

YOUTH LEVEL OF SERVICE/CASE MANAGEMENT INVENTORY:  
THE PREDICTIVE VALIDITY OF POST-COURT INVOLVEMENT ASSESSMENT

By

Ashlee R. Barnes

A THESIS

Submitted to  
Michigan State University  
in partial fulfillment of the requirements  
for the degree of

Psychology- Master of Arts

2013

## **ABSTRACT**

### **YOUTH LEVEL OF SERVICE/CASE MANAGEMENT INVENTORY: THE PREDICTIVE VALIDITY OF POST-COURT INVOLVEMENT ASSESSMENT**

By

Ashlee R. Barnes

Juvenile risk assessments are becoming increasingly popular in jurisdictions across North America. Court officials use risk assessment scales to predict future crime, identify youth needs, and inform case planning. If risk assessment tools are to be useful, they must demonstrate predictive validity overall as well as demonstrate predictive validity across gender and racial subgroups. Currently, the literature shows that juveniles are typically assessed when they enter court jurisdiction. This initial risk assessment score is the only one used to predict recidivism. This study sought to determine the predictive accuracy of the composite risk score youth received following dismissal from court jurisdiction. The entry/initial and exit/dismissal composite scores were compared to identify their relative validity. Differential predictive validity across race/ethnicity and gender was also explored. Theoretical and policy implications and the impact of court supervision were then discussed.

**Copyright by  
ASHLEE R. BARNES  
2013**

**DEDICATION**  
To Kamran for motivating me

## **ACKNOWLEDGMENTS**

Thanking God for keeping His promise

## TABLE OF CONTENTS

<b>LIST OF TABLES .....</b>	<b>vii</b>
<b>INTRODUCTION .....</b>	<b>1</b>
<i>History of Risk Assessment .....</i>	<i>4</i>
<i>Risk Assessment Utility .....</i>	<i>9</i>
Predictive Validity of Risk Assessments .....	12
<i>Adult Offenders .....</i>	<i>13</i>
<i>YLS/CMI.....</i>	<i>15</i>
<i>Variety of Juvenile Risk Assessments.....</i>	<i>19</i>
Disproportionate Minority Contact.....	22
<i>Differential Predictive Validity by Race/ethnicity .....</i>	<i>24</i>
Gender and Risk Assessment.....	28
<i>Validity of Reassessment Risk Scores .....</i>	<i>31</i>
Current Study .....	41
<i>Significance of Current Study .....</i>	<i>41</i>
<b>METHODS .....</b>	<b>44</b>
<i>Sample.....</i>	<i>44</i>
<i>Training and Procedures .....</i>	<i>45</i>
<i>Measures .....</i>	<i>45</i>
<b>RESULTS .....</b>	<b>47</b>
<i>Do initial and exit YLS/CMI risk scores differ in mean level and variability?.....</i>	<i>48</i>
<i>Are YLS/CMI risk scores assessed at exit from court supervision differentially valid predictors of recidivism? .....</i>	<i>49</i>
<i>Do race/ethnicity and gender moderate the relationship between risk and recidivism for initial and exit scores? .....</i>	<i>50</i>
<i>What is the relative predictive validity of change in risk scores and exit risk scores? .....</i>	<i>52</i>
<i>Does time under court supervision moderate the relationship between risk and recidivism? Does time under court supervision moderate the relationship between change in risk score and recidivism?.....</i>	<i>53</i>
<b>DISCUSSION .....</b>	<b>56</b>
<i>Limitations .....</i>	<i>60</i>
<i>Future Directions and Implications.....</i>	<i>61</i>
<b>APPENDIX.....</b>	<b>64</b>
<b>REFERENCES.....</b>	<b>66</b>

## LIST OF TABLES

Table 1. Risk Assessment Validation Studies.....	36
Table 2. Risk Assessment Validation Studies Summary .....	39
Table 3. Sample Characteristics.....	44
Table 4. Correlations for Independent and Dependent Variables.....	47
Table 5. YLS/CMI Initial and Exit Mean Risk Scores, Risk Levels, and Recidivism Rates	48
Table 6. Area Under the Curve for Initial and Exit Risk Scores .....	49
Table 7. Logistic Regression Predicting Recidivism by Initial and Exit Risk Scores.....	50
Table 8. Subgroup Area Under the Curves for Initial and Exit Risk Scores .....	51
Table 9. Logistic Regression Predicting Recidivism by Gender and Race/Ethnicity.....	52
Table 10. Logistic Regression Predicting Recidivism by Change Scores and Exit Scores...	53
Table 11. Risk Level Composition for Time Under Supervision Categories .....	54
Table 12. Logistic Regression Predicting Recidivism by Time Under Supervision .....	54
Table 13. Logistic Regression Predicting Recidivism by Change in Risk Scores.....	55

## INTRODUCTION

Assessing the risks and needs of juvenile offenders may play a key role in reducing the number of youth that become criminals as they enter adulthood. The lifetime costs of a career criminal in the United States totals 2.1 to 3.7 billion dollars (Cohen, Piquero, & Jennings, 2010). According to the most recent national juvenile court statistics, the number of delinquency cases processed by juvenile courts rose from 1, 200,000 in 1985 to 1, 500,000 in 2009; a 30% increase (Puzzanchera, Adams, & Hockenberry, 2012). Nationwide, juvenile offenders were arrested for 10% of murders and 25% of property crimes in 2009 (Puzzanchera & Adams, 2011). Although only a small percentage of juvenile offenders become repeat offenders (Cottle, Lee, & Heilbrun, 2001), delinquency continues to be a prevalent issue as much time and resources are allocated to processing young offenders.

In the last 20 years, risk assessment tools have been developed to distinguish between youth more and less likely to reoffend. Risk assessment measures are becoming increasingly important tools in juvenile jurisdictions across the United States (Onifade et al., 2008). “Risk assessment utilization has grown from 33% of state juvenile justice systems in 1990 to 86% by 2003” (Schwalbe, 2007, p.449).

Risk assessments are comprised of criminogenic risk factors that predict recidivism, inform decision-making, and assess needs (Onifade et al., 2008; Tyda, 2011). The literature contains extensive information on the validity and utility of risk assessment measures, the likelihood of specific types of offenses occurring (e.g., violent offense, sex offense), and differential predictive validity based on race/gender (Olver, Stockdale, & Wormith, 2009; Onifade, Davidson, & Campbell, 2009). To date, research has focused on assessing risk of recidivism based on the risk score a juvenile receives upon entering the juvenile justice system.



Specifically, of the 34 validation studies of the most investigated risk assessment measures to date, only three studies have examined reassessments of risk during court supervision and/or assessment of risk upon dismissal from court jurisdiction (Flores, Travis, & Latessa, 2003; Schlager & Pacheco, 2011; Vose, Lowenkamp, Smith, & Cullen, 2009;). Hence, the predictive validity of risk assessments post court intervention is not well documented. Recent studies have indicated that risk scores from initial assessments can accurately predict recidivism (Catchpole & Gretton, 2003; Olver et al., 2009;); they have also shown risk scores from assessments following court supervision can be valid predictors of recidivism (Baglivio & Jackowski, 2012). However, there is not a study that has directly assessed the relative validity of the two. Currently, initial risk scores appear to be the most studied assessment of recidivism (Betchel, Lowenkamp, & Latessa, 2007; Olver, Stockdale, & Wong, 2011; Onifade et al., 2009;). On the other hand, assessment scores following court intervention may offer additional information about recidivism risk. This type of research may also help disentangle the relative effects of initial risk level and subsequent intervention.

Embedded in the issue of crime and delinquency is the overrepresentation of minority youth in the juvenile justice system. A recent report stated that minority youth 10-17 years of age comprised 23% of the total United States population, yet they constituted 52% of incarcerated youth (McGhee & White, 2010). Furthermore, Black juvenile offenders are confined on average for 61 days longer than White youth, and Latino youth are confined 112 days longer than White youth, even when accounting for offense seriousness (Piquero, 2008). Disproportionate minority contact in the justice system is a prominent social issue in the United States. Many view risk assessment scales as a strategy to guarantee equal treatment for all offenders. It has been argued that risk assessment measures have the means to reduce discretionary biases by increasing the

consistency of assessment through a structured process (Schwalbe, Fraser, Day, & Cooley, 2006). Ideally, if court personnel are employing level of risk to make decisions on how to sanction (community-based programming, residential treatment facility, detention center) a juvenile offender, then all offenders should be managed equally regardless of race or ethnic background. However, this ideal is premised on the assumption that risk assessment scales are equal predictors of juvenile recidivism across race/ethnicity. It is also important to note that risk assessment measures were validated against many forms of recidivism (i.e. subsequent arrest, incarceration, or violation). Since any form of recidivism may be based on a biased system response (e.g. differential surveillance in neighborhoods of color, racial profiling, differential processing rates, harsh and unequal sentences for similar offenses across race/ethnicity), it may be that risk assessment itself is systematically biased. In other words, it would be problematic to assume that risk assessment measures could predict an unbiased outcome variable (recidivism), when the outcome variable itself, is systematically inequitable.

In addition to the issue of race/ethnicity in the juvenile justice system, the role that gender plays in processing youthful offenders also merits discussion. Although males make up the large majority of the juvenile justice system caseloads (about 75%), females are among the fastest growing subpopulation of juvenile offenders (Puzzanchera et al., 2012). Female offenders are becoming more visible on juvenile court caseloads, however their risk and needs are being assessed by risk assessment tools that have been developed on all male, or nearly all male samples. While some scholars do not believe that there are gendered pathways to crime, others suspect that female offenders may have different patterns of risk than their male counterparts (Chesney-Lind, Morash, & Stevens, 2008). Specifically, Hoge and Andrews recently developed an updated version of the Youth Level of Service/Case Management Inventory, designed to be

“gender-informed.” The role of gender in risk assessment is both scarce and controversial, therefore it is important to examine how gender impacts predictive validity.

The current study helped quantify the importance of assessing risk over time utilizing a widely validated risk assessment scale, Youth Level of Service/Case Management Inventory (YLS/CMI). Risk scores assessed at entry to the court and upon exiting the court were compared to identify their relative validity in predicting one-year recidivism for a sample of juvenile probationers. In order to detect the impact of court supervision on recidivism, the current study compared the entry and exit risk scores to identify significant differences. This study also examined the differential predictive validity of the post-intervention risk scores based on race/ethnicity and gender.

For purposes of organization, the following literature review summarizes the history of risk assessment as it relates to the contributions of clinical judgment and criminogenic risk factors; describes risk assessment utility for offender populations in the criminal and juvenile justice systems; presents a review of validation studies that investigated popular juvenile risk assessment measures; describes disproportionate minority contact and its connection to differential predictive validity of risk assessment for non-White offenders; briefly addresses gender and risk assessment; and describes the importance of reassessment. Following the literature review, research questions and pertinent methods are presented. This will be followed by the results and discussion section.

### *History of Risk Assessment*

Dating back to the first risk assessment methods, crime and delinquency has remained a prevalent social issue. Juvenile justice systems are continuously developing ways to manage growing numbers of juvenile offenders. Risk assessment scales are widely used in juvenile

justice systems to predict recidivism risk, to assess an offender's rehabilitative needs, and to inform decisions on how to process juvenile delinquents. Risk assessment as it is commonly used today underwent an evolution from subjective opinions to empirically validated techniques. Before actuarial risk assessment scales were developed, courts solely relied on the professional judgment of clinicians and court officials to make decisions about how to process offenders (Andrews, Bonta, & Wormith, 2006). At the time, this method appeared to be reliable because officials used their professional experiences as a basis to decide which offenders were most likely to make future contact with the justice system. Yet, unaided clinical judgment proved to be detrimental as court officials had varying levels of experience, personal biases, and different opinions as to the most effective way to predict recidivism and to inform the decision-making process (Schwalbe, Fraser, Day, & Arnold, 2004).

More generally, the specific issues related to the use of risk assessment tools in the juvenile justice system lies within a larger tradition of clinical versus actuarial prediction in the field of psychology and across several disciplines. For example, researchers have examined the accuracy of informal versus formal risk assessments in education, employment selection, college admissions, medical diagnoses, and psychiatry. Several studies found that compared to systematic assessment, which is based on statistical calculation, unstructured clinical/professional judgment was less than satisfactory (Grove & Meehl, 1996; Krysick & LeCroy, 2002; Schwalbe, 2008; Schwalbe et al., 2004; Shaffer, Kelly, & Lieberman, 2011; Tyda, 2011). A meta-analysis examining the prediction of human behavior and health diagnoses found that only in a small number of cases (8 out of 136) were clinical judgments able to outperform risk assessment tools (Grove & Meehl, 1996; Shaffer et al., 2011). While the clinical judgment method is still used today, it has been suggested that actuarial measures are better than using

clinical judgment (Grove & Meehl, 1996), as it is more accurate in predicting risk and reduces bias by building structure and uniformity into the decision-making process.

The history of risk assessment appears to be largely atheoretical. However, by inference, risk assessment lies within a framework that supposes that chronic offenders have unique characteristics that are not shared by one-time offenders. Ernest Burgess, who examined case characteristics related to parole violations in a sample of adult offenders, developed the first actuarial scale in the late 1920s (Onifade et al., 2008; Schwalbe, Fraser, & Day, 2007). Actuarial risk assessment scales are comprised of empirically derived criminogenic risk factors. In order to develop these scales Burgess used risk factors to determine who was most likely to violate parole. This process involved creating levels of risk that helped parole boards decide who should be released and who should remain incarcerated (Burgess, 1928).

While determining level of risk for adult offenders on parole was essential, it quickly became evident that it was also important to design risk assessment scales for young offenders. An exemplar meta-analysis, conducted by Cottle and colleagues (2001) compared effect sizes of 22 independent samples and identified 30 risk factors most strongly predictive of juvenile recidivism. The strongest predictors included “age at first commitment, age at first contact with the law, non-severe pathology, family and conduct problems, effective use of leisure time, and delinquent peers” (Cottle et al., 2001, p. 385). Psychometricians drew on these factors when designing actuarial risk assessment scales, increasing their predictive validity.

Criminogenic risk factors range from individual to macro-level characteristics that increase the likelihood that an offender will make contact with the justice system. These risk factors have been categorized as either static or dynamic. Static risk factors are predictors of recidivism that are related to an offender’s past including gender, age at first arrest, or number of

prior arrests (Tyda, 2011). These risk factors are characterized by their “fixed” nature in that there is nothing that treatment can do to alter the offender’s history or demographic characteristics (McGrath & Thompson, 2012). Static factors have been documented as some of the strongest predictors of reoffending. For instance, out of the 30 risk factors Cottle et al. (2001) identified, the two most predictive variables were static factors (age at first commitment and age at first contact with the law).

On the other hand, dynamic risk factors, also referred to as criminogenic needs, are predictors of recidivism that are characterized by their potential to change (Fass, Heilbrun, Dematteo, & Fretz, 2008). Intervention efforts usually target dynamic risk factors such as family instability, association with delinquent peers, and poor use of leisure time (Vincent, Paiva-Salisbury, Guy, & Perrault, 2012). Less research has been conducted on dynamic factors. However, currently, dynamic risk factors are becoming increasingly popular as emphasis is shifting from risks to needs (Fass et al., 2008).

