PILOTING THE MONTANA JUVENILE PROBATION SCREENER

June 2020

Patrick McKay
Dusten Hollist
Jackson Bunch
Jessica Mayrer
Jason Liechty
Chuck Harris

The Criminology Research Group
Social Science Research Laboratory
University of Montana, Missoula

ACKNOWLEDGEMENTS

The researchers would like to recognize and thank all who enabled the work that this report is built upon. This report was provided to the Office of the Court Administration by the University of Montana's Criminology Research Group (CRG). The points of views, opinions, findings, and conclusions in this document are those of the authors and do not necessarily represent the policies or positions of any state or federal agency.

A debt of gratitude is extended to Montana Supreme Court Office of the Court Administrator for funding this project. CRG extends a special thanks to Tom Billteen and Kelly Elder from the Office of the Court Administrator and to all juvenile probation departments that piloted the Montana Juvenile Probation Screener (MJPS) and responded to the associated Juvenile Probation Survey. Support from the University of Montana, in particular the Office of Research and Sponsored Programs, the College of Arts and Sciences, and the Department of Sociology is recognized and appreciated. Finally, CRG recognizes former Criminology Research Group Director, Professor Dusten Hollist, who passed away as research for this analysis was ongoing. Professor Hollist created the CRG and devoted much of his life and career to increasing evidence-based data-driven practices in Montana. All members of his team strive to continue his legacy knowing that Montana is a better place because of him.

The Criminology Research Group

Social Science Research Laboratory

The University of Montana, Missoula

Social Science Building, Room 259

Missoula, Montana 59812

(406) 243-5381 (Office)

(406) 243-5951 (Fax)

CONTENTS

LIST OF TABLES

LIST OF FIGURES

EXECUTIVE SUMMARY

In July 2018, the Criminology Research Group (CRG) through the Social Science Research Lab (SSRL) at the University of Montana partnered with Montana Youth Court Services through the Office of the Court Administrator to pilot the Montana Juvenile Probation Screener (MJPS). The pilot test was employed to evaluate the predictive accuracy of the MJPS in a prospective study design. The following report is divided into three sections. The first section describes the process of piloting the MJPS and discusses initial data analyses, including demographics of screened youth, risk levels, dispositions associated with MJPS risk levels, and overriding the indicated decision. The second section discusses results from a survey distributed to probation officers across Montana about their experience using currently employed risk and needs assessments. Pilot sites were asked additional questions about their experience with the MJPS. The final section discusses to what extent the MJPS predicts recidivism and how these results compared to predictions from the current risk assessment used in the field (Pre-screen BOT).

MONTANA JUVENILE PROBATION SCREENER PILOT TEST

Early in 2018, six Montana Youth Court Services districts agreed to pilot the MJPS. The sites included Lewis and Clark (District 1), Beaverhead (District 5), Flathead (District 11), Yellowstone (District 13), Lincoln (District 19), and Ravalli (District 21) counties. The pilot began on July 2, 2018 and concluded on Jan. 31, 2019.

During the pilot test, 1,408 youth were cited with an offense at the six pilot sites. From this total, 34% or 444 youth were administered the MJPS. Administration rates between sites ranged from a high of 82% to the low of 23%. Youth cited for felony offenses were unlikely to be screened. Only 9% of youth cited for a felony were administered the MJPS, compared to 35% of youth cited for misdemeanors, and 34% of youth cited for status, technical, or city ordinance violations. Finally, non-white youth were less likely to be screened. Only 20% of non-white youth were administered the MJPS, while 35% of white youth were given the MJPS. American Indian youth had the lowest administration rate, at 17%.

Results presented in the first section of this report explore risk factor frequency, as chronicled by the MJPS, in addition to risk score distributions and face and concurrent validity. Additionally, we analyze youth disposition compared to assessment-designated risk levels, common errors found during MJPS administration, and probation officer justifications for desired overrides of MJPS scores.

