-axis. AUC statistics can help us to select cutoff values for an instrument's scores (here, estimating what cutoff levels to use for YLS/CMI Scores) provided that we know the value of hits and the costs of misses. The AUC statistic (Area Under the Curve) is conceptually (and mathematically) related to the "r" measure, both measures of the strength of a relationship or effect size. The AUC is the probability of an outcome (here failure) for an individual who is selected at random from the population of individuals defined in a specific manner. If a risk instrument has an AUC of .50 then, an

individual selected at random from the group predicted to recidivate has 50% change of recidivating and a 50% change of not recidivating – the prediction is of no value. AUC values must be greater than .50 for the instrument to predict beyond chance. Values of AUC that are between .50 and .56 are small effect sizes, values between .56 and .67 are moderate effect sizes, and those between .67 and .79 are large effect sizes.

One more consideration is the problem of identifying cutoffs from one sample without knowing if they will generalize to a new sample. In other words it is important to compare cutoffs across at least two samples to feel confident that they are not simply the product of a unique ROC curve, one that will be difficult to reproduce. To get around this problem, UNLLP divided the sample of juvenile probationers in the database into two new samples, each determined at random through a random numbers generator: Random Sample 1 with 3075 youths and Random Sample 2 with 3077 youths, all of whom together populated the original database. (Note that this analysis does not control for PTIS for the purpose of studying cutoff points.)

Figure 19 displays the ROC curve for Random Sample 1 with the probability of missing a failure on the X axis and the probability of a hit on the Y axis for the YLS/CMI scores so that each point on the diagonal line represents one of the possible YLS/CMI scores. The diagonal line itself represents a situation where at each level of the probability of a hit, the probability of a miss is the same. If the diagonal line is the real outcome, then the probability of a youth drawn at random failing is actually .50, so that having any information about the YLS/CMI is of no value at all. The area of the square that contains the graph is equal to 1.00, so that the area that the diagonal cuts off is exactly .50. The blue line on the graph is the hit to miss ratio that the ROC analysis calculated for Random Sample 1 for the YLS total scores and shows a moderate effect of .664, which is significantly greater than .50.

Figure 19: The ROC Curve for Random Sample 1: Predicting Failure for Pure Success with YLS/CMI Total Scores (N = 3077)



The ROC AUC analysis computes the probability of hit (True Positive or Sensitivity) and the probability of a miss (False Negative or 1 – Specificity) for all points on the predictor – here scores on the YLS/CMI. Table 14 shows these values for the cutoff points that Nebraska Probation currently uses for the YLS. The first column is the Risk Level Class, the second is the range with the endpoint of the range in bold, the third column is the size of the class for Random Sample 1 at that level of YLS/CMI risk, the fourth column is the probability of a hit for the endpoint score of the class, and the fifth column is the probability of a miss for the endpoint score of the class.

Table 14. The Probability of True Positives and False Negatives for Nebraska Cutoff points on the YLS/CMI for Random Sample 1

Class	Range	N	*p* (True/Pos.) Hit	*P* (False/Neg.) Miss
Low	0-**8**	780	.841	.653
Low Moderate	9-**15**	1255	.449	.228
High Moderate	16-**22**	845	.089	.038
High	23-**34**	194	.001	.000
Very High	35-**42**	1	.000	.000

Thus the probability of a true positive (i.e., accurately predicting failure) for an individual with a score of 8 or higher on the YLS/CMI is .841 but the probability of a false negative (i.e., inaccurately predicting success) is .653 for the same individuals with a score of 8 or higher. Notice that as the cutoffs increase the probability of a hit goes down, that is the probability of successfully predicting a failure for a score of 8 or higher is almost twice that of successfully predicting a failure for a score of 15 or higher. At the same time the probability of a miss goes down so that probability of unsuccessfully predicting success for an individual with a score of 15 or higher is almost 1/3 the false negative rate of making the same prediction of for an individual of a score of 8 or higher. Thus, the probability of predicting a failure for an individual with a score of 22 or higher is .089 but the probability of error is only .038. The goal is to maximize the probability of hits and minimize the probability of misses and distribute the cutoff points to reflect how much error one is willing to accept to obtain a hit. Here there is some asymmetry because there are too few scores in the high and very high group to maximize the predictive validity of the YLS/CMI scores. Figure 20 shows the ROC curve for Random Sample 2, which confirms the AUC effect in a new sample (AUC = .661) and Table 15 shows the probability of

hits and misses for the current Nebraska cutoffs for Random Sample 2, which is nearly the same as in sample 1.