In a recent study, authors used the Australian Adaptation of the Youth Level of Service/Case Management Inventory to identify which category of criminogenic risk factors was most predictive of juvenile recidivism (McGrath & Thompson, 2012). Using logistic regression, the first Model contained the YLS/CMI-AA’s static factors (Prior Offenses domain), the second Model contained dynamic factors (the remaining 7 domains) and the third Model contained both static and dynamic factors. The results indicated that while the first Model (Prior Offenses domain) was a significant predictor of recidivism, only four (Education/Employment, Peer Relations, Substance Abuse, Attitudes/Beliefs) of the seven domains were significant predictors of recidivism in the second Model (McGrath & Thompson, 2012). Overall, the third Model containing a combination of static and dynamic domains explained the most variance in the

outcome variable. These findings illustrated the importance of both static and dynamic criminogenic factors in accurately predicting future reoffending (McGrath & Thompson, 2012). Identifying risk factors that are most predictive of recidivism is essential to the development of actuarial risk assessment scales.

To stay current, it was important for risk assessment tools to advance by including variables and techniques that best predict recidivism. Several iterations of risk assessment scales have been designed and can be classified into different generations. First generation risk assessment solely relied on the professional judgment and intuition of clinicians and court officials to make decisions about how to handle offenders (Brennan, Dieterich, & Ehret, 2009). Also known as unaided clinical judgment, this method of risk prediction has proven inadequate for reasons such as “users being unaware of (or ignoring) recidivism base rates, weighting factors in a manner inconsistent with research, and classification based on erroneous mental heuristics...” (Shaffer et al., 2011, p. 168). In other words, clinical judgment was equivalent to one’s best guess.

Second generation scales employed static criminogenic risk factors (Andrews et al., 2006) to predict future crime. Second generation scales such as the Psychopathy Checklist-Revised (PCL-R) have demonstrated predictive ability superior to unaided clinical judgment (Brennan et al., 2009). However, use of second generation tools have been criticized for its limited coverage of relevant risk/need factors, its use of atheoretically derived risk factors, its failure to include dynamic risk factors, and oversight of treatment implications (Brennan et al., 2009).

The development of third generation scales such as the Level of Service Inventory-Revised (LSI-R), addressed the limitations of second-generation measures. Third generation risk

assessment tools include empirically derived static and dynamic risk factors that allow court officials to address offenders' needs. Although these measures have the ability to predict recidivism, they are criticized for their narrow theoretical focus, failure to highlight offender's strengths, and absence of guidance in regards to case management (Andrews et al., 2006; Brennan et al., 2009; Shaffer et al., 2011).

Fourth generation risk assessment measures built upon third generation scales by including a responsivity component. The responsivity principle involves recognizing that individual-level characteristics such as IQ, mental health, personality, and gender should be considered when delivering treatment (Vitapolous, Peterson-Badali, & Skilling, 2012). Fourth generation risk assessments also include protective factors, inclusion of more theoretically derived risk factors, improved case management guidelines, and ease of integration with court computer information systems (Brennan et al., 2009). The Level of Service/Case Management Inventory (LS/CMI) and Correctional Offender Management Profiling for Alternative Sanctions (COMPAS) are the most well known fourth generation risk assessment measures (Andrews et al., 2006; Brennan et al., 2009).

#### *Risk Assessment Utility*

For various reasons, risk assessment is an integral component in the juvenile justice system decision-making process. Risk assessment instruments can measure the likelihood that an offender will "reoffend, violate probation, or fail to appear in court" (Tyda, 2011, p. 3). Risk assessments can also predict rearrest, reincarceration, technical violations, program completion, and reconvictions (Baglivio & Jackowski, 2012; Flores et al., 2003). Drawing on validated risk factors to identify level of risk in juvenile offenders is essential both theoretically and in practice. The average young offender may commit one or two new offenses, yet approximately 6-8% of



young offenders are responsible for more than half of all juvenile crimes committed (Cottle et al., 2001; Onifade et al., 2008; Schmidt, McKinnon, Chattha, & Brownlee, 2006).

Utilizing risk assessment scales to classify recidivists and non-recidivists is advantageous for both court officials and juvenile delinquents. Accurate classification affords court officials the opportunity to allocate intensive services for youth that are most likely to make future contact with the court (Schwalbe et al., 2004). Reserving costly services such as intensive probation and residential placement will protect the court's resources and enable them to focus on offenders that require the most attention (Meyers & Schmidt, 2008). Similarly, young offenders assessed as low risk to reoffend can be appropriately sanctioned with community-based programs, or diverted out of the justice system entirely (Krysick & LeCroy, 2002). More crucially, studies have confirmed that by sentencing low-risk youth to detention facilities populated with high-risk offenders can actually increase delinquency among low-risk youth (Gatti, Tremblay, & Vitaro, 2009). For example, low-risk offenders can be exposed to riskier delinquent behaviors displayed by high-risk offenders such as violence, glamorizing illegal activity, or sharing negative attitudes towards authority; which could potentially interfere with the characteristics of being a low-risk offender (e.g. lack of delinquent peers/associates). Utilizing risk assessment measures may help prevent iatrogenic effects because court officials would rely on level of risk to decide how to process juvenile offenders; decreasing the likelihood that low-risk youth are exposed to peers displaying higher risk for deviance (Gatti et al., 2009).

Another useful feature of risk assessment scales is the ability to predict time to new offense. The literature supports that risk scores are associated with the number of days it takes to commit a new offense (Catchpole & Gretton, 2003). Meaning, as the risk score increases, the time to new offense is shorter. The association between composite risk scores and time to re-

offense can help court personnel predict how quickly an offender is most likely to reoffend. Recognizing that higher risk youth are likely to reoffend sooner will prompt court officials to prioritize securing treatment services to the highest risk youth in a timely manner. However, it is important to note that finding significant differences between each risk level for time to re-offense has often varied by research study. For example, Catchpole & Gretton, (2003) examined the ability of three risk assessment tools (YLS/CMI, PCL:YV, SAVRY) to predict time to new offense for young violent offenders. While the authors found significant differences in time to new offense for all three risk levels (low, moderate, high risk) using the YLS/CMI, the remaining risk assessment tools only showed significant differences between high risk and low risk offenders. Similar results have been demonstrated elsewhere (Onifade et al., 2008).

Risk assessment inclusion of dynamic risk factors is an additional benefit for the juvenile justice system. Utilizing dynamic risk factors to assess offenders can serve a dual (Vincent, Chapman, & Cook, 2011) role in case planning as these factors provide information about both level of risk and criminogenic needs. For instance, when court personnel identify that offenders are scoring high in Education risk/need, the court can provide academic services that address that domain. Furthermore, there is evidence to support that courts failing to target offenders' specific needs results in higher recidivism rates. (Bonta, Rugge, Scott, Bourgon, & Yessine, 2008; Luong & Wormith, 2011; Vieira, Skilling, & Peterson-Badali, 2009; Vincent et al., 2012).

In addition to serving as a needs assessment, predicting risk to recidivate, and time to new offense, these tools can predict specific recidivism outcomes. For instance, the Juvenile Sex Offender Protocol—II (J-SOAP-II) can predict the likelihood that a youth will sexually re-offend (Worling & Langstrom, 2003). The Youth Level of Service/Case Management Inventory (YLS/CMI) was designed to predict general recidivism (Bonta et al., 2008), while the Structured

Assessment of Violent Risk in Youth (SAVRY) was formed to predict violent offenses (Martinez, Flores, & Rosenfeld, 2007). Furthermore, many studies have been able to use risk assessment tools to predict nonviolent recidivism (Welsh et al., 2008), institutional violations (Flores et al., 2003), technical violations, and successful program completion (Shaffer et al., 2011). Identifying a youth as high risk to commit a future violent or sexual offense is critical to providing relevant treatment services that will reduce risk and target specific needs.

Risk assessment instruments are central to the decision-making process in many jurisdictions across North America. As previously stated, these measures are used to guide case management and to predict various outcomes such as sexual re-offense and successful program completion. It is important to note that the utility of risk assessments would be unknown without validation studies. Research studies that seek to investigate the predictive validity of risk assessment instruments provide evidence as to whether these measures can accurately predict intended outcomes. The next section will present a review of adult and juvenile risk assessment validation studies.

### Predictive Validity of Risk Assessments

Relevant studies were identified through several avenues. Electronic databases (PsychInfo, ProQuest, Web of Science, Google Scholar) were searched for published studies using the following search terms: juvenile recidivism, risk assessment, predictive validity, YLS/CMI, LSI-R. Of those articles selected, the references sections were examined to identify additional risk assessment validation studies. While a few published studies from 2012 were identified, the majority of the research studies included were published in 2011 and earlier. The literature review based on adult offenders is non-comprehensive as it only includes nine validation studies based on various versions of the LSI. It was important to include a brief review

of adult studies because the current study's measure (YLS/CMI) was adapted from the LSI. Adult studies were identified while searching the electronic databases listed above for any research conducted on reassessment scores (Keywords: reassessment, risk assessment change, risk scores over time). The references section of the first study identified on reassessment scores that used an adult sample was perused to identify additional studies with the same measure (LSI). The review of the juvenile risk assessments includes a comprehensive number of studies investigating the validity of the YLS/CMI. It also includes a small number of studies examining other popular juvenile risk assessments (i.e. SAVRY).

In order to efficiently present the methodological details of this literature review, Table 1 was constructed. For each article reviewed for this study, Table 1 presents the sample size; percentage of the sample that was non-White; risk measure employed; the Area Under the Curve (AUC) statistic; percentage of studies that examined the validity of the risk score assessed at entry to court (intake validity); percentage of studies that examined the validity of the risk scores assessed upon exiting the court (exit validity); percentage of studies that investigated the differential predictive validity of risk scores by race/ethnicity (race validity); percentage of studies that investigated the differential predictive validity of risk scores by gender (gender validity) and the follow-up length for each study. Table 2 illustrates a summary of the information found in Table 1. At the end of this review, both Table 1 and Table 2 will be used in the methodological critique.

### *Adult Offenders*

Several risk assessment scales have been developed over the past few decades. The most widely-investigated and implemented risk assessment tool for adult offenders is the Level of Service Inventory-Revised (LSI-R) (Holsinger, Lowenkamp, & Latessa, 2006). The LSI-R is a

third-generation risk assessment designed to predict general recidivism. The LSI-R has demonstrated accurate predictive validity across several studies (Fass et al., 2008; Holsinger et al., 2006; Luong & Wormith, 2011; Schlager & Simourd, 2007). In a recent study, authors conducted a meta analysis that sought to identify the sources of variability in the strength of the LSI-R predictive validity estimates. Although the bivariate correlations ranged between .16 and .46 (mean  $r = .36$ ), each correlation reached statistical significance demonstrating the predictive utility of the LSI-R's across 42 studies (Andrews et al., 2011).

The LSI-R has been found to better predict recidivism than more recently developed risk assessment scales. Fass et al. (2008) were the first to publish a validation study on the Correctional Offender Management Profiling for Alternative Sanctions (COMPAS), a fourth-generation risk assessment. In order to see how the COMPAS measured up to more established risk assessment tools, the authors compared the predictive validity of the COMPAS to the LSI-R in an all male sample ( $N = 975$ ) of recently released prisoners. Recidivism was coded as any new arrest during a 12 month follow-up period. For the total sample, the Area Under the Curves (AUCs) for the LSI-R and COMPAS was .60 and .53, respectively. While the AUC for the LSI-R composite score was substantial, the AUC for the COMPAS was only slightly above chance (Fass et al., 2008).

Although the LSI-R was initially developed on Canadian offenders, the scale has since been shown to obtain robust results across diverse samples including Australian offenders (Hsu et al., 2009), Native American offenders (Holsinger et al., 2006), African American offenders, and Hispanic offenders (Schlager & Simourd, 2007). In a large sample ( $N = 78,052$ ) of Australian offenders, Hsu and colleagues (2009) sought to identify the predictive validity of the LSI-R across gender and sentence order. Sentence order was classified as either community or

custodial, with offenders sentenced to community orders being recommended to complete community service, while offenders sentenced to custodial orders were typically incarcerated. The LSI-R total score was found to be moderately predictive of recidivism with the highest bivariate correlations for male and female offenders sentenced to custodial orders ( $r = .20$  and  $.23$ , respectively) (Hsu et al., 2009). In a study that explored the predictive validity of the LSI-R on a sample of male and female Native American offenders ( $N = 405$ ), the bivariate correlation for the overall sample was  $r = .18$ ,  $p < .05$ . (Holsinger et al., 2006). Similarly, moderate predictive validity was demonstrated for a predominantly non-White sample of male adult offenders with an AUC of  $.60$  (Fass et al., 2008).

Unfortunately, the LSI-R is not always robust when predicting across diverse samples. When investigating the predictive validity of the LSI-R with an African American and Hispanic sample, Schlager & Simourd (2007) concluded that while the statistical relationship between the composite scores and reconviction rates at a two-year follow up for African American offenders was significant, the predictive validity estimates for the overall sample was not. Furthermore, the authors concluded that overall, the LSI-R was not an effective risk assessment tool for their sample and that their results were weak in comparison to previous findings on diverse samples (Schlager & Simourd, 2007).

#### *YLS/CMI*

The LSI-R and many other risk assessment measures were developed for adult offenders, however the prevalence of juvenile delinquency created an impetus to design tools to predict recidivism risk for young offenders. The most widely-used and validated risk assessment tool for juvenile offenders is the Youth Level of Service/Case Management Inventory (YLS/CMI) (Schmidt, Campbell, & Houlding, 2011; Schwalbe, 2007). The YLS/CMI is a third generation

risk assessment that was adapted from the structure and components of the LSI-R. The YLS/CMI was created by Hoge & Andrews, (2002) in order to predict general recidivism for young offenders aged 12-18; as well as assist juvenile court officers in needs assessment/case management with the inclusion of dynamic risk factors. The YLS/CMI was originally developed on a Canadian sample of young probationers and is currently utilized in juvenile jurisdictions across North America (Betchel et al., 2007; Jung & Rawana, 1999). The YLS/CMI has 42 items that are divided into eight subscales. The subscales are Prior/Current Offenses, Education, Leisure & Recreation, Peer Relations, Substance Abuse, Family & Parenting, Attitudes & Orientation, and Personality (Schmidt et al., 2005). Each item is scored dichotomously (yes or no) indicating whether or not risk is present. The items are totaled and the composite score is translated into a level of risk; low, moderate, high, and very high (Flores et al., 2003).