Results. Risk factor frequencies were examined to ensure that no risk factor was occurring too often in the population of screened youth or any subsamples. The most common risk factor was *current anti-social friends* with over half (51.7%) of youth having this risk factor. The least common risk factor was *one or more felony referrals,* with only 5.2% of youth classified as having that risk factor. The remaining six risk factors occurred between 11.7% to 35.1% of the time. Risk factor frequencies were similar between sexes. Minor variations discovered between male and female risk factors were expected and do not present a concern. Risk factor frequencies among races were similar overall, but two risk factors stood out as being markedly different. First, 50% of white youth were deemed to have *current anti-social friends,* compared to almost 70% of non-white youth. Second, 21% of white youth were reported to have a *belief in fighting,* compared to 62% of non-white youth.

Most screened youth were indicated to have low recidivism risk (68%), with significantly fewer categorized as moderate risk (25%), and even fewer high risk (6%). Non-white youth were more likely than white youth to be deemed moderate risk (31.8% compared to 24.8%) and significantly more likely to be classified as high-risk (13.6% to 4.8%). Risk scores between males and females were comparable.

Face validity relates to whether officers administering the MJPS believe the tool is correctly predicting risk level. Out of the 444 youth administered the MJPS, officers agreed with 92% (400) of the predicted risk levels. This provides significant evidence of face validity. Alternatively, officers only agreed with the Pre-screen BOT 83% of the time. An examination of concurrent validity entailed an investigation into to what extent the MJPS predicts risk compared to a previously validated risk assessment (Pre-screen BOT). The MJPS and the Pre-screen BOT indicated the same risk level 80% of the time with a moderate to strong positive correlation ($r = .627$).

Risk levels, as indicated by the MJPS, were also highly predictive of disposition outcome for youth. Most low-risk youth (80%) were given the least restrictive disposition with few (15%) given informal probation, no low-risk youth were given the most restrictive disposition of formal probation. Moderate-risk young people were diverted less frequently (65%), 29% were placed on informal supervision and 1% were put on formal probation. Similarly, high-risk youth were the least likely to receive a diversion disposition and much more likely to be given informal probation and formal probation, relative to low and moderate-risk youth.

Few (10%) administrator errors were found on completed MJPS instruments. The most common error was the omission of information about practitioner judgement related to a youth's recidivism risk. The second most common error was the omission of youth disposition. Other errors included risk factors left blank, incorrect Juvenile Court and Assessment Tracking System (JCATS) numbers, lacking override justifications, miscalculated risk scores, and incorrect rater IDs. Overall, however, errors were uncommon, especially for a tool administered completely by hand.

Overrides occurred when officers administering the MJPS did not agree with the MJPS-indicated risk level. This occurred roughly 8% of the time. When an override was made, officers were prompted to explain why they believed risk was higher or lower than the MJPS indicated. For instances in which officers believed youth posed less risk than the MJPS indicated, they attributed their thinking to mitigating and protective factors not chronicled by the MJPS. Similarly, when officers believed a youth posed a higher risk than designated by the MJPS, they attributed that thinking to aggravating factors and risk factors not included on the MJPS.

JUVENILE PROBATION SURVEY

The Montana Juvenile Probation Survey invited Youth Court Services officers administering risk assessments to voice their professional opinions about instruments currently in use. Inclusion of the probation officer perspective constitutes an important step to increase buy-in for utilization of any risk assessment instrument. In total, 80 probation officers completed the Montana Juvenile Probation Survey. Of those respondents, 20 individuals from the six participating MJPS pilot districts completed the survey. This section details Juvenile Probation Survey results.

Results. Officers were asked to share their ideas about strategies for improving the current risk assessment used in the field, the Pre-screen BOT. The following thoughts and ideas were reported:

- Increase training
- Increase research
- Believe Pre-screen is overly invasive
- Believe Pre-screen is too long
- Does not collect certain important information

Officers were asked how they collected answers to MJPS questions. Youth interviews constituted the most common way MJPS information was collected, with all respondents reporting this mode of data collection. Additionally, most (90%) officers reported interviewing parents or guardians. Roughly half (45%) collected data from JCATS. Finally, few (10%) officers reported speaking with school officials or teachers and 10% reporting using "law enforcement notes" to collect MJPS risk factor information.