Figure 20: The ROC Curve for Random Sample 2: Predicting Failure for Pure Success with YLS/CMI Total Scores (N = 3077)

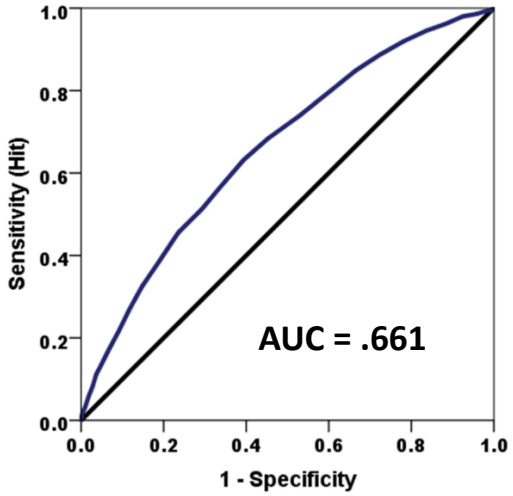


Table 15. The Probability of True Positives and False Negatives for Nebraska Cutoff points on the YLS/CMI for Random Sample 2

Class	Range	N	*p* (True/Pos.) Hit	*P* (False/Neg.) Miss
Low	0-**8**	748	848	.664
Low Moderate	9-**15**	1263	.456	.235
High Moderate	16-**22**	836	.112	.037
High	23-**34**	229	.001	.000
Very High	35-**42**	1	.000	.000

In summary, the Nebraska Cutoff scores stack the hits and errors in the low range of the scale, such that there could be some improvement in prediction of failures if the cutoff scores

assigned lower values of the YLS/CMI greater risk and in general lowered the boundaries of the higher risk levels. Table 16 displays the TP and FN figures for an alternative distribution of cutoff scores for Random Sample 1.

Table 16. The Probability of True Positives and False Negatives for Lower True Positives and False Negative points on the YLS/CMI for Random Sample 1

Class	Range	N	*p* (True/Pos.) Hit	*P* (False/Neg.) Miss
Low	0-**5**	326	.941	.847
Low Moderate	6-**10**	836	.741	.503
High Moderate	11-**16**	1036	.379	.191
High	17-**22**	680	.089	.038
Very High	23-**42**	195	.000	.000

By lowering the cutoffs, this new system distributes hits and misses more evenly across the YLS/CMI scores and distributes the youth more evenly across the risk levels, reflecting more accurately a normal distribution of risk. Table 17 shows confirmation of this outcome for Random Sample 2.

Table 17. The Probability of True Positives and False Negatives for Lower True Positives and False Negative points on the YLS/CMI for Random Sample 2

Class	Range	N	*p* (True/Pos.) Hit	*P* (False/Neg.) Miss
Low	0-**5**	338	.945	.835
Low Moderate	6-**10**	785	.739	.529
High Moderate	11-**16**	1078	.381	.186
High	17-**22**	646	.112	.037
Very High	23-**42**	230	.000	.000

The question that remains for the new set of cutoff scores is whether it alters the predictive validity of the YLS/CMI in the full sample predicting failure on the pure success

outcome. UNLLP computed a new logistic regression analysis using the new cutoff scores as the predictor and failure on pure success, now controlling for PTIS. The results of this analysis appear in Table 18, which should be compared to the original risk level analysis in Table 3. Here the label YLS/CMI Risk Level measures the effect of the four higher YLS levels against the low risk level with each subsequent line showing the effect of that level (low moderate, high moderate, high, and very high) compared to the lowest level. Table 16 shows that after controlling for PTIS, the five risk level factor significantly predicted failure on the pure success variable ($p < .001$). The full model is significant and an r value can be estimated at $r = .27$ ($r = .26$ for the original cutoffs) for the risk level factor. Most importantly, the odds ratios show that the odds of failing for a youth in the low moderate risk level is 1.77 times greater than one in the low risk level, 3.19 times for the low moderate risk youth compared to the low risk youth, 5.39 for a youth in the high level and 7.68 times more likely for a youth in the very high level. Comparing the logistic regression statistics for the original cutoff scores in Table 3 to those in Table 16 shows that the modified cutoff has slightly more predictive validity than the original cutoff scores. Most notably, the odds of a youth in the very high risk level to fail is 7.68 more likely than for one in the low risk level using the modified cutoff, while the odds of failing for a youth in the high/very high risk level using the original cutoffs is only 5.19 times greater than it is for one in the low risk category with those cutoffs. Thus, it would appear that the original cutoffs dilute some of the predictive validity of the YLS/CMI by "bunching" youth into the lower risk categories.