The YLS/CMI has been examined with a variety of samples that provide evidence of its predictive validity. For example, United States and Canadian studies on the YLS/CMI demonstrated that it accurately predicted recidivism 7-18% better than chance (Onifade et al., 2011; Onifade et al., 2009; Onifade et al., 2008; Schmidt et al., 2005). Several authors have sought to validate the YLS/CMI's ability to predict recidivism and time to new offense by risk level. In one sample of juvenile offenders, the YLS/CMI correctly identified 59% of the offenders as either recidivists or non-recidivists, with an AUC of .62 (Onifade et al., 2008). With a 12-month follow-up, they also found a significant relationship between time to re-offense and YLS/CMI composite score for high and low risk offenders (Onifade et al., 2008). Another research study yielded significant correlations between YLS/CMI total score and "any re-offense" and "serious re-offense" recidivism outcomes (Schmidt et al., 2005). Although the AUC statistics varied in magnitude across the recidivism outcomes, the YLS/CMI maintained

predictive accuracy for both any re-offense and serious re-offense with .67 and .61, respectively (Schmidt et al., 2005). Furthermore, evidence of strong predictive validity of the YLS/CMI has been illustrated elsewhere (Onifade et al., 2011; Schwalbe, 2007).

The YLS/CMI has been shown to accurately predict various recidivism outcomes such as rearrest/reincarceration (Flores et al., 2003), reconviction (Olver et al., 2011), and treatment success (Vieira et al., 2009). While the YLS/CMI has made valuable contributions to the literature and juvenile justice practices, research validating the YLS/CMI are not without limitations. Studies investigating the YLS/CMI have often employed small sample sizes (Catchpole & Gretton, 2003; Vitopolous et al., 2012), retrospective coding, lack of diversity within samples (Flores et al., 2003; Schmidt et al., 2011), neglected to address potential gender and race/ethnicity differences (Schmidt et al., 2005; Welsh et al., 2008), and utilized inadequate follow up periods (Jung & Rawana, 1999).

It has been suggested that researchers should employ a follow-up period of at least 12 months or more, as this guarantees an adequate amount of time for youth to reoffend (Onifade et al., 2008). The predictive validity of the YLS/CMI has been confirmed with longer follow up times in several exemplar studies (Catchpole & Gretton, 2003; Flores et al., 2003; McGrath & Thompson, 2012; Onifade et al., 2008). Specifically, McGrath & Thompson (2012) examined reconviction rates among a large sample ( $N = 3,568$ ) of Australian young offenders and identified an AUC of .65 after a 12 month follow-up. Likewise, two studies investigated the predictive validity of three major risk assessments tools with follow up times up to 10 years (Schmidt et al., 2011; Welsh et al., 2008). The YLS/CMI achieved AUCs of .64 and .66, further illustrating sound predictive validity. On the other hand, studies that tracked youth for less than 12 months also obtained comparable results. Jung & Rawana (1999) examined how well the



YLS/CMI could predict reoffense rates at a 6-month follow up period for a sample of Canadian youth. ANOVA and MANOVA analyses revealed accurate predictive ability with a main effect of each subscale and YLS/CMI total score.

Predictive validity of the YLS/CMI was also demonstrated for different categories of juvenile offenders including offenders at intake (Onifade et al., 2011; Onifade et al., 2009), offenders referred to mental health assessments (Schmidt et al., 2011; Schmidt et al., 2005; Vieira et al., 2009), violent offenders (Catchpole & Gretton, 2003), youth on probation (Jung & Rawana, 1999; Onifade et al., 2008), and juveniles with custodial dispositions (Flores et al., 2003; Vitopolous et al., 2012).

In one study, Onifade et al. (2008) used the YLS/CMI subscales and performed a cluster analysis in order to identify differences within categories of offenders traditionally classified into low, moderate, and high-risk groups (Onifade et al., 2008). The authors cross-validated the risk patterns across an intake and probation sample; results revealed five unique clusters. These cluster types improved prediction rates above and beyond that of the traditional risk groups. This finding highlights the ability of the YLS/CMI to accurately predict recidivism by level of risk for probationers and intake youth (Onifade et al., 2008).

Another study compared the predictive utility of the YLS/CMI across groups of offenders sentenced to a detention facility, rehabilitation center, and probation (Flores et al., 2003). Analyses revealed that the YLS/CMI total score was significantly related to recidivism across the three agencies. For the juvenile offenders sentenced to rehabilitation, the YLS/CMI significantly predicted re-arrest, re-arrest seriousness, and re-incarceration. The composite scores significantly predicted program completion, technical violations, and re-incarceration for probationers.

Finally, the YLS/CMI accurately predicted re-arrest and re-incarceration for youth placed in detention (Flores et al., 2003).

#### *Variety of Juvenile Risk Assessments*

Although the YLS/CMI is the most investigated and commonly used risk assessment for juvenile offenders, there are several other tools that have the ability to predict juvenile recidivism. Among the most popular are the Structured Assessment of Violence Risk in Youth (SAVRY), Psychopathy Checklist: Youth Version (PCL: YV), North Carolina Assessment of Risk (NCAR) and the Juvenile Sex Offender Protocol-II (J-SOAP-II). These scales have similar characteristics. These assessments are composed of 9-42 items that are divided into domains or subscales that typically measure: prior offense history, family circumstances, education, peer group, drug involvement, leisure activity, personality, impulsivity and attitudes (Cottle et al., 2001). Risk assessments are “additive models of risk, where probability of re-offense conceivably increases as a function of risk score” (Onifade et al., 2011, p. 841). Each tool has specified cut-off scores to indicate an offender’s level of risk (e.g. low, moderate, high, very high).

One of the key differences among risk assessment measures is the type of recidivism each were designed to predict. The J-SOAP-II was created to predict the probability juvenile offenders would repeat a sexual offense. The NCAR and YLS/CMI were created to predict general recidivism. The SAVRY was designed to predict the likelihood that an offender would commit a violent offense. Comparative analyses have identified that some scales have the ability to predict offenses outside of what they were originally intended to. For example, the PCL: YV was designed to predict general recidivism and to detect psychopathic tendencies, yet it has demonstrated the ability to predict nonviolent and technical offenses (Schmidt et al., 2011).

Researchers have sought to identify which risk assessment scales have the ability to predict juvenile recidivism the most accurately. These comparative studies have resulted in mixed findings. For instance, one study compared the SAVRY, PCL: YV and youth adaptations of the LSI (e.g. YLS/CMI) and found that the risk assessment tools did not outperform each other in predicting any recidivism outcome (Olver et al., 2009). Catchpole & Gretton (2003) compared the predictive utility of the same risk assessment tools (SAVRY, PCL: YV, YLS/CMI) with a small sample of violent offenders ( $N = 74$ ), with general and violent recidivism as outcome measures. For all three assessments, the AUCs for general recidivism ranged from .74 to .78, with the PCL: YV demonstrating the strongest predictive validity (Catchpole & Gretton, 2003). Nonetheless, similar to the results shown in Olver et al. (2009), the scales did not outperform each other in predicting violent recidivism with AUCs of .73 for all three scales (Catchpole & Gretton, 2003). Although these results appeared promising, the authors were criticized for their limitations (e.g. small sample size, short follow up).

In an effort to address the limitations in Catchpole & Gretton's (2003) study, Welsh et al. (2008) conducted a comparative study with the same tools (SAVRY, PCL: YV, YLS/CMI), a larger sample ( $N = 133$ ), longer follow up (3 year avg.), and a broader range of statistical analyses (i.e. logistic regression, bivariate correlations). The YLS/CMI demonstrated the weakest predictive validity for general recidivism, violent recidivism, and nonviolent recidivism. Although the SAVRY demonstrated the strongest validity in predicting violent recidivism with an AUC of .81, the PCL:YV was dominant in predicting nonviolent recidivism (Welsh et al., 2008). The logistic regression analysis confirmed these results. The SAVRY offered the most incremental validity by accounting for the most variance in general and violent recidivism while holding the YLS/CMI and PCL:YV constant. Similarly, the PCL:YV offered more incremental

validity than the YLS/CMI and accounted for more variance in explaining general and violent recidivism than the YLS/CMI (Welsh et al., 2008).

Other juvenile measures include variations of the LSI-R. With the permission of Multi-Health Systems, the Saskatchewan Department of Corrections (SDOC) developed the LSI-SK, a revised version of the LSI-R that was tailored to the SDOC jurisdictional needs (Luong & Wormith, 2011). The sample contained a large percentage of Aboriginal youth whose new conviction rates were tracked for an average of almost two years. Predictive validity estimates for the LSI-SK total score demonstrated a significant point biserial correlation of .39 and an AUC of .73, indicating predictive utility (Luong & Wormith, 2011).

Up to this point, the development of risk assessment tools for young offenders has been acclaimed by juvenile jurisdictions across North America. A variety of measures have been developed to measure specific recidivism outcomes including general, sexual, and violent reoffending. While research has indicated that some instruments are more valid and reliable than others (Welsh et al., 2008), the most widely investigated assessments (e.g. LSI-R, YLS/CMI) are also among the most commonly used (Andrews et al., 2011). Recently, risk assessment utility has extended beyond merely predicting future crime and has been suggested as a practice to standardize decision-making processes (Cabaniss, Frabutt, Kendrick, & Arbuckle, 2007). Specifically, of risk assessment has been recommended as a strategy to reduce the overrepresentation of minority offenders in the justice system. Supporters posit that risk assessment scales equally assign level of risk based on criminogenic factors, limiting the opportunity for treatment recommendations being influenced by race/ethnicity. Validation studies provide evidence that scales have the ability to predict recidivism, however, whether these tools predictions vary by race/ethnicity warrant further exploration. The next section will

present theoretical explanations and empirical support for disproportionate minority contact, its relationship to risk assessment, and highlight studies that have explored the differential predictive validity of risk assessment measures.

### Disproportionate Minority Contact

Racial disparities in the juvenile justice system are a prevalent social issue in the United States. Disproportionate minority contact (DMC) occurs at every level of the justice system: arrest, referral, petition to court, adjudication, detention, out-of-home placement following adjudication, and transfer to adult court (Bridges & Steen, 1998; Kakar, 2006; Werling, Cardner, & University-Austin, 2011; Wordes, Bynum, & Corley, 1994). The literature indicates that the highest instances of disparity occur in arrest rates and subsequent stages within the criminal justice system are affected as a result (Fitzgerald & Carrington, 2011; Freiburger & Jordan, 2011; Werling et al., 2011). For instance, “Black youth are more likely than Whites to be formally charged in juvenile court and to be sentenced to out-of-home placement, even when referred for the same offense” (Piquero, 2008). In addition, Piquero (2008) reported that national rates of incarceration for Hispanics was double that for Whites in 2005.

Two prominent theoretical explanations for disproportionate minority contact are differential involvement and differential processing (Bridges & Steen, 1998; Fitzgerald & Carrington, 2011; Hindelang, 1973; Werling et al., 2011). The theory of differential involvement posits that non-Whites simply commit crimes in higher frequency, severity, and more variety than their White counterparts (Piquero, 2008). Differential processing claims that racial disparities exist because of higher levels of profiling and police surveillance in communities of color, and discrimination in the handling of non-White offenders once they enter the court system (Bridges & Steen, 1998; Freiburger & Jordan, 2011; Kakar, 2006). Studies have found

mixed results on the underlying causes of DMC. While some studies support differential involvement, others provide evidence of differential processing at various stages in the criminal justice system.

Hindelang (1973) was one of the first to explore the relationship between race/ethnicity and crime. He compared official arrest records to victimization survey data and found evidence that supports differential involvement of Black offenders for rape and robbery. Similar to Hindelang's study, Elliot & Ageton (1980) found that Black offenders were disproportionately represented among high frequency offenders. The authors concluded, "while we do not deny the existence of official processing biases, it does appear that official correlates of delinquency also reflect real differences in frequency and seriousness of delinquent acts" (Elliot & Ageton, 1980, p. 107). On the other hand, Piquero & Brame (2008) investigated this issue using both official police records and self-reported delinquency for a sample of serious adolescent offenders. Overall, they found no significant racial bias in the frequency or variety of official or self-reported prior offenses.

Other authors found evidence that supported the theory of differential processing. Wordes et al. (1994) sought to identify whether race had an impact on police, court intake, and court preliminary hearings decisions to detain youth. Their results indicated that race had an independent effect on detention decisions, even after controlling for agency, felony offense seriousness (e.g. use of weapon), and social factors (e.g. socioeconomic status) (Wordes et al., 1994). Other studies have found evidence of DMC when making decisions to petition a case to court (Freiburger & Jordan, 2011), racially discriminatory policing (Fitzgerald & Carrington, 2011), and sentencing recommendations (Bridges & Steen, 1998).

It is apparent that previous research has not conclusively isolated the root causes of DMC. Frieburger & Jordan (2011) asserted that mixed findings might be due to particular contexts being overlooked in the analysis. The authors looked at overrepresentation in regards to which youth were more likely to be petitioned to court while controlling for county-level (e.g. female-headed households) and individual-level factors (e.g. offense type). Results demonstrated that race did not significantly impact the odds of being petitioned to court. However, structural and individual interactions indicated that Black youth in high poverty areas were more likely to be petitioned (Freiburger & Jordan, 2011). While it is not precisely clear why minority offenders are overrepresented, juvenile justice agencies have begun to develop strategies to prevent it from occurring (Cabaniss et al., 2007).

In an effort to address the racial disparities in the juvenile justice system, the Office of Juvenile Justice and Delinquency Prevention recommended standardized risk assessments as a best practice tool (Onifade et al., 2009). Utilizing risk assessment instruments to reduce disproportionate minority contact supports the theory of differential processing. Therefore, if non-White offenders are disparately charged and receiving harsher sentences than White offenders for similar crimes, employing a tool that standardizes court decision-making may be one way to address this issue. However, and potentially more important, is whether these tools accurately and reliably predict recidivism equally for White and non-White offenders. With this notion, examining the differential predictive validity of risk assessment is essential. In the next section, literature will be summarized that examines the ability of risk assessment scales to predict across race/ethnicity.

#### *Differential Predictive Validity by Race/ethnicity*

As stated earlier, the overrepresentation of minority offenders in the juvenile justice system is a longstanding social issue; risk assessment researchers are increasingly beginning to study this phenomenon. The differential predictive validity of risk assessment tools has produced mixed findings. Many studies that examined predictive validity across race/ethnicity found that race moderated the relationship between risk score and recidivism (e.g., Fass et al., 2008; Onifade et al., 2009; Schmidt et al., 2006; Schwalbe et al., 2007; Vincent et al., 2011); while others did not find race/ethnicity to be a significant moderator (Jung & Rawana, 1999; Meyers & Schmidt, 2008; Schwalbe, 2009).