Officers responding to the survey indicated greater confidence in the MJPS than the Pre-screen BOT. Of those who used the MJPS, 90% reported to having *a lot* or *a moderate amount* of confidence in the screener's ability to predict risk. Seventy four percent of responding officers reported the same amount of confidence in the Pre-screen BOT. Moreover, 74% of those who used the MJPS and the Pre-screen BOT reported the MJPS was an improvement over the Pre-screen.

In an open-ended question format, officers were asked what MJPS questions they had trouble collecting. *Positive adult role model, history of running away, belief in fighting,* and *problem-solving* risk factors were flagged as the most challenging to collect.

PREDICTIVE VALIDITY

On Feb. 10, 2020, one year after the Montana Juvenile Probation Screener (MJPS) pilot test concluded, the Criminology Research Group (CRG) received from the Montana Office of the Court Administrator (OCA) Youth Court Services, recidivism data on youth administered the MJPS. This final section examines MJPS predictive validity. Predictive validity refers to a risk assessment's ability to accurately predict an outcome of interest. The MJPS was created to predict recidivism risk. Several measurements of predictive validity are presented here and when possible, MJPS results are compared to those from the Pre-screen BOT. Additionally, CRG examined how the MJPS and professional discretion can work together to maximize prediction accuracy.

Youth administered the MJPS were, in general, less likely to have a recidivating offense compared to the entire population of youth going through intake at pilot sites. Of youth administered the MJPS, only 22% had *any recidivating* offense compared to 32% of the general population.

Results. Risk scores were first investigated. Each point increase in the MJPS risk score should indicate an actual increase in recidivism risk. Thus, each point increase should be associated with an increase in the recidivism rate. Overall, this pattern was discovered for all three recidivism outcomes. Low MJPS risk scores present lower levels of recidivism compared to high risk scores. Below, misdemeanor/felony recidivism rates are presented by MJPS risk level (i.e., low, moderate, high).

- Low: 14%
- Moderate: 26%
- High: 33%

When officers agreed with the MJPS-indicated risk level, recidivism rates were improved. Misdemeanor/felony recidivism rates are presented below for only those cases where officers agreed with the MJPS risk level.

- Low: 14%
- Moderate: 30%
- High: 44%

When officers believed the MJPS risk level was too high, recidivism rates were lower than their associated risk level as designated by the MJPS. Similarly, when officers believed the MJPS risk level was too low, recidivism rates were higher than their associated risk level as designated by the MJPS. Taken together, evidence presented here demonstrates an increase in prediction performance when professional judgement is included.

Logistic regression was utilized to examine the effect of the total MJPS risk score (0 to 7) on predicting recidivism, while holding gender and race constant. Results from three logistic regression models indicate the MJPS risk score is a statistically significant predictor for each type of recidivating outcome (*any*, *misdemeanor/felony*, and *felony*), while holding race and gender constant. For each MJPS risk score point increase, there was a 43% increase in the likelihood of *any recidivism*, a 40% increase in the likelihood of *misdemeanor/felony recidivism*, and an 89% increase in the likelihood of *felony recidivism*. Also according to the models, being male increased the likelihood of *any recidivism* and *misdemeanor/felony recidivism,* being white decreased the likelihood of *any recidivism* while holding all else constant.

Finally, receiver operating characteristic (ROC) analysis was employed to analyze MJPS accuracy. Full sample results show the MJPS increased prediction accuracy compared to the Pre-screen BOT for all recidivism categories. Area under the curve (AUC) statistics derived from ROC analysis of the MJPS ranged from a high of .76 (large effect) for prediction of *felony recidivism* to a low of .65 (moderate effect) for the prediction of *misdemeanor/felony recidivism*. Pilot site results match closely to data collected between 2010 and 2015 for MJPS creation. In the prospective analysis, there was minimal "shrinkage," a term used to compare results from data used to create a tool to results collected live during the pilot test. A smaller amount of shrinkage is evidence that the results are consistent from one sample to the next and from one time period to the next. The MJPS also demonstrated superior performance to the Pre-screen when the sample was broken down into gender and race-specific categories. When prediction accuracy was measured for the American Indian sample, however, results indicated prediction performance no better than chance for both the MJPS and Pre-screen. The non-random selection of American Indian youth during the pilot test and a small sample size may have affected the prediction performance. Continued analysis of to what extent risk assessments used by Youth Court Services accurately predict recidivism among young Native Americans is recommended.