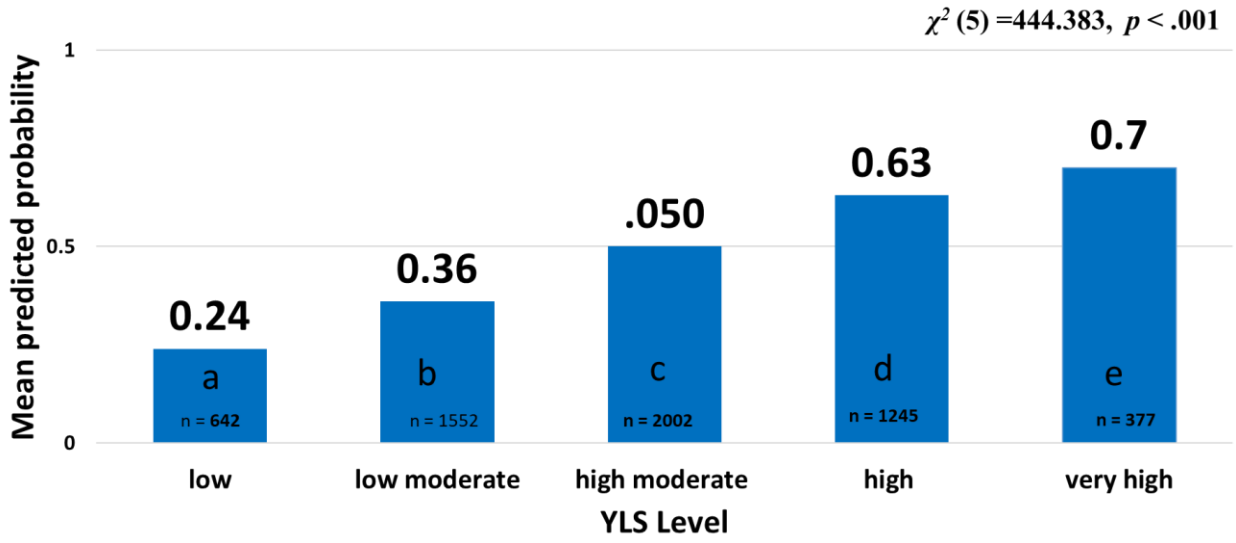
Table 18. Predictive Validity of the YLS/CMI Risk Levels with modified cutoffs for Pure Success Outcome– Predicting Failure (N=5818)

Predictor	Beta	S.E.	Wald	d.f.	O.R.
Possible Time in System	.000	.000	6.381*	1	1.000
YLS/CMI Risk Level			406.802**	4	
Low Moderate vs. Low	.574	.107	29.018**	1	1.776
High Moderate vs. Low	1.161	. .103	127.360**	1	3.192
High vs. Low	1.686	.110	236.272**	1	5.399
Very High vs. Low	2.039	.146	194.488**	1	7.680
Constant	-1.252	.101	154.948**	1	.286

Note: Model χ^2 (5) =444.383 $p < .001$; Nagelkerke $R^2 = .098$; $r = .28$; *$p < .05$, **$p < .001$

Finally, Figure 21 presents the results of the logistic regression analyses in Table 18 graphically showing the probability of failure on the pure success outcome for each level of the YLS/CMI modified cutoff risk levels among Nebraska Youth in Probation. Each risk level produces a probability of failure that is significantly different from each other level such that youth in the low risk level have a probability of failure at 24% while youth in the high and very high level have a probability of failure at 70%. Notice the step function of the curve demonstrating increases in level of risk steadily increases as do the associated likelihood of failure. This graph is comparable to the Figure 9, which demonstrates a similar function for the original cutoff scores.

Figure 21: Mean Probability of Failure at Each YLS/CMI Risk Level with the Modified cutoffs for the Pure Success Outcome (N = 5488)



One final note about setting cutoffs is that there are a very large number of ways to set cutoffs each with their own probability of true positives and false negatives. UNLLP recommends that NAOP and the researchers try several other models – based upon Probation's own values of predicting failure accurately balanced against Probation's tolerance for missing youth who are likely to fail.