One study investigated the predictive validity of the YLS/CMI with a sample of White and African American young offenders (Onifade et al., 2009). The YLS/CMI composite risk score was a valid predictor of recidivism for the full sample and by subgroup, however when subscale scores were examined, the risk assessment did not perform equally across race/ethnicity (Onifade et al., 2009). While six of the eight YLS/CMI subscale (offense history, education, peer group, substance abuse, leisure activity, personality) scores demonstrated a statistically significant relationship with re-offense for White males, four out of eight subscale (education, peer group, substance abuse, attitudes) scores reached statistical significance for African American males. Similarly, the subscale scores for White females reached significance for five out of eight factors (family circumstances, education, peer group, personality, attitude) while three out of eight subscale (family circumstances, education, personality) scores were significant predictors of re-offending for African American females (Onifade et al., 2009). Although recidivism rates were higher for African American juvenile offenders, this study indicated that the YLS/CMI had statistically weaker predictive validity for this subgroup when compared to White juvenile offenders. The authors concluded that the YLS/CMI might be performing



differently across race/ethnicity because the measure may be lacking risk factors critical for accurate prediction for African Americans. This concept, omitted variable bias, asserts that variables of risk such as socioeconomic status, differential reporting, and increased police surveillance in neighborhoods with high densities of African Americans, should be considered when seeking equal prediction rates (Onifade et al., 2009).

Schwalbe et al. (2006) sought to describe the impact of race/ethnicity on the NCAR's predictive validity using a large adjudicated sample. The authors found that African American males and females recidivated at significantly higher rates than their White counterparts. Although NCAR's eight out of nine risk factors predicted recidivism for the entire sample, only five out of nine risk factors were significant for Black females (Schwalbe et al., 2006). In a separate study, Schwalbe and colleagues (2004) found similar results. Specifically, the study revealed differential predictive validity by race/ethnicity as African American offenders "attained the base rate of recidivism at an approximate risk score of five, compared to non-Latino White youths who attained it an approximate risk score of eight" (Schwalbe et al., 2004, p. 13). This means that the NCAR underpredicted future crime for African Americans. For example, when White youth reached a risk score of 12, there was a sharp increase in recidivism, however recidivism rates for African American youth plateaued at a risk score of 11. African American offenders' recidivism rates were consistently higher than their overall risk scores predicted they should be (Schwalbe et al., 2004).

Although the YLS/CMI and NCAR appear to be the most investigated in regards to differential predictive validity, less common risk assessment instruments have been examined as well. Although the Joint Risk Matrix was specifically designed to increase the predictive validity of risk assessment for diverse populations, it was found to predict differently across race/ethnicity

(Schwalbe et al., 2007). Likewise, the Positive Achievement Change Tool (PACT) (Baglivio & Jackowski, 2012; Baglivio, 2009), the SAVRY (Vincent et al., 2011), and the PCL:YV demonstrated substantial, yet, unequal prediction rates (Schmidt et al., 2006).

On the other hand, a few studies found that risk assessment scales were robust to race/ethnicity (Jung & Rawana, 1999; Meyers & Schmidt, 2008; Schwalbe, 2009;). The inconsistent findings may be due to limitations evident in these studies. A small sample size ( $N = 121$ ) may have contributed to Meyers & Schmidt (2008) varying findings. The authors suggested that a small number of youth in each subgroup may have decreased their statistical power (Meyers & Schmidt, 2008). Jung & Rawana (1999) used a shorter follow-up (6 months) when compared to the studies that found race as a moderator. Although this study demonstrated accurate predictive validity for the full sample, more time to recidivate may have been necessary to see differences across race/ethnicity emerge. Perhaps more curious, is the fact that the sample was drawn from a city in Canada. It is possible that since the risk assessment measure (YLS/CMI) was developed on a Canadian sample that it predicted recidivism well for both Native and non-Native youth in this region (Andrews et al., 2011). However, researchers have found differential validity present in both Canadian and American samples of juvenile offenders (Schmidt et al., 2006; Schwalbe et al., 2006). Evidence of differential predictive validity found in Schmidt et al. (2006) and Schwalbe et al. (2006), may be due to their use of different risk assessment instruments (PCL: YV and NCAR, respectively).

Researchers conducted a validation study of the Arizona Risk/Needs Assessment (ARNA) and found that the measure was robust to race/ethnicity (Schwalbe, 2009). This study is noteworthy because the authors used a suitable follow-up period (12 months), and a large sample of 29, 711 juvenile offenders. However, a limitation of this study is that the researchers did not

conduct reliability checks with the court officers that administered the assessments. Neglecting to assess the ARNA's reliability could account for erroneous outcomes (Schwalbe, 2009).

Risk assessment measures not only help assess needs and risk, guide case planning, and predict future offending, juvenile justice advocates have recommended the adoption of risk assessment tools as a strategy to potentially decrease disproportionate minority contact. Results from studies that examined differential predictive validity demonstrated mixed results. While some researchers found risk assessment scales to be robust to race/ethnicity, others found evidence of unequal predictive performance. However, it is important to keep in mind that if risk assessment instruments are validated by accurately predicting official recidivism in a disproportionate system, then risk assessment measures themselves may be systematically inequitable; therefore making the task of identifying root causes to DMC nothing less than challenging. Potential race/ethnicity bias is not the only issue within the juvenile justice system. Risk assessment measures should have equal prediction rates for all subpopulations, including female offenders.

### Gender and Risk Assessment

It is important to mention the role of gender in the criminal and juvenile justice systems. In 2009, females were involved in 28% of delinquency cases handled by juvenile courts; up from 19% in 1985 (Puzzanchera et al., 2012). Interestingly, between 1985 and 2009, the female delinquency caseload grew at an average rate of 3% per year, while the average rate of increase for males was 1% per year (Puzzanchera et al., 2012). "This average annual growth in the female caseload outpaced that for males for all offense categories between 1985 and 2009" (Puzzanchera et al., 2012, p. 12). These statistics demonstrate that while male offenders are largely overrepresented in the juvenile justice system, the number of female offenders making

contact with the court is occurring at a higher rate. Some scholars believe that this increase in female involvement is due to an increase in risk to engage in delinquent and criminal acts, while others assert that the juvenile and criminal justice systems are exercising harsher sanctions on women and girls, even for less serious offenses (Stevens, Morash, & Chesney-Lind, 2011). This debate could potentially be resolved by examining girls' and women's risk to commit crime. Unfortunately, there is no consensus concerning whether the same patterns of risk emerge for males and females in regards to their criminal trajectories to engage in delinquency or crime.

Risk assessment has been one way to test for patterns of risk among male and female offenders. The idea is that if risk assessment measures can predict recidivism equally across gender, there would be reason to believe that risk factors do not vary for male and female offenders. Studies with separate female analyses have found that widely-used risk assessment measures predict recidivism the same across gender (Flores et al., 2003; Jung & Rawana, 1999; Meyers & Schmidt, 2008; Olver et al., 2012; Olver et al., 2009; Onifade et al., 2010; Schmidt et al., 2011; Schwalbe, 2008). Conversely, other studies demonstrated that risk assessment tools performed poorly for female offenders when compared to males (Betchel et al., 2007; Onifade et al., 2008; Schmidt et al., 2005; Schmidt et al., 2006; Schmidt et al., 2011). Andrews et al. (2012) found that the LS/CMI performed considerably better for female offenders (Males: mean  $r = .39$ ; mean AUC = .75; Females: mean  $r = .53$ ; mean AUC = .83).

Although some studies have found that existing risk assessment tools have the ability to predict risk similarly across gender, some scholars assert that these tools fail to include risk factors that are related specifically to female criminality (Bloom et al., 2002; Chesney-Lind et al., 2008). For example, Voorhis, Wright, Salisbury, & Bauman (2010) tested the relationship between gender-responsive risk factors and recidivism. They found that parental stress, self-

esteem, self-efficacy, family support, mental health, relationship dysfunction, and child abuse predicted reoffending among female offenders (Voorhis et al., 2010). A study conducted by Simpson, Yahner, & Dugan (2008) sought to replicate female criminal pathways to crime identified by Daly (1994) and found evidence that suggested criminal pathways may be gendered. Specifically, a principle components analysis was conducted with a sample of high-risk, mostly African American women revealing a substantial overlap with the typologies identified in Daly's (1994) study (as cited in Simpson et al., 2008). The typologies included *Harmed and harming women*, *Drug-connected women*, and *Battered women*. *Harmed and harming women* experienced a chaotic home life, abuse and/or neglect as children, early drug/alcohol abuse, and showed symptoms of emotional and psychological damage. *Drug-connected women* did not have extensive criminal histories and were characterized as having used drugs experimentally and sold drugs in association with family. Finally, *Battered women* were in a violent relationship and their criminal activity was connected to their violent relationship (Simpson et al., 2008).

It is imperative that more research is conducted examining the potential differences for male and female offenders. Not only to obtain accurate, and equal predictions of risk, but because these differences may also indicate a difference in needs. In Bloom et al. (2002), the authors recommended that gender-responsive policies be implemented. The suggested strategies included acknowledging that gender makes a difference by prioritizing issues specific to female offenders; develop practices, programs, and policies that promote healthy relationships by developing community and peer-support networks; and improve women's economic and social conditions by developing their capacity to be self-sufficient (Bloom et al., 2002). Implementing

these strategies (among others) will ensure that female offenders' needs are addressed and that they are receiving appropriate treatment while under court supervision.

One way to improve overall prediction rates for all groups is by assessing risk over time. Currently, the bulk of the literature utilizes the initial risk score assigned to youth to predict recidivism. The next section will discuss the paucity of studies that examined the validity of reassessment scores.

### *Validity of Reassessment Risk Scores*

Although it is evident that several dynamic and static criminogenic risk factors predict reoffending behavior, they do not account for the total variance in recidivism. According to Andrews and colleagues (2011), derivations of the LS/CMI demonstrated a range of 15-35% total variance explained based on the AUC statistic. Validation studies generally assess risk based on the risk score youth received once they entered the juvenile court system (Krysick & LeCroy, 2002; Onifade et al., 2008; Schlager & Simourd, 2007). This initial risk score is used to predict outcomes such as future recidivism, time to new offense, and successful program completion. While the initial risk score has proved to be an accurate predictor of recidivism (Betchel et al., 2007; Brennan et al., 2009; Catchpole & Gretton, 2003; Luong & Wormith, 2011; Meyers & Schmidt, 2008; McGrath & Thompson, 2012; Olver et al., 2011; Onifade et al., 2009; Onifade et al., 2008; Schmidt et al., 2011; Schmidt et al., 2006; Schmidt et al., 2005; Schwalbe, 2009; Schwalbe et al., 2007; Shaffer et al., 2011; Welsh et al., 2008), another way to account for more variance may be to reassess youths' level of risk during court supervision and when they exit the court system, post-supervision. There are two important issues to consider. First, is the proximity of the predictor variable (risk assessment) to the criteria (re-offending). It is understood that as time progresses risk to reoffend may change. Exclusively utilizing the initial

risk score as the predictor variable opposed to the risk score most proximal to the re-offense may lead to less accurate predictions. Second, utilizing risk assessment conducted at entry to predict re-offending, and hence before court intervention, may confound the relationship between court intervention and re-offending, with initial risk assessment and re-offending. It is important to disentangle these two effects.

To date, only two studies have attempted to assess the validity of a composite risk score following court supervision (Baglivio & Jackowski, 2012; Flores et al., 2003). Employing the Positive Achievement Change Tool (PACT), Baglivio & Jackowski (2012) investigated the measure's ability to predict new arrests exclusively using the exit risk assessment score. In other words, no comparative analyses of entry and post assessment risk scores were completed. These authors also explored the differential predictive validity across race/ethnicity and gender using a 12-month follow up and a sample of 15, 072 young offenders (Baglivio & Jackowski, 2012). The PACT's post supervision score modestly predicted recidivism for each subgroup with AUCs ranging from of .57 to .62. The authors suggested that their modest findings might be due to their use of a selective sample of successful probation completions.

Flores et al. (2003) sought to identify the YLS/CMI as a valid predictor of case outcome for juvenile delinquents under correctional supervision. In addition, the authors sought to identify whether correctional treatment was associated with reductions in recidivism by comparing initial, and a combination of reassessment and post-supervision risk scores (Flores et al., 2003). The reassessments were administered either one year following the initial assessment or at the time of program completion. Results indicated that the relationship between treatment delivery and change in risk score was not statistically significant. Likewise, the relationship between treatment completion and change in score was not significant (Flores et al., 2003). Meaning, that having

participated or having completed treatment under correctional supervision did not produce a statistically significant score reduction. However, the authors noted that scores from the first assessment to the reassessment did decrease, indicating that although nonsignificant, supervision may have made an impact in a “theoretically relevant” direction (Flores et al., 2003). Finally, the authors confirmed the predictive utility of the post-supervision/reassessment scores as it was significantly correlated with recidivism outcomes. Conversely, it is important to note that these findings should be interpreted with caution, as the number of youth with both initial and post-supervision/reassessment scores was small ( $n = 87$ ). In addition, the authors failed to conduct analyses verifying there were no systematic differences between youth with reassessments and those without (Flores et al., 2003). Nevertheless, this study is pivotal as it attempted to provide evidence of the predictive utility of post-supervision/reassessment scores with a juvenile sample.

Many researchers have discussed the importance of reassessment (Andrews et al., 2006; Baglivio & Jackowski, 2012; Baglivio, 2009; Bonta et al., 2008; Flores et al., 2003; Lowenkamp & Betchel, 2007; Olver et al., 2009; Schlager & Pacheco, 2011; Schmidt et al., 2011; Schmidt et al., 2005). The importance of implementing and validating reassessments is four-fold. First, the use of reassessments can assist practitioners in monitoring the appropriateness of their case management plan for a particular offender. If the reassessment provides evidence that a particular intervention is not meeting an offender’s needs, the practitioner can adjust the level of supervision or intervention plan accordingly (Lowenkamp & Betchel, 2007). Second, utilizing reassessments, especially post-court supervision, can serve as an evaluative tool. For example, recognizing that participation in certain programs lead to higher levels of risk post-supervision may trigger court officers to conduct an in-depth evaluation on treatment effectiveness (Bonta et al., 2008). Third, Andrews et al. (2006) discussed unpublished dissertations that provided



support that reassessment of dynamic risk factors can substantially improve the predictive criterion validity of risk assessment. They claimed that “[b]ased on the available evidence, we anticipate reassessments will double and, perhaps, triple the outcome variance explained by intake assessments” (Andrews et al., 2006). Ultimately, implementing reassessments could lead to better predictions of reoffending behaviors. Finally, examining exit assessments can help practitioners gain a more accurate picture of reoffending by using the risk score that includes the incidental influence of court supervision. Given this recommendation, it is unfortunate that only 1 out of 34 studies (see Table 1) have exclusively examined exit assessment scores. This gap in the extant literature was examined in the current study.

Recently, two studies explored the feasibility of measuring risk to recidivate over time (Schlager & Pacheco, 2011; Vose et al., 2009). These studies were less focused on the validity of the reassessment score and highlighted the relationship to recidivism with change in risk scores from two points in time. While Schlager & Pacheco’s (2011) adult offenders were reassessed six months after the initial assessment, the Vose and colleagues (2009) sample was reassessed an average of 12 months following the initial assessment. As a result of court supervision, there were reductions in risk score from Time 1 to Time 2 in both studies. One of the studies demonstrated that the relationship between risk score and recidivism was stronger for the Time 2 assessment (Vose et al., 2009). Considering that the magnitude of the relationship between risk score and recidivism increased from Time 1 to Time 2, this may provide evidence of the increased predictive utility of reassessment scores.