CONCLUSION

This report details the first prospective validation study of the Montana Juvenile Probation Screener (MJPS). The MJPS was designed to serve as a brief and accurate risk assessment that would increase fidelity compared to more time-consuming instruments. Probation officers offered their recommendations to improve the Pre-screen BOT, and the MJPS already addresses several shortcomings. For example, the MJPS is significantly shorter, with risk factors that are not overly

intrusive. Brevity allows for easier training and understanding of the instrument, which, in turn, will increase inter-rater reliability. The MJPS was built upon data collected in Montana and has established itself as an accurate risk predictor during a retrospective test with comparable results during this prospective analysis. The Criminology Research Group believes these improvements and others will increase instrument fidelity.

Analyses of data collected during the MJPS pilot test demonstrate promising results. MJPS face validity was high. Juvenile probation officers agreed with the predicted risk level 92% of the time and reported—via survey—confidence in the screener's ability to predict recidivism. When compared to the Pre-screen BOT, officers indicated a higher degree of confidence in the MJPS, agreeing with the MJPS-indicated risk level to a greater extent than they did with the Pre-screen. During the pilot test, the MJPS successfully located low-risk youth, most of whom were given the least restrictive disposition. Such identification and disposition application satisfies a primary goal associated with MJPS introduction. Further, prediction validation analyses demonstrated the MJPS comparable and, at times more accurate than the Pre-screen BOT when predicting recidivism. Overall findings from this study support MJPS validity and its use as a brief tool to determine recidivism risk.

INTRODUCTION

In July 2018, the Criminology Research Group (CRG) through the Social Science Research Lab (SSRL) at the University of Montana partnered with Montana Youth Court Services through the Office of the Court Administrator to pilot the Montana Juvenile Probation Screener (MJPS). The pilot test was employed to evaluate the predictive accuracy of the MJPS in a prospective study design. The investigation was built upon five years of research examining the process by which recidivism risk is measured during Montana Youth Court Services intake.

Background

Montana juvenile probation currently utilizes two assessments: (1) the Back On Track (BOT) Pre-screen; and (2) the BOT Full-screen. The BOT Pre-screen is approximately 40 questions designed to predict recidivism risk. The BOT Full-screen, alternatively, is a significantly longer needs assessment. The Full-screen is intended to be administered to youth who score moderate to high on the Pre-screen BOT to assist in service decisions for those youth found to be in most need.

In May 2015, the CRG released a report titled "The Montana Back on Track Risk Assessment Instrument: an Assessment and Validation Study" (see McKay, Hollist, Bunch, Acton, Tillman, and Harris, 2015). The report detailed findings from an analysis of the Montana Back on Track Pre-screen related to how well it predicted recidivism among a variety of subsamples. Subsamples analyzed included gender, race, and levels of offense severity. The analysis showed the Pre-screen BOT performed comparably to other risk assessment instruments in prediction accuracy for first-time offending youth. Further, the inquiry explored possible reasons certain youth were administered the Pre-screen BOT while others were not. Different administration rates were found based on race, age, offense type, and, most notably, the district in which the offense occurred. Findings from that inquiry indicated the Pre-screen and Full-BOT were not being utilized with fidelity and consistency.

To address the problem of low or uneven fidelity identified in the 2015 report, CRG set out to determine if a brief instrument could be created from risk factors found on the Full-screen BOT that maintained, if not increased, prediction accuracy and could replace the current Pre-screen BOT. Findings from that inquiry were detailed in a report titled "Montana Back on Track (BOT): Developing a Risk Assessment Screener and Utilizing the BOT to Assist in Service Placement." (See McKay, Hollist, and Ranger, 2017). As discussed in that report, a shorter screening instrument stands to increase fidelity to screening and encourage evidence-based decision making among probation officers. Of the 246 risk factors on the Full-screen BOT, eight were found to be the most important for determining youth recidivism risk. These eight risk factors were compiled together to create the new brief screener titled the Montana Juvenile Probation Screener (MJPS). The screener demonstrated comparable, and at times improved, prediction accuracy over the Pre-screen BOT using retrospective data