EXECUTIVE SUMMARY AND CONCLUSIONS

Introduction. This report details the work that the University of Nebraska/Law Psychology Program (UNLLP) completed to analyze the validity of the Youth Level of Service/Case Management Inventory (YLS/CMI) as Nebraska Probation uses it to assess risk levels of youth in the juvenile justice system. Hoge and Andrews (2002) developed the YLS/CMI as an assessment tool in accordance with Andrew and Bonta's Risk-Need-Responsivity (RNR) model, which emphasizes the importance of tailoring offender treatment to

the needs of adults and youth in the justice system. The YLS/CMI, an adaptation of the LSI-R modified to measure risk and needs of youth in the juvenile justice system, is a 42-item standardized instrument administered in a semi-structured 60 to 90 minute interview with youth that results in individual scores (ranging from 3 to 9) and corresponding risk levels (i.e., low, moderate, and high) for each of eight criminogenic domains (prior and current offenses, family circumstances/parenting, education/employment, peer relations, substance abuse, leisure/recreation, personality/behavior, and attitudes/orientation).

There are a large number of studies that have tested the validity of the YLS/CMI all summarized in Olver, Stockdale and Wormith's 2014 meta-analysis of the then existing studies examining the validity of the LSI measures including the Adult LS/CMI and the YLS/CMI. The meta-analysis included 36 studies of the youth YLS scales, which overall showed a moderate and statistically significant effect size (r = .25). In Canada, the effect size was r =. 33, in the United States, r = .22 and outside North America, r = .28. Effect sizes calculated in this report of the YLS/CMI in Nebraska probation that are at or above r = .22 show that the validity of the YLS/CMI in predicting failures for Nebraska youth is comparable to those established in the rest of the United States.

Description of the Sample and Methodology. Data for this analysis came from the Nebraska Probation archival database, which included 6,158 youth assessed with the YLS/CMI between May, 2007 and November 2015. The median age of the youth at the time of the assessment was 16 years old. Approximately 64% of the sample were boys and 36% were girls. With regard to race and ethnicity, 52% were White, 22% were African American and 20% were Hispanic. The remaining 6% were Native American, Asian/Pacific or other. To test the validity of the YLS/CMI UNLLP calculated several indices of probation outcome identifying success and

failure for each measure and conducted logistic regression analyses to examine the predictive validity of the instrument.

Validity Results. Pure success included any youth with a success at the first disposition and no further occurrences in the database so that a successful youth completed one probation and never returned to probation. UNLP assigned a 0 to every youth who satisfied these requirements and a 1 to any youth that did not (failure). The YLS/CMI predicted failure on the pure success outcome at a moderate and significant level (r = .29), which demonstrates moderate predictive validity for the tool exceeding the aggregated effect size across the United States. As Nebraska computed risk level increased so does the probability of failure demonstrating strong evidence for the validity of the YLS/CMI as Nebraska Probation officers administer and interpret it.

Gender and Race/Ethnicity. Many criminology researchers including Monahan and Skeem (2016) maintain that the field has understudied the instruments that measure criminal risk in youth and adults given the current surge of interest in using risk assessment in adjudication and most importantly in diverting low risk offenders from jail and prison into community corrections. The current view is that decision makers ought to carefully consider whether existing risk instruments may give rise to disparities in imprisonment, probation and parole because of the racial, gender or economic limitations or even bias in these instruments. Indeed, Olver et al.'s (2014) meta-analysis of the LS family of instruments showed differences in recidivism effect sizes for the LS instruments across males (r = .30), females (r =.30), non-minorities (r = .32) and especially minorities (r = .23).

UNLLP conducted several additional logistic regression analyses to examine disparate predictions of probation outcomes for the YLS/CMI in Nebraska. With respect to gender, the

YLS/CMI predicts failure on the pure success outcome similarly for boys and girls, such that the effect size is not statistically different for boys versus girls and the probability functions are similar (except that boys are more likely to fail and/or return to probation), and in short, there is no evidence for disparate impact of the YLS/CMI due to sex of the youth. With respect to race/ethnicity the YLS/CMI predicts failure on the pure success outcome similarly for White, African American and Hispanic youth. Again the difference in effect sizes is not statistically different for minority status and the probability functions are similar (except that White youth are less likely to fail probation and/or return to probation than are African American and Hispanic youth). In short, there is no evidence for disparate impact of the YLS due to race or ethnicity of the youth.