The literature review contained information on both adult and juvenile risk assessment validation studies. The articles reviewed set the framework for the development of the current study. Table 2 illustrates a summary of the methodological details presented in Table 1. Of the

34 validation studies reviewed, 25 studies used juvenile samples. The average number of participants for these studies was 3,378 (median  $N = 480$ ). Of the studies that reported AUC statistics (16 out of 25), the average weighted AUC was .63. More than half of these studies reported a follow up length of longer than 12 months. Most important to this study is the summary information regarding intake validity, exit validity, race validity, and gender validity also found in Table 2. In the subsequent section, details of the current study--including research questions, the study's significance, and methodological critiques, are presented.

Table 1. Risk Assessment Validation Studies

<b>Author(s)</b>	<b># of participants</b>	<b>% non-White</b>	<b>% female</b>	<b>Risk Measure</b>	<b>AUC</b>	<b>Intake Validity</b>	<b>Exit Validity</b>	<b>Validity by race</b>	<b>Validity by Gender</b>	<b>Time</b>
Jung, 1999	263	51	34	MRNAF (YLS/CMI)	N/A	Yes	No	Yes	Yes	6 mos
Krysick et al., 2002	7,001	45	35	NCCD	N/A	Yes	No	No	No	1 year
Catchpole et al., 2003	74	45	15	YLS/CMI SAVRY PCL	.74 .74 .78	Yes	No	No	No	1 year
Flores et al., 2003	1,679	30	21	YLS/CMI	N/A	Yes	No	Yes	Yes	2 years
Schwalbe et al., 2004	464	50	25	NCAR	N/A	Yes	No	No	No	1 year
Gavazzi et al., 2005	399	89	61	GRAD	N/A	Yes	No	Yes	Yes	1 year
Schmidt et al., 2005	107	29	37	YLS/CMI	.61	Yes	No	No	No	3 years (avg.)
Holsinger et al., 2006	403	35	35	LSI-R	N/A	Yes	No	Yes	Yes	17 mos

Table 1 (cont'd)

Schmidt et al., 2006	130	30	30	PCL	.71	Yes	No	Yes	Yes	3 years (avg.)
Schwalbe et al., 2006	9,534	54	38	NCAR	N/A	Yes	No	Yes	Yes	1 year
Betchel et al., 2007	4,482	47	14	YLS/CMI	.60	Yes	No	Yes	Yes	3 years
Schlager et al., 2007	446	100	0	LSI-R	N/A	Yes	No	Yes	No	2 years (avg.)
Schwalbe et al., 2007	536	45	32	NCAR JRM	.68 .71	Yes	No	Yes	Yes	9 mos
Fass et al., 2008	975	86	0	LSI-R COMPAS	.60 .53	Yes	No	Yes	No	1 year
Gavazzi et al., 2008	711	35	39	GRAD	N/A	Yes	No	Yes	Yes	1 year
Meyers, et al., 2008	121	31	34	SAVRY	.75 .76	Yes	No	Yes	Yes	1 year 3 years
Onifade et al., 2008	328	69	27	YLS/CMI	.62	Yes	No	No	No	1 year
Welsh et al., 2008	133	28	36	YLS/CMI SAVRY PCL	.60 .77 .74	Yes	No	No	No	3 years (avg.)

Table 1 (cont'd)

Baglivio, 2009	8,132	39	30	PACT	.59	Yes	No	No	Yes	1 year
Brennan et al., 2009	2,328	24	19	COMPAS	.68	Yes	No	Yes	Yes	4 years
Hsu et al., 2009	78,502	Did not specify	15	LSI-R	N/A	Yes	No	No	Yes	N/A
Onifade et al., 2009	968	47	26	YLS/CMI	.63	Yes	No	Yes	Yes	2 years
Schwalbe, 2009	29, 711	52	36	ARNA	.65	Yes	No	Yes	Yes	1 year
Vose et al., 2009	2,849	15	14	LSI-R	N/A	Yes	No	No	Yes	N/A
Hsu et al., 2011	71,122	Did not specify	15	LSI-R	N/A	Yes	No	No	Yes	N/A
Luong et al., 2011	192	64	27	LSI-SK	.73	Yes	No	No	No	2 years (avg.)
Olver et al., 2011	167	62	44	YLS/CMI	.77	Yes	No	Yes	Yes	7 years (avg.)
Schlager, et al., 2011	179	90	11	LSI-R	N/A	Yes	No	No	No	N/A

Table 1 (cont'd)

Shaffer et al., 2011	830	47	18	RMS	.67	Yes	No	No	Yes	3 years (avg.)
Schmidt et al., 2011	112	31	27	YLS/CMI SAVRY PCL	.66 .74 .79	Yes	No	No	Yes	10 years (avg.)
Vincent et al., 2011	480	64	0	SAVRY	N/A	Yes	No	Yes	No	5 years
Baglivio et al., 2012	15,072	N/A	14	PACT	.59	No	Yes	Yes	Yes	1 year
McGrath et al., 2012	3,568	56	15	YLS/CMI	.65	Yes	No	No	No	1 year
Vitopolous, 2012	76	63	49	YLS/CMI	N/A	Yes	No	No	Yes	3 years

Table 2. Risk Assessment Validation Studies Summary

<b>Population</b>	<b># of studies</b>	<b>N</b>	<b>Non-White</b>	<b>Female</b>	<b>* AUC</b>	<b>** AUC</b>	<b>Intake Validity</b>	<b>Exit Validity</b>	<b>Gender Validity</b>	<b>Race validity</b>	<b>1 year</b>	<b>&gt; 1 year</b>
Juveniles	25	3,378	49	27	.63	.69	96	.04	68	56	.48	.52
Adults	9	15,782	37	7	.63	.62	100	0	55	44	20	80

*Note.* Represents average percentages for each category. \* Weighted and \*\* unweighted AUCs calculated based on the number of adult and juvenile studies that reported AUCs (k = 4 and 16, respectively); Percentage of non-Whites participating based on the number of adult and juvenile studies that reported racial composition (k = 5 and 24, respectively)

## Current Study

The current study increased our knowledge of risk assessment utility by investigating the relative predictive validity of Youth Level of Service/Case Management Inventory (YLS/CMI) risk assessment scores prior and subsequent to court supervision. This study also examined the differential predictive validity of risk assessment scores by race/ethnicity and gender. The following research questions were addressed.

1. Do initial and exit YLS/CMI risk scores differ in mean level and variability?
2. Are YLS/CMI risk scores assessed at exit from court supervision differentially valid predictors of recidivism?
3. Do race/ethnicity and gender moderate the relationship between risk and recidivism for initial and exit scores?
4. What is the relative predictive validity of change in risk scores and exit risk scores?
5. Does time under court supervision moderate the relationship between risk and recidivism? Does time under court supervision moderate the relationship between change in risk scores and recidivism?

### *Significance of the Current Study*

The literature demonstrates that there is a substantial amount of unexplained variation in recidivism. Although many researchers have argued that assessing risk scores over time may improve our ability to predict future reoffending (Andrews et al., 2006; Bonta et al., 2008; Lowenkamp & Betchel, 2007; Onifade et al., 2011), the literature is limited to 3 (out of 34) studies that examined risk scores over time. While these studies have made valuable contributions to the risk assessment literature by examining the validity of change in reassessment scores over time, they do not directly align with the goals of the current research.



First, the two most recent studies utilized adult populations (Schlager & Pacheco, 2011; Vose et al., 2009). Second, all three studies did not exclusively examine the validity of the risk assessment scores following court supervision.

Most relevant to the current research, only one study utilized a juvenile sample (Flores et al., 2003). Similar to the current project, the authors examined reassessment scores; however, the study was limited because only a portion of the sample was reassessed upon dismissal from the court. Another critical drawback was the potentially non-representative nature of the study's sample. Due to the small proportion of youth who completed both entry and exit assessments when compared to those youth who completed only an entry assessment, the authors were unable to conduct analyses that would detect statistical differences between the two groups. Put simply, the authors did not investigate potential systematic differences between the youth who had two scores and those youth who only had one (Flores et al., 2003).

The current study contributes to the risk assessment literature in several important ways. First, this study increases our understanding of the relationship between reassessment scores and recidivism among youth by employing a sample of juvenile offenders. As previously stated, of 34 studies, only 1 (Flores et al., 2003) utilized a juvenile sample to examine reassessment risk scores. Next, the current study builds on existing research by highlighting the predictive accuracy of the risk score received upon exiting the court. Table 2 shows that while 96% of studies reviewed utilized the initial risk score to predict recidivism, only one study (Baglivio & Jackowski, 2012) exclusively examined the validity of the scores youth received upon exiting the court. Moreover, by utilizing improved proportions of youth with both an entry and exit assessment, this study seeks to draw improved conclusions from the results. Since 2004, the court in question has assessed every juvenile offender at entry to the court as a part of the intake

process (1,559 youth in the delinquency division). Of those youth assessed at entry, 306 received an assessment upon exiting the system. In Flores and colleagues (2003), out of the total 1,507 assessed at entry to the system, 87 were reassessed. In other words, while the current study's sample represents 20% of possible entry and exit score matches, Flores et al. (2003) sample represented 6%. While investigating 20% of the total sample is not ideal, it is an attempt to draw more accurate conclusions.

This research study also fills a gap in the existing literature by examining whether the validity of the post-supervision risk scores varies across race/ethnicity and gender. As shown in Table 2, exploring race/ethnicity subgroup differences only occurred for 56% of the reviewed studies; all of which utilized the initial risk score to predict future reoffending. Validation studies investigated gender as a moderator more often, 68% of the time. Of the three studies that investigated reassessment scores, none of them explored potential subgroup differences. Exploring whether the relationship between exit risk score and reoffending is moderated by race and/or gender may indicate that court supervision has varying effects on recidivism based on subgroup membership.

Finally, the current research provides an initial step towards identifying the impact of court supervision on recidivism outcomes. Theoretically, juvenile offender risk scores should decrease as a result of court programming and services. There is evidence to support that if an offender's treatment needs are met by participating in court programming, reoffending rates will decrease (Bonta et al., 2008; Luong & Wormith, 2011; Vieira et al., 2009). This study can offer local county court officials comprehensive information about the comparative predictive accuracy of entry and exit risk assessment scores, the impact of court programming on reoffending, and the potential differential effects of court supervision across race/ethnicity.

## METHODS

### *Sample*

The study was conducted using secondary data. The data came from the delinquency division of a Midwestern, mid-sized juvenile court. There were no refusals or duplicate cases. The overall juvenile population includes 1,559 youth assessed at entry to the court. The court data manager identified 266 youth with corresponding initial and exit YLS/CMI assessments based on the initial petition. Several tests were conducted in order to identify any systematic differences between the study's sample and the total sample by age, gender, race/ethnicity, YLS/CMI initial risk score, risk level, and one-year recidivism rates. With the exception of risk level ( $\chi^2 (2, N = 1,557) = 7.86, p < .05$ , Cramer's  $V = .07$ ), which indicated that youth with both initial and exit scores were more serious offenders, there were no systematic differences between the current sample and the overall juvenile population for the court in question. The current sample ( $N = 211$ ) represents probationers who received an YLS/CMI assessment at both entry and upon dismissal from court supervision between August 2005 and September 2011. This timeframe allowed recidivism to be assessed for 12 months following both the initial and exit risk assessment score (55 youth were not included in subsequent analyses because they had not reached the 12-month follow up criteria). Table 3 provides descriptive information on the sample.

Table 3. Sample Characteristics

Variable	N	Mean
Age	211	14.6
Gender		
Male	150	71.1
Female	61	28.9
Race/Ethnicity		
White	84	39.8

Table 3 (cont'd)

Non-White	127	60.2
-----------	-----	------

### *Training and Procedures*

Over the course of four days, each juvenile court officer received 32 hours of training on how to administer and score the YLS/CMI. This training took place prior to using the instrument. Training activities included providing definitions, clarifying the protocol and scoring guide, explaining what each item measures, mock interviews, and coding previously taped cases. Inter-rater reliability checks were performed quarterly and consistently reached at least 90% exact agreement.

### *Measures*

The Youth Level of Service/Case Management Inventory (YLS/CMI) is a third-generation risk assessment tool that was created by Hoge & Andrews, (2002) to predict general recidivism for young offenders aged 12-18. The YLS/CMI has 42 items that are divided into eight subscales. The subscales are Prior/Current Offenses (5 items), Education (7 items), Leisure & Recreation (3 items), Peer Relations (4 items), Substance Abuse (5 items), Family & Parenting (6 items), Attitudes & Orientation (5 items), and Personality (7 items) (Schmidt et al., 2005). Each item is scored dichotomously (yes or no) indicating whether or not risk is present. The items are totaled and the composite score is translated into a level of risk; low, moderate, high, and very high (Flores et al., 2003). The Appendix includes the 42 items of the YLS/CMI.

The current study's outcome of interest is recidivism. Recidivism was defined as any new court petitions received subsequent to the administration of the YLS/CMI. Recidivism was coded as either 1 (new petition) or 0 (no new petitions) based on whether the juvenile offender reoffended during the 12-month follow up period subsequent to the date he/she received an

initial and exit YLS/CMI assessment. Recidivism data were collected through the court data management system. Both juvenile and adult records were checked in order to track any new petitions/arrests acquired in the event that an offender reoffended upon aging out of the juvenile justice system.

## RESULTS

As a precursor to examining research questions directly, bivariate correlations of initial composite score, exit composite score, change in risk scores, time under supervision, recidivism, gender, and race are illustrated in Table 4. In addition, descriptive statistics were calculated for composite risk scores, risk levels, and recidivism rates (see Table 5). While the large majority of offenders were categorized as moderate risk at entry to the court, those youth categorized as moderate risk upon exiting the court dropped to a little more than half. There were also notable changes in the remaining categories as the percentage of youth represented in the low risk level considerably increased over time. Similarly, a sharp decline was demonstrated among youth assigned a high-risk level. A significant McNemar-Bowker chi-square  $X^2(3, N = 211) = 63.63$ ,  $p < .05$ , Cramer's  $V = .33$  identified significant changes in risk level from entry to exit from court supervision. Crosstabs were conducted to identify differences in one-year recidivism rates based on the entry and exit composite risk scores. A significant McNemar chi-square  $X^2(2, N = 211) p < .05$ ,  $\phi = .35$  demonstrated that recidivism rates post-court involvement were statistically higher than post-initial YLS/CMI recidivism rates.

Table 4. Correlations for Independent and Dependent Variables

	Initial	Exit	Change	Time	Rec 1	Rec 2	Gender	Race
Initial	---							
Exit	.49**	---						
Change	.16**	-.46**	---					
Time	.30**	.25**	.25**	---				
Rec 1	.14*	.31**	.23**	.30**	---			

Table 4 (cont'd)

Rec 2	.11	.13	.04	-.05	.35**	---		
Gender	-.18**	.01	.05	.03*	.13	.17*	---	
Race	-.07	.11	.06	-.02	.09	.03	.02	---

*Note.* Rec 1 = Recidivism post-initial; Rec 2 = Recidivism post-exit; Change = Difference Score

Table 5. YLS/CMI Initial and Exit Mean Risk Scores, Risk Levels, and Recidivism Rates

	Initial YLS/CMI	Exit YLS/CMI
<i>M (SD)</i>	16.6 (7.2)	11.2 (6.6)
Risk Level	<i>n (%)</i>	<i>n (%)</i>
Low	29 (13.7)	86 (40.8)
Moderate	138 (65.4)	109 (51.7)
High	44 (20.9)	16 (7.6)
Recidivism Rates		
Recidivists	135 (64)	163 (77.3)
non-Recidivists	76 (36)	48 (22.7)

*Do initial and exit YLS/CMI risk scores differ in mean level and variability?*

A paired-samples t-test was conducted to examine mean differences between the entry and exit scores. The analysis indicated that the entry and exit composite risk scores were significantly different from each other  $t(210) = 11.53, p < .05$ , Cohen's  $d = .77$ . Specifically, the mean of the total scores decreased by 5.4 between the time offenders entered and exited court supervision. In order to test for homogeneity of variance across the two composite scores, the Pitman-Morgan test was conducted. While SPSS provides homogeneity of variance tests (such as Levene's) when running an ANOVA or independent sample's t-test, no such analysis is offered when comparing means in a dependent sample's t-test. The Pitman-Morgan test was designed to identify differences in variability when using paired-samples (Gardner, 2001). The test revealed

no significant differences in variability between the entry and exit risk scores  $t(209) = 1.35, p > .05$ .

*Are YLS/CMI risk scores assessed at exit from court supervision differentially valid predictors of recidivism?*

In order to investigate the predictive validity of the initial and exit composite risk scores, a Receiver Operating Characteristic/Area Under the Curve (ROC/AUC) analysis was implemented. This test specifies the proportion of true positives, or the number of offenders predicted to reoffend that did in fact commit a future offense, to the number of true negatives, or the number of youth predicted to not reoffend that indeed did not commit a future crime. This statistic is useful when comparing the predictive validity across samples because it controls for base rates of the criterion variable (Rice & Harris, 1995). The AUC can range from 0.0 indicating no predictive validity, to 1.0 demonstrating perfect validity. This statistic calculates the probability that a randomly selected recidivist would score higher on a risk assessment scale than a randomly selected non-recidivist. In other words, an AUC above .50 indicates that the predictive validity of the measure is better than chance (Rice & Harris, 1995). Rice & Harris (2005) described AUC values of .556 as small, .639 as moderate, and .714 as large predictive validity effect sizes. The AUCs for both composite scores are presented in Table 6.

Table 6. Area Under the Curve for Initial and Exit Risk Scores

Assessment	AUC	SE	P value	Confidence Intervals
Initial Scores	.59	.04	.03	.51 - .67
Exit Scores	.58	.05	.08	.50 - .67

$p < .05$

As illustrated, only the initial total scores yielded an Area Under the Curve with a  $p < .05$ . It is important to note the significant overlap of the confidence intervals across the two AUC



statistics. Additional analyses to compare the predictive validity of each score were conducted and results indicated that the AUCs did not statistically differ. Therefore, risk scores assessed at exit from court supervision are not differentially valid predictors of recidivism.

To confirm the results of the ROC/AUC analysis, a binary logistic regression was employed in which recidivism one-year post-initial assessment was regressed on initial YLS/CMI scores for the first model. In the second model, the exit YLS/CMI scores were included as the independent variable to predict recidivism one-year post-exit assessment. Table 7 shows the regression results for both models. As illustrated, initial YLS/CMI scores significantly predicted the outcome variable with an  $OR = 1.07$ ,  $CI [1.00, 1.08]$  indicating that for every one point increase in initial risk score, offenders are 1.07 times more likely to reoffend. As demonstrated, the exit YLS/CMI scores approached significance, and were not differentially valid predictors of recidivism, as indicated by the significant overlap of the CIs for both composite scores ( $OR = 1.05$ ,  $CI [1.00, 1.10]$ ).

Table 7. Logistic Regression Predicting Recidivism by Initial and Exit Risk Scores

Variable	<i>B</i>	<i>SE</i>	Wald	P value	<i>Exp(B)</i>
Initial risk	.04	.02	3.98	.05	1.04
Constant	-1.23	.38	11.07	.01	.28
<i>-2 Log Likelihood = 271.71; <math>X^2 = 12.33</math>; Cox &amp; Snell <math>R^2 = .02</math>; Nagelkerke <math>R^2 = .03</math></i>					
Exit risk	.05	.02	3.53	.06	1.05
Constant	-1.75	.34	27.02	.01	.17
<i>-2 Log Likelihood = 222.79; <math>X^2 = 19.88</math>; Cox &amp; Snell <math>R^2 = .02</math>; Nagelkerke <math>R^2 = .03</math></i>					

*Do race/ethnicity and gender moderate the relationship between risk and recidivism for initial and exit scores?*

The ROC/AUC analysis was used to test the differential predictive validity of the YLS/CMI initial and exit risk scores across race and gender. Race/ethnicity was recoded into a dichotomous variable where youth were divided into White and non-White categories. Juvenile

offenders identified as Caucasian (40%) during initial YLS/CMI administration were coded as “White.” Youth in the non-White category were identified as one of the following: African-American (36%), Hispanic/Latino (12%), Multi-racial (11%), Other (1%). Female offenders were coded 0 and male offenders were coded 1. The AUCs for each subgroup can be found in Table 8. As illustrated, every subpopulation yielded an AUC above .50. When examining the initial composite scores, only the AUC statistics for non-White youth and males reached statistical significance ( $p < .05$ ). No AUC statistics for the exit composite scores reached significance for any of the other subgroups ( $p > .05$ ). Post-hoc comparison analyses were conducted to test the statistical differences of subgroup AUCs within and between initial and exit composite scores. None of the AUC statistics were significantly different from the other.

Table 8. Subgroup Area Under the Curves for Initial and Exit Risk Scores

	Males	Females	White	Non-White
AUC ( <i>SE</i> )				
Initial Scores	.61 <sup>*</sup> (.05)	.62 (.07)	.58 (.07)	.61 <sup>*</sup> (.05)
Exit Scores	.59 (.05)	.55 (1.00)	.51 (.08)	.61 (.06)

<sup>\*</sup>  
 $p < .05$

A moderated binary logistic regression was employed to confirm the results of the ROC/AUC analysis (See Table 9). Initial risk scores, gender, and the product of gender and initial risk scores were included as variables in the first model. Initial risk scores, race/ethnicity, and the product of initial risk scores and race/ethnicity were included as variables in the second model. Consistent with the results of the ROC/AUC analyses, gender and race/ethnicity did not moderate the relationship between initial YLS/CMI risk scores and recidivism. To determine whether gender or race/ethnicity moderated the exit risk-recidivism relationship, exit risk scores, gender and exit risk scores by gender were entered as covariates in the first model. Exit risk

scores, race/ethnicity, and the product of exit risk scores and race/ethnicity were entered as variables into the second model. Similarly, gender and race/ethnicity were not found to moderate the relationship between exit risk scores and the outcome variable, and none of the interaction variables significantly differed from each other.

Table 9. Logistic Regression Predicting Recidivism by Gender and Race/Ethnicity

Variable	<i>B</i>	<i>SE</i>	Wald	P value	<i>Exp(B)</i>
Initial risk	.03	.03	8.44	.36	1.03
Race	.06	.80	.01	.94	1.07
Initial X Race	.02	.04	.24	.63	1.02
Constant	-1.34	.64	4.41	.04	.26
<i>-2 Log Likelihood = 269.50; <math>X^2 = 5.50</math>; Cox &amp; Snell <math>R^2 = .03</math>; Nagelkerke <math>R^2 = .04</math></i>					
Initial risk	.06	.05	1.40	.24	1.06
Gender	.94	1.07	.77	.38	2.56
Initial X Gender	-.01	.05	.03	.88	.99
Constant	-2.12	.98	4.62	.03	.12
<i>-2 Log Likelihood = 266.29; <math>X^2 = 5.05</math>; Cox &amp; Snell <math>R^2 = .04</math>; Nagelkerke <math>R^2 = .06</math></i>					
Exit risk	.04	.05	.58	.45	1.04
Race	-.11	.71	.02	.88	.90
Exit X Race	.01	.05	.07	.80	1.01
Constant	-1.67	.56	8.75	.01	.19
<i>-2 Log Likelihood = 222.70; <math>X^2 = 23.41</math>; Cox &amp; Snell <math>R^2 = .02</math>; Nagelkerke <math>R^2 = .03</math></i>					
Exit risk	.05	.06	.66	.42	1.05
Gender	1.16	.95	1.49	.22	3.18
Exit X Gender	-.01	.07	.01	.91	.99
Constant	-2.64	.88	9.08	.01	.07
<i>-2 Log Likelihood = 216.04; <math>X^2 = 17.45</math>; Cox &amp; Snell <math>R^2 = .05</math>; Nagelkerke <math>R^2 = .07</math></i>					

*p* < .05

*What is the relative predictive validity of change in risk scores and exit risk scores?*

A paired samples t-test showed that the exit scores and change in raw scores were statistically different from each other  $t(210) = 7.22, p < .001$ , Cohen's  $d = .50$ . A binary logistic regression was used to compare the predictive utility of overall change in scores and the exit risk

scores. Prior to running the logistic regression, raw change scores were standardized and converted into a categorical variable. Z-scores more than one standard deviation below the mean denoted risk scores that increased, or worsened over time. Z-scores that fell between -1 and 1 standard deviation indicated that risk scores did not change over time. Z-scores greater than one standard deviation above the mean denoted risk scores decreased, or improved over time.

YLS/CMI exit scores were used to predict one-year recidivism following dismissal from the court in the first model. Standardized change scores were entered as a covariate to predict recidivism in the second model. The odds ratio for the change scores variable did not statistically predict recidivism (see Table 10). The odds ratios for the exit composite scores approached significance ( $OR = 1.05$ ,  $CI [1.00, 1.10]$ ). In other words, exit risk scores demonstrated predictive validity better than the change scores, but not in a statistically reliable way.

Table 10. Logistic Regression Predicting Recidivism by Change Scores and Exit Scores

Variable	<i>B</i>	<i>SE</i>	Wald	P value	<i>Exp(B)</i>
Change			.45	.80	
Change (1)	.04	.47	.01	.93	1.04
Change (2)	.33	.49	.45	.50	1.44
Constant	-1.27	.19	42.77	.01	.28
<i>-2 Log Likelihood = 225.85; <math>X^2 = .01</math>; Cox &amp; Snell <math>R^2 = .002</math>; Nagelkerke <math>R^2 = .003</math></i>					
Exit score	.05	.02	3.53	.06	1.05
Constant	-1.75	.34	27.02	.01	.17
<i>-2 Log Likelihood = 222.79; <math>X^2 = 19.88</math>; Cox &amp; Snell <math>R^2 = .02</math>; Nagelkerke <math>R^2 = .03</math></i>					

Note. (1) = Better score; (2) = Worse score;  $p < .05$

*Does time under court supervision moderate the relationship between risk and recidivism? Does time under court supervision moderate the relationship between change in risk scores and recidivism?*

The number of days spent under supervision ranged from 0 to 1,705 days ( $M = 360$ ). A categorical variable was computed based on the number of days each youth spent under court

supervision; the sample was divided into three groups: shortest, medium, and longest length of time. This variable was dummy coded with the shortest length of time under supervision being treated as the reference category. Levels of risk for each group can be found in Table 11.

Table 11. Risk Level Composition for Time Under Supervision Categories

Time Categories	Short	Medium	Long
Range of Days	0-179	181-409	411-1705
Mean ( <i>SD</i> )	121 (47)	269 (69)	686 (307)
Low Risk (%)	25.7	7.1	8.56
Moderate Risk (%)	62.9	70.0	63.4
High Risk (%)	11.4	22.9	28.2

To investigate whether the amount of time spent under court supervision moderates both the risk-recidivism and change in risk-recidivism relationship, a moderated binary logistic regression was implemented. YLS/CMI exit scores, time under supervision and an interaction variable (exit scores by time under supervision) were entered as covariates to predict one-year recidivism following dismissal from the court. The results of the logistic regression revealed that the exit scores by time under supervision variable did not reach statistical significance (see Table 12), indicating that the relationship between risk and recidivism is not moderated by time under supervision.

Table 12. Logistic Regression Predicting Recidivism by Time Under Supervision

Variable	<i>B</i>	<i>SE</i>	Wald	P value	<i>Exp(B)</i>
Exit risk	.05	.05	1.17	.28	1.05
Time			.88	.64	
Time (1)	.34	.75	.209	.65	3.22
Time (2)	-.56	.95	.34	.56	.58
Exit X Time			.02	.99	
Exit X Time (1)	-.004	.06	.005	.94	1.00
Exit X Time (2)	.004	.07	.003	.95	1.00
Constant	-3.13	.68	21.40	.00	.04

$-2 \text{ Log Likelihood} = 219.05$ ;  $X^2 = 7.73$ ;  $\text{Cox \& Snell } R^2 = .034$ ;  $\text{Nagelkerke } R^2 = .051$

*Note.* Time (1) = medium time; Time (2) = long time;  $p < .05$

One-year recidivism was regressed on standardized change scores, time under supervision, and an interaction variable (change scores by time under supervision). As illustrated in Table 13, the regression model indicated that the interaction variable, change scores by time under supervision, did not reach statistical significance. Therefore, time under supervision was not found to moderate the change in risk-recidivism relationship.

Table 13. Logistic Regression Predicting Recidivism by Change in Risk Scores

Variable	<i>B</i>	<i>SE</i>	Wald	P value	<i>Exp(B)</i>
Change			2.14	.34	1.13
Change (1)	1.45	1.05	1.90	.17	4.25
Change (2)	.75	1.27	.35	.55	2.13
Time			3.19	.20	
Time (1)	.63	.45	1.99	.16	1.88
Time (2)	-.19	.52	.13	.72	.83
Change X Time			2.78	.83	
Change (1) X Time (1)	-2.13	1.35	2.51	.11	.12
Change (1) X Time (2)	-1.28	1.30	.97	.33	.28
Change (2) X Time (1)	-.34	1.46	.06	.82	.71
Change (2) X Time (2)	-.73	1.54	.22	.64	.48
Constant	-1.45	.32	20.34	.01	.24

$-2 \text{ Log Likelihood} = 219.43$ ;  $X^2 = .01$ ;  $\text{Cox \& Snell } R^2 = .03$ ;  $\text{Nagelkerke } R^2 = .05$

*Note.* Time (1)= medium; Time (2)= long; Change (1) = better scores; Change (2) = worse scores;  $p < .05$

## **DISCUSSION**

To date, risk assessment research has a primary focus on the predictive value of the risk score youth receive when making initial contact with the justice system. This exploratory study was an attempt to identify the relative predictive validity of the YLS/CMI risk assessment scores youth received post-court involvement. The author first sought to determine whether initial and exit risk scores differed in mean level and variability. As expected, mean level risk scores significantly decreased from entry to upon exiting the court and there was no significant difference in variability between the initial and exit scores. While the expectation that the initial and exit YLS/CMI risk scores would significantly differ from each other was supported, this study did not confirm any other research hypotheses.