Domain Status – Predictive Validity. UNLLP next conducted a series of new logistic regressions to determine which of the LS/CMI risk level domains contributed the most to the outcomes. These logistic regressions tested the effects all eight criminogenic domains separately on the prediction of failure and/or return to the system. Results showed that the five following domains, in order of their importance optimally predict failure on the pure success outcome: Prior and Current Offences, Education and Employment, Peer Relations, Family Circumstances and Parenting, and finally Personality and Behaviors. The first of these domains, Prior and Current Offenses, is a static factor while the other four are dynamic factors that are amenable to change. One interpretation of this result is that that officers should prioritize Education and Employment, Peer Relations, Family Circumstances and Parenting, and Personality and Behaviors in their work with the youth in Nebraska.

Alternative Cutoffs. The final section of this report examines the existing cutoff scores that Nebraska probation uses to classify youths into 5 levels of risk on the YLS/CMI (low, low

moderate, high moderate, high and very high) and explores the utility of an alternative cutoff system. Two points of qualification must accompany these analyses. First, the outcome factor that we use for this work was the factor we termed, "pure success" or more accurately failure at pure success. While returned to probation is a type of recidivism, it is not recidivism in the typical sense, which normally refers to new juvenile adjudications or adult convictions for some specific type of crime. UNLLP is in the process of analyzing recidivism data gleaned from the JUSTICE database in Nebraska which will allow for a more typical definition of recidivism for future analyses of cutoff scores.

Second, setting cutoff scores is an inductive rather than a deductive process, which depends upon one's willingness to accept error balanced against the probability of accurate identification of finding a result, here a failure in probation. There are many potential cutoffs for the YLS/CMI depending upon the value of predicting failure and the tolerance of error. UNLLP tried one alternative among many possibilities using AUC statistic or area under the ROC (reviewer operating characteristic) curve, which plots the false negatives rate of misses (here predicting a success when a failure occurs) on the X-axis of a graph and the true positive rate of a hit (here predicting a failure when a failure actually occurs) on the Y-axis.

The analyses measured the hit and miss rates for the current cutoffs that Nebraska Probation utilizes for the YLS/CMI and found that they adequately distributed hits and misses across the levels of risk. Using a bootstrap method, UNLLP developed one possible alternative cutoff system and measured that system's hit and miss rates and found it to be more evenly distributed. The validity scores of the alternative system produced a mathematical solution that was slightly more predictive than the existing system of cutoffs. Many other cutoff systems are possible.

Conclusion. This report tested the validity of YLS/CMI in predicting failure in probation and found strong evidence for the validity of the instrument as Nebraska Probation currently uses the tool with juveniles in the system. There was no evidence for disparate impact in predicting outcomes due to either sex of the youth or race/ethnicity of the youth. Five of the 8 criminogenic domains contribute heavily to the predictive validity of the instrument. AUC/ROC analysis measured the hit and miss rates of the existing risk level cutoffs and tested one of many possible cutoff alternatives. UNLLP strongly endorses the continued use of the YLS/CMI with existing or moderated cutoff scores for risk level based upon our analysis of the existing probation data which show that the YLS/CMI is a valid tool for measuring risk of failure in Nebraska youth.

References

Andrews, D. A., & Bonta, J. L. (2003). The psychology of criminal conduct (3rd ed.). Cincinnati, OH: Anderson Publishing.

Andrews, D.A., & Bonta, J. (2010). Rehabilitating criminal justice policy and practice. *Psychology, Public Policy, and Law, 16,* 39-55. DOI: 10.1037/a0018362

Andrews, D.A. & Dowden, C. (2006). Risk principle of case classification in correctional treatment: A meta-analytic investigation. *International Journal of Offender Therapy and Comparative Criminology, 50,* 88-100. DOI: 10.1177/0306624X05282556

Andrews, D. A., Bonta, J. L., & Wormith, J. S. (2004). Level of Service/Case Management Inventory (LS/CMI): An offender assessment system. Toronto: Multi-Health Systems.