The author also sought to examine whether exit risk scores were differentially valid predictors of recidivism. Differential validity was expected under the assumption that exit scores may have provided a more reliable measure of risk. Exit scores were presumed to be a more accurate measure of risk primarily because Juvenile Court Officers received quarterly trainings in administering the YLS/CMI. Welsh and colleagues (2008) discussed the importance of “quality control checks or booster sessions” for court officials as it relates to risk assessment validity and reliability (p.112). Evidence of the importance of well-trained court staff can be found elsewhere (Bonta et al., 2008; Vincent, 2012). Furthermore, differential validity was anticipated because post-initial assessments would increase the likelihood that Juvenile Court Officers become more familiar with the youth’s personality and behavior through multiple face-to-face contacts, and that more information (i.e. relevant records, family observations, school visits) is made available when conducting the assessments. While it is not clear why the initial

and exit score AUC statistics were not significantly different from each other, descriptive statistics offered some interesting insights.

Although the average initial risk score was 17, and the average exit risk score was 11, they did not predict one-year recidivism differently. In fact, whereas initial risk scores were significantly higher than exit scores, the post-exit scores recidivism rates were significantly higher. Onifade and colleagues (2008) found that the predictive validity of YLS/CMI scores decreased as the raw risk score increased. Specifically, scores under 17 were better able to correctly predict reoffenders (Onifade et al., 2008). One explanation for an increase in recidivism rates is that the deterrent effect of court supervision is removed and youth may be more likely to engage in delinquent acts upon exiting the system. Some may argue that being under court supervision increases your chances of getting into trouble as you are constantly under the court's scrutiny; however, this study does not measure probation violations, which may be an indicator of the effect of court supervision. Future studies should consider probation violations as a measure to examine court supervision's impact on recidivism.

Another plausible explanation for decreased levels of risk, but increased recidivism rates may be that Juvenile Court Officers' interventions are not targeting the youth's criminogenic needs. As a result of court intervention, merely being involved in the system could decrease risk domains in which youth obtained low or moderate scores. For example, a youth may decrease his/her moderate risk score in the Education domain with the knowledge that the Juvenile Court Officers can (and actually do) conduct random school visits. On the other hand, if a given offender's high risk scores in the Peer Relations or Leisure/Recreation domains for any reason (e.g. lack of relevant programming, novice court officer) is not addressed, recidivism rates are likely to increase (Bonta et al., 2008; Mears, Cochran, Greenman, Bhati, & Greenwald, 2011;



Vieira et al., 2009). Still plausible is that youth's recidivism rates increased upon exiting the system as a result of undergoing court intervention (Bonta et al., 2008).

The relative predictive validity of exit risk scores and change in risk scores were also examined. Change scores were hypothesized to have superior predictive validity however; results indicated that the exit risk scores performed better at predicting recidivism, although not in a statistically reliable way. Change scores were thought to represent a more nuanced picture of the impact of supervision between entry and upon exiting the court as knowledge of an offender's risk score changes over time would offer more information than the exit score alone. This was not the case, as change in scores did not emerge as a significant predictor of recidivism. One explanation for non-significant findings may stem from the use of difference scores (Initial score – Exit score). Proponents of difference scores argue for their use because they are 1) intuitive and easy to interpret and 2) they are assumed to represent information distinct from its components (Griffin, Murray, & Gonzalez, 1999). However, it is important to note that the use of difference scores have been criticized for several reasons including 1) unreliable scales produce unreliable difference scores that may obscure real effects, 2) highly-correlated variables produce unreliable difference scores, and 3) meaningless differences in variability can dramatically change results/conclusions (Furr, 2011; Griffin et al., 1999). Although the difference scores components in this study are reliable and only moderately correlated (see Table 4), this type of variable does not always produce the best model for the data and utilizing alternative statistical methods (e.g. multilevel modeling, growth modeling, partialing, multiple regression) is recommended (Furr, 2011; Griffin et al., 1999). As this study was exploratory, the most feasible variable and analysis was employed, however future studies should use improved statistical analyses.

Unexpectedly, race/ethnicity, gender, and length of time under supervision did not moderate the relationship between risk (initial or exit scores) and recidivism during a one-year follow up. In addition to evidence of disproportionate minority contact, race/ethnicity was expected to moderate the risk-recidivism relationship consistent with past research conducted with this population (e.g. Onifade et al., 2009) and with research conducted in other jurisdictions (e.g. Schwalbe, 2006). However, these negative findings could support past research that the YLS/CMI can equally predict recidivism across race/ethnicity (Jung & Rawana, 1999). As this moderator was tested with a less than ideal sample size, it is likely that the analysis lacked enough power for significant differences to emerge; therefore this finding should be interpreted with extreme caution. It is also possible that following youth for longer than one-year post-YLS/CMI could have produced different results, as the increased predictive validity of the YLS/CMI over time has been documented (Schmidt et al., 2011).

There was also an expectation that gender would moderate the relationship between risk and recidivism. However, a significant interaction did not emerge. Lack of significant findings could support the hypothesis that the YLS/CMI is gender-responsive and predicts recidivism equally for males and females (Flores et al., 2003; Jung & Rawana, 1999; Meyers & Schmidt, 2008; Olver et al., 2012; Olver et al., 2009; Onifade et al., 2010; Schmidt et al., 2011; Schwalbe, 2008). Similar to the explanation given above, lack of significant findings may be due to small sample size and a short follow-up period. Given this limitation, this finding should also be interpreted with extreme caution and future studies should extend the follow-up beyond one year and employ a larger sample of female offenders.

It was hypothesized that the length of time under supervision and its relationship with both exit risk scores and change in risk scores would impact the likelihood of re-offense. This

expectation grew from findings that suggested that depending on level of risk, spending too much time under court supervision could be iatrogenic, leading to negative outcomes (i.e. increased exposure to delinquent peers) (Lipsey, Howell, Kelly, Chapman, & Carver, 2010). Time under supervision was not found to moderate the exit risk-recidivism relationship. It may be that length of time is not as important as level of contact. Studies have investigated how offenders' frequency of contact with probation officers impact recidivism and have found more contact increases likelihood of reoffending (Bonta et al., 2008; Gatti et al., 2009). Future studies may identify whether length of time, and number of contacts interact leading to an increase in risk of recidivism. For instance, what is the effect on risk if an offender is under supervision for a short length of time, yet the juvenile court officer is heavily involved and makes frequent contact with the youth? It would also be worth examining whether length of time under supervision has differential affects by subgroups (e.g. high-risk youth, youth of color). While several explanations have been presented to describe this non-significant interaction, this finding should be interpreted with caution, as this variable's lack of variability may be the cause for lack of significant findings emerging. Specifically, 63% of the sample spent less than 12 months under supervision.

### *Limitations*

This study was not without its limitations. Use of archival data is a well-known limitation as one cannot guarantee there are not systematic errors in the way that data is collected (Vieira et al., 2009). In addition, employing official records of delinquency do not portray an accurate picture of how often an offender truly engages in delinquent acts. In addition, there are several ways to define recidivism (e.g. arrest, conviction, commitment, etc.), and this study employed official court contact. While other forms of recidivism may have led to different results, the use

of court petitions was viewed as the most reliable “middle-ground” outcome variable. In other words, the use of petitions can be viewed as a more conservative measure than arrest, and a more liberal measure than conviction. It is also important to note that research examining the use of different forms of recidivism found that they tend to strongly correlate with each other.

As previously mentioned, there were also issues with the use of difference scores to calculate change in risk scores from entry to upon exiting the juvenile justice system. To address this limitation, the author calculated bivariate correlations and both components were found to have adequate reliability. Next, this study could improve by using a larger sample size to increase statistical power and to better represent the population of the court in question. Although there were no systematic differences between the current sample and the overall juvenile population (with the exception of risk level), this sample only represented 20% of juvenile cases. However, this percentage is an improvement over studies with similar analyses (Flores et al., 2003). This study had to sacrifice the length of the follow up period to maximize the sample size as only offenders with entry and exit assessment matches were included in the study. Although past studies have cited a one-year follow up as an adequate amount of time for youth to reoffend (Onifade et al., 2008), Andrews and colleagues (2006) argue that shorter follow-up periods may reduce predictive validity.

#### *Future Directions and Implications*

Exploring the predictive utility of exit risk assessment scores adds to our knowledge of risk assessment, however more work is to be done. Future research on the relative predictive validity of initial and exit risk scores should be completed with follow-up periods longer than one year to increase predictive validity. As exit scores and change in scores did not yield significant findings, researchers should consider examining change in risk over more than two

points in time by way of reassessments. Not only can the use of reassessment risk scores permit the use of improved statistical methods, researchers believe their use could improve predictive validity (Andrews et al., 2006; Baglivio & Jackowski, 2012; Baglivio, 2009; Douglas & Skeem, 2005; Flores et al., 2003; Lowenkamp & Betchel, 2007; Olver et al., 2009; Onifade et al., 2011; Schmidt et al., 2011; Schmidt et al., 2005; Viljoen, Elkovitch, Scalora, & Ullman, 2009). Instead of limiting investigations solely to change in composite risk scores, future research should examine changes across YLS/CMI subscales as each subscale represents a criminogenic need. Exploring changes in criminogenic needs can allow researchers and practitioners to identify which areas are being most impacted by court supervision.

In order for future research to identify the impact of supervision, court practitioners must provide systematic information on the type of programming that youths receive, the amount of time youths spend in programs, and the number of programs youths are involved in at any given time. This future direction is consistent with the juvenile court “best practice” literature, which asserts that risk assessment should be used as a guideline for case management. This is especially important because research has shown that effectively targeting youths’ criminogenic needs during supervision leads to recidivism risk reduction (Bonta et al., 2008).

According to the current findings, the YLS/CMI appears to be gender and race/ethnicity-neutral. However, this finding should be interpreted with extreme caution. Researchers should continue examining the differential predictive validity of risk assessment across gender and race/ethnicity with larger sample sizes. Specifically because it is well documented that the juvenile justice system contains bias (Schwalbe et al., 2004), and that risk assessment instruments have been introduced to reduce this bias (Shepherd et al., 2013; Schwalbe et al., 2006) As a result, it is important to continue studying the impact that subgroup membership has

risk of recidivism. In addition to race/ethnicity and gender, future research should examine other variables (e.g. SES, family involvement, court officer characteristics) that could potentially moderate the risk-recidivism relationship.

Conducting validation studies that focus on assessing an offender's change in risk over time has implications for both research and practice. This work allows researchers to better understand the reliability of risk assessment instruments and how to improve them to maximize validity. The newest generations of risk assessment tools include dynamic risk factors whose very nature is to be examined over multiple points in time. Court practitioners allocate time and resources to purchasing and administering assessments, training court personnel, and using assessments to inform decision-making. As court practitioners place a high value on risk assessment implementation, exploring change in risk scores keeps them privy to whether their interventions and programs are effective at reducing risk. Additional research can provide evidence of the relative predictive utility of risk scores that may encourage court personnel to improve implementation policies assuring that every youth receive initial, process, and exit assessments. If the literature provides evidence that reassessment scores do not improve predictive validity, courts can save resources by focusing their attention on using the initial risk score to predict recidivism.

Finally, future research should examine the impact of court supervision as it relates to length of time, intensity of programming, and frequency of contact with juvenile court officers. This relationship should also be examined across juvenile subpopulations as these factors may have differential impacts. In closing, risk assessment research should go beyond the validation of initial risk scores and examine the validity and reliability of scores youth receive in the process of and upon dismissal from court supervision.

## **APPENDIX**

## Youth Level of Service/Case Management Inventory (YLS/CMI) Items

### **Prior/Current Offenses**

1. Three or More Prior Convictions
2. Two or more failures to comply
3. Prior Probation
4. Prior Custody
5. Three or More Current Convictions

### **Education**

6. Low Achievement
7. Problems with Teachers
8. Problems with Peers
9. Disruptive Classroom Behavior
10. Disruptive Behavior on School Property
11. Truancy

### **Leisure/Recreation**

12. Lack of Organized Activities
13. Could Make Better Use of Time
14. No Personal Interests

### **Peer Relations**

15. Lack of Positive Peer Acquaintances
16. Lack of Positive Friends
17. Some Delinquent Peer Acquaintances
18. Some Delinquent Friends

### **Substance Abuse**

14. Occasional Drug Use
15. Chronic Drug Use
16. Chronic Alcohol Use
17. Substance Abuse Interferes with Life
18. Substance Use Linked to Offense(s)

### **Family & Parenting**

19. Inadequate Supervision
20. Difficulty in Controlling Behavior
21. Inappropriate Discipline
22. Inconsistent Parenting
23. Poor Relations (Father-Youth)
24. Poor Relations (Mother-Youth)

### **Attitudes & Orientation:**

30. Not Seeking Help
31. Actively Rejecting Help
32. Defies Authority
33. Antisocial/Procriminal Attitudes
34. Callous, Little Concern for Others

### **Personality & Behavior**

35. Short Attention Span
36. Poor Frustration Tolerance
37. Verbally Aggressive/Verbally Intimidating
38. Explosive Episodes
39. Physically Aggressive
40. Inadequate Guilt Feelings
41. Inflated Self-Esteem
42. \* Unemployment/Not Looking for Work

\* Note: The variable Unemployment/Not looking for Work was omitted from the measure. This item was not relevant to this sample due to average age and had no variation.



## REFERENCES

## REFERENCES

- Andrews, D., Bonta, J., & Wormith, J. (2006). The recent past and near future of risk and/or need assessment. *Crime and Delinquency*, 52 (1), 7-27.
- Andrews, D., Bonta, J., Wormith, J., Guzzo, L., Brews, A., Rettinger, J., et al. (2011). Sources of variability in estimates of predictive validity: A specification with Level of Service general risk and need. *Criminal Justice and Behavior*, 48 (5), 413-432.
- Andrews, D., Guzzo, L., Raynor, P., Rowe, R., Rettinger, L., Brews, A., et al. (2012). Are the major risk/need factors predictive of both female and male reoffending? A test with the eight domains of the Level of Service/Case Management Inventory. *International Journal of Offender Therapy and Comparative Criminology*, 56 (1), 113-133.
- Baglivio, M. (2009). The assessment of risk to recidivate among a juvenile offending population. *Journal of Criminal Justice*, 37, 596-607.
- Baglivio, M., & Jackowski, K. (2012). Examining the validity of a juvenile offending risk assessment instrument across gender and race/ethnicity. *Youth Violence and Juvenile Justice*, 0(0), 1-18.
- Betchel, K., Lowenkamp, C., & Latessa, E. (2007). Assessing the risk of re-offending for juvenile offenders using the Youth Level of Service/Case Management Inventory. *Journal of Offender Rehabilitation*, 45 (3/4), 85-108.
- Bloom, B., Owen, B., Covington, S., & Raeder, M. (2002). Gender responsive strategies: Research, practice, and guiding principles for women offenders. National Institute of Corrections.
- Bonta, J., Rugge, T., Scott, T., Bourgon, G., & Yessine, A. (2008). Exploring the black box of community supervision. *Journal of Offender Rehabilitation*, 47(3), 248-270.
- Brennan, T., Dieterich, W., & Ehret, B. (2009). Evaluating the predictive validity of the compas risk and needs assessment system. *Criminal Justice and Behavior*, 36(1), 21-40.
- Bridges, G., & Steen, S. (1998). Racial disparities in official assessments of juvenile offenders: Attributional stereotypes as mediating mechanisms. *American Sociological Review*, 63(4), 554-570.
- Cabaniss, E., Frabutt, J., Kendrick, M., & Arbuckle, M. (2007). Reducing disproportionate minority contact in the juvenile justice system: Promising practices. *Aggression and Violent Behavior*, 12(4), 393-401.

- Catchpole, R., & Gretton, H. (2003). The predictive validity of risk assessment with violent young offenders: A 1- year examination of criminal outcome. *Criminal Justice and Behavior*, 30, 668-708.
- Chesney-Lind, M., Morash, M., & Stevens, T. (2008). Girls troubles, girl's delinquency, and gender responsive programming: A review. *Australian & New Zealand Journal of Criminology*, 41(1), 162-189.
- Cohen, M., Piquero, A., & Jennings, W. (2010). Estimating the costs of bad outcomes for at-risk youth and the benefits of early childhood interventions to reduce them. *Criminal Justice Policy Review*, 21(4), 391-434.
- Cottle, C. C., Lee, R. J., & Heilbrun, K. (2001). The prediction of criminal recidivism in juveniles: A meta-analysis. *Criminal Justice and Behavior*, 28, 367-294.
- Douglas, K., & Skeem, J. (2005). Violence risk assessment. *Psychology, Public Policy, and Law*, 11(3), 347-383.
- Elliot, D., & Ageton, S. (1980). Reconciling race and class differences in self-reported and official estimates of delinquency. *American Sociological Review*, 45(1), 95-110.
- Fass, T., Heilbrun, K., Dematteo, D., & Fretz, R. (2008). The LSI-R and the COMPAS: Validation data on two risk-needs tools. *Criminal Justice and Behavior*, 35(9), 1095-1108.
- Fitzgerald, R., & Carrington, P. (2011). Disproportionate minority contact in Canada: Police and visible minority youth. *Canadian Journal of Criminology and Criminal Justice*, 53(4), 449-486.
- Flores, A., Travis, L., & Latessa, E. (2003). Case classification for juvenile corrections: An assessment of the Youth Level of Service/Case Management Inventory (YLS/CMI). *Center for Criminal Justice Research, Division of Criminal Justice*, Cincinnati, OH.
- Freiburger, T., & Jordan, K. (2011). A multilevel analysis of race on decision to petition a case in the juvenile court. *Race and Justice*, 1(2), 185-201.
- Furr, R. M. (2011). *Scale construction and psychometrics for social and personality psychology*. Los Angeles, CA: Sage Publications.
- Gardner, R. (2001). *Psychological statistics using SPSS for windows*. New Jersey: Prentice Hall.
- Gatti, U., Tremblay, R., & Vitaro, F. (2009). Iatrogenic effect of juvenile justice. *The Journal of Child Psychology and Psychiatry*, 50(8), 991-998.
- Griffin, D., Murray, S., & Gonzalez, R. (1999). Difference score correlations in relationship research: A conceptual primer. *Personal Relationships*, 6(4), 505-518.

- Grove, W., & Meehl, P. (1996). Comparative efficiency of informal (subjective, impressionistic) and formal (mechanical, algorithmic) prediction procedures: The clinical-statistical controversy. *Psychology, Public Policy, and Law*, 2, 293-323.
- Hanley, J., & McNeil, B. (1983). A method of comparing the areas under receiving operating characteristic curves derived from the same cases. *Radiology*, 148(3), 839-843.
- Hindelang, M. (1973). Race and involvement in common law personal crimes. *American Sociological Review*, 43(1), 93-109.
- Holsinger, A., Lowenkamp, C., & Latessa, E. (2006). Exploring the validity of the Level of service Inventory-Revised with Native American offenders. *Journal of Criminal Justice*, 34, 331-337.
- Hsu, C., Caputi, P., & Byrne, M. (2009). The Level of Service Inventory-Revised (LSI-R) : A useful risk assessment measure for Australian offenders? *Criminal Justice and Behavior*, 36(7), 728-740.
- Jung, S., & Rawana, E. (1999). Risk and need assessment of juvenile offenders. *Criminal Justice and Behavior*, 26(1), 69-89.
- Kakar, S. (2006). Understanding the causes of disproportionate minority contact: Results of focus group discussions. *Journal of Criminal Justice*, 34, 369-381.
- Krysick, J., & LeCroy, C. (2002). The empiracal validation of an instrument to predict risk of recidivism among juvenile offenders. *Research on Social Work Practice*, 12(1), 71-81.
- Lipsey, M., Howell, J., Kelly, M., Chapman, G., & Carver, D. (2010). *Improving the effectiveness of juvenile justice programs: A new perspective on evidence-based practice*. Georgetown University. Washington DC: Center for Juvenile Justice Reform.
- Lowenkamp, C., & Betchel, K. (2007). The predictive validity of the LSI-R on a sample of offenders drawn from the records of the Iowa department of corrections data management system. *Federal Probation*, 71(3), 25-29.
- Luong, D., & Wormith, J. (2011). Applying risk/need assessment to probation practice and its impact on the recidivism of young offenders. *Criminal Justice and Behavior*, 38(12), 1177-1199.
- McGhee, H., & White, G. (2010). *A way out: Creating partners for our nation's prosperity by expanding life paths of young men of color*. Washington, DC: Joint Center for Political and Economic Studies.
- McGrath, A., & Thompson, A. (2012). The relative predictive validity of the static and dynamic domain scores in risk-need assessment of juvenile offenders. *Criminal Justice and Behavior*, 39(3), 250-263.

- Mears, D., Cochran, J., Greenman, S., Bhati, A., & Greenwald, M. (2011). Evidence on the effectiveness of juvenile court sanctions. *Journal of Criminal Justice*, 39(6), 509-520.
- Meyers, J., & Schmidt, F. (2008). Predictive validity of the Structured Assessment for Violence Risk in Youth (SAVRY) with juvenile offenders. *Criminal Justice and Behavior*, 35(3), 344-355.
- Olver, M., Stockdale, K., & Wormith, S. (2009). Risk assessment with young offenders: A meta-analysis of three assessment measures. *Criminal Justice and Behavior*, 36(4), 329-353.
- Olver, M., Stockdale, K., & Wong, S. (2011). Short and long-term prediction of recidivism using the Youth Level of Service/Case Management Inventory in a sample of serious young offenders. *Law and Human Behavior*, 1(1), 1-15.
- Onifade, E., Davidson, W., & Campbell, C. (2009). Risk assessment: The predictive validity of the Youth Level of Service/Case Management Inventory with African Americans and girls. *Journal of Ethnicity in Criminal Justice*, 7(3), 205-221.
- Onifade, E., Davidson, W., Campbell, C., Turke, G., Malinowski, J., & Turner, K. (2008). Predicting recidivism in probationers with the Youth Level of Service Case Management Inventory (YLS/CMI). *Criminal Justice and Behavior*, 35(3), 474-483.
- Onifade, E., Davidson, W., Livsey, S., Turke, G., Horton, C., Malinowski, J., et al. (2008). Risk assessment: Identifying patterns of risk in young offenders with the Youth level of Service/Case Management Inventory. *Journal of Criminal Justice*, 36, 165-173.
- Onifade, E., Wilkins, J., Davidson, W., Campbell, C., & Peterson, J. (2011). A comparative analysis of recidivism with propensity score matching of informal and formal juvenile probationers. *Journal of Offender Rehabilitation*, 50(8), 531-546.
- Piquero, A. (2008). Disproportionate minority contact. *Juvenile Justice*, 8(2), 59-79.
- Piquero, A., & Brame, R. (2008). Assessing the race-crime and ethnicity-crime relationship in a sample of serious adolescent delinquents. *Crime and Delinquency*, 54, 1-33.
- Puzzanchera, C., & Adams, B. (2011). *Juvenile arrests 2009*. Washington, DC: Office of Juvenile Justice and Delinquency Prevention.
- Puzzanchera, C., Adams, B., & Hockenberry, S. (2012). *Juvenile Court Statistics 2009*. Pittsburgh, PA: National Center for Juvenile Justice.
- Puzzanchera, C., Adams, B., & Sickmund, M. (2011). *Juvenile court statistics 2008*. Pittsburgh, PA: National Center for Juvenile Justice.

- Rice, M., & Harris, G. (1995). Violent recidivism: Assessing predictive validity. *Journal of Consulting and Clinical Psychology, 63*(5), 737-748
- Schlager, M., & Pacheco, D. (2011). An examination of changes in LSI-R scores over time: Making the case for needs-based case management. *Criminal Justice and Behavior, 38*(6), 541-553.
- Schlager, M., & Simourd, D. (2007). Validity of the Level of Service Inventory-Revised (LSI-R) among African American and Hispanic male offenders. *Criminal Justice and Behavior, 34*(4), 345-354.
- Schmidt, F., Campbell, M., & 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*(1), 23-42.
- Schmidt, F., Hoge, R. D., & Gomes, L. (2005). Reliability and validity analyses of the youth Level of Service/Case Management Inventory. *Criminal Justice and Behavior, 32*(3), 329-344.
- Schmidt, F., McKinnon, L., Chattha, H., & Brownlee, K. (2006). Concurrent and predictive validity of the Psychopathy Checklist: Youth Version across gender and ethnicity. *Psychological Assessment, 18*(4), 393-401.
- Schwalbe, C. (2004). Re-visioning risk assessment for human service decision making. *Children and Youth Services Review, 26*, 561-576.
- Schwalbe, C. (2007). Risk assessment for juvenile justice: A meta-analysis. *Law and Human Behavior, 31*, 449-462.
- Schwalbe, C. (2008a). A meta-analysis of juvenile justice risk assessment instruments: Predictive validity by gender. *Criminal Justice and Behavior, 35*(11), 1367-1381.
- Schwalbe, C. (2008b). Strengthening the intergration of actuarial risk assessment with clinical judgment in evidence based practice framework. *Children and Youth Services Review, 30*, 1458-1464.
- Schwalbe, C. (2009). Risk assessment stability: A revalidation study of the Arizona Risk/Needs Assessment instrument. *Research on Social Work Practice, 20*, 205-213.
- Schwalbe, C., Fraser, M., & Day, S. (2007). Predictive validity of the Joint Risk Matrix with juvenile offenders: A focus on gender and race/ethnicity. *Criminal Justice and Behavior, 34*(3), 348-361.
- Schwalbe, C., Fraser, M., Day, S., & Arnold, E. (2004). North Carolina Assessment of Risk (NCAR): Reliability and predictive validity with juvenile offenders. *Journal of Offender Rehabilitation, 40*(1/2), 1-22.

- Schwalbe, C., Fraser, M., Day, S., & Cooley, V. (2006). Classifying juvenile offenders according to risk of recidivism: Predictive validity, race/ethnicity, and gender. *Criminal Justice and Behavior*, 33, 305-324.
- Shaffer, D., Kelly, B., & Lieberman, J. (2011). An exemplar-based approach to risk assessment: Validating the Risk Management Systems instrument. *Criminal Justice Policy Review*, 22(2), 167-186.
- Shepherd, S., Luebbers, S., & Dolan, M. (2013). Gender and ethnicity in juvenile risk assessment. *Criminal Justice and Behavior*, 40(4), 388-408.
- Simpson, S., Yahner, J., & Dugan, L. (2008). Understanding women's pathways to jail: Analysing the lives of incarcerated women. *Australian & New Zealand Journal of Criminology*, 41(1), 84-108.
- Stevens, T., Morash, M., & Chesney-Lind, M. (2011). Are girls getting tougher, or are we tougher on girls? Probability of arrest and juvenile court oversight in 1980 and 2000. *Justice Quarterly*, 28(5), 719-744.
- Tyda, K. (2011). *Screenings and assessments used in the juvenile justice system: Evaluating risks and needs of youth in the juvenile justice system*. San Francisco, CA: Judicial Council of California.
- Welsh, J., Schmidt, F., McKinnon, L., Chattha, H., & Meyers, J. (2008). A comparative study of adolescent risk assessment instruments: Predictive and incremental validity. *Assessment*, 15(1), 104-115.
- Werling, R., Cardner, P., & University-Austin, P. (2011). Disproportionate minority/police contact: A social service perspective. *Applied Psychology in Criminal Justice*, 7(1), 47-58.
- Wordes, M., Bynum, T., & Corley, C. (1994). Locking up youth: The impact of race on detention decisions. *Journal of Research in Crime and Delinquency*, 31(2), 149-165.
- Worling, J., & Langstrom, N. (2003). Assessment of criminal recidivism risk with adolescents who have offended sexually: A review. *Trauma, Violence, and Abuse*, 4(4), 341-362.
- Vieira, T., Skilling, T., & Peterson-Badali, M. (2009). Matching court-ordered services with treatment needs: Predicting treatment success with young offenders. *Criminal Justice and Behavior*, 36(4), 385-401.
- Viljoen, J., Elkovitch, N., Scalora, M., & Ullman, D. (2009). Assessment of reoffense risk in adolescents who have committed sexual offenses: Predictive validity of the ERASOR, PCL:YV, YLS/CMI, and Static-99. *Criminal Justice and Behavior*, 36(10), 981-1000.

- Vincent, G., Chapman, J., & Cook, N. (2011). Risk-needs assessment in juvenile justice: Predictive validity of the SAVRY, racial differences, and the contribution of needs factors. *Criminal Justice and Behavior*, 38(1), 42-62.
- Vincent, G., Paiva-Salisbury, C. N., Guy, L., & Perrault, R. (2012). Impact of risk/needs assessments on juvenile probation officers' decision making: Importance of implementation. *Psychology, Public Policy, & Law*, 18(4), 549-576 .
- Vitopolous, N., Peterson-Badali, M., & Skilling, T. (2012). The relationship between matching service to criminogenic need and recidivism in male and female youth: Examining the RNR principles in practice. *Criminal Justice and Behavior*, 39(8), 1025-1041.
- Voorhis, P., Wright, E., Salisbury, E., & Bauman, A. (2010). Women's risk factors and their contributions to existing risk/needs assessment: The current status of a gender-responsive supplement. *Criminal Justice and Behavior*, 37(3), 261-288.
- Vose, B., Lowenkamp, C., Smith, P., & Cullen, F. (2009). Gender and the predictive validity of the LSI-R: A study of parolees and probationers. *Journal of Contemporary Criminal Justice*, 25(4), 459-471.