Andrews, D. A., Bonta, J., & Wormith, J. S. (2006). The recent past and near future of risk and/or need assessment. *Crime & Delinquency, 52*, 7-27. doi.org/10.1177/0011128705281756

Andrews, D.A., Zinger, I., Hoge, R.D., Bonta, J., Gendreau, P., & Cullen, F.T. (1990). Does correctional treatment work? A clinically relevant and psychologically informed meta-analysis. *Criminology, 28*, 369-404. DOI: 10.1111/j.1745-9125.1990.tb01330.x

Childs, K. K., Ryals, J., Frick, P. J., Lawing, K., Phillippi, S. W., & Deprato, D. K. (2013). Examining the validity of the Structured Assessment of Violence Risk in Youth (SAVRY) for predicting probation outcomes among adjudicated juvenile offenders. *Behavioral Sciences and the Law, 31*, 256-270. DOI: 10.1002/bsl.2060

Clarke, M. C., Peterson-Badali, M., & Skilling, T. A. (2017). The relationship between changes in dynamic risk factors and the predictive validity of risk assessments among youth offenders. *Criminal Justice and Behavior, 44*, 1340-1355. doi.org/10.1177/0093854817719915

Dowden, C., & Andrews, D. A. (1999). What works for female offenders: A meta-analytic review. *Crime and Delinquency, 45*, 438-452. doi.org/10.1177/0011128799045004002

Dowden, C., & Andrews, D. A. (1999). What works in young offender treatment: A meta-analysis. *Crime and Delinquency, 45*, 438-452. doi.org/10.1177/0011128799045004002

Dowden, C., & Andrews, D. A. (2000). Effective correctional treatment and violent reoffending: A meta-analysis. *Canadian J. Criminology*, *42*, 449-467.

Dowden, C., & Andrews, D. (2003). Does family intervention work for delinquents? Results of a meta-analysis. *Canadian Journal of Criminology and Criminal Justice*, *45*(3), 327-342. doi.org/10.3138/cjccj.45.3.327

Garcia-Gomis, A., Villanueva, L., & Jara, P. (2017). Risk factors and youth recidivism prediction in general and property offenders. *Psychiatry, Psychology and Law*, *24*, 308-318. doi.org.libproxy.unl.edu/10.1080/13218719.2016.1247419

Gendreau, P., & Andrews, D. A. (1990). Tertiary prevention: What the meta-analyses of the offender treatment literature tell us about what works. *Canadian Journal of Criminology, 32*, 173-184.

Hoge, R. D., & Andrews, D. A. (2002). Youth Level of Service. *Case Management Inventory*.

Luther, J. B., Reichert, E. S., Holloway, E. D., Roth, A. M., & Aalsma, M. C. (2011). An exploration of community reentry needs and services for prisoners: a focus on care to limit return to high-risk behavior. *AIDS patient care and STDs, 25*(8), 475-481. doi: 10.1089/apc.2010.0372

Nebraska Supreme Court Administrative Operations, Article 10, §1-1001

Olver, M.E., Stockdale, K.C., & Wormith, J. S. (2014). Thirty years of research on the Level of Service Scales: A meta-analytic examination of predictive accuracy and sources of variability. *Psychological Assessment, 26,* 156-176. doi: org/10.1037/a0035080

Parent, G., Guay, J. P., & Knight, R. A. (2012). Can we do better? The assessment of risk of recidivism by adult sex offenders. *Criminal Justice and Behavior, 39,* 1647-1667. DOI: 10.1177/0093854812451680

Schmidt, F., Campbell, M.A., & Houlding, C. (2011). Comparative analyses of the YLS/CMI, SAVRY, and PCL:YV in adolescent offenders: A 10-year follow-up into adulthood. *Youth Violence and Juvenile Justice, 9,* 23-42. doi.org/10.1177/1541204010371793

Schmidt, F., Sinclair, S. M., & Thomasdóttir, S. (2015). Predictive validity of the Youth Level of Sercie/Case Management Inventory with youth who have committed sexual and non-sexual offenses. *Criminal Justice and Behavior, 43,* 413-430. doi.org/10.1177/0093854815603389

Taylor-Gooby, P., & Zinn, J. O. (2006). Current directions in risk research: New developments in psychology and sociology. *Risk Analysis, 26,* 397 – 411. doi: 10.1111/j.1539-6924.2006.00746.x

Viljoen, J. L., Shaffer, C. S., Gray, A. L., & Douglas, K. S. (2017). Are adolescent risk assessment tools sensitive to change? A framework and examination of the SAVRY and the YLS/CMI. *Law and Human Behavior*, *41*, 244-257. doi.org/10.1037/lhb0000238

Ward, T., Melser, J., & Yates, P. M. (2007). Reconstructing the Risk-Need-Responsivity model: A theoretical elaboration and evaluation. *Aggression and Violent Behavior*, *12*, 208-228. doi.org/10.1016/j.avb.2006.07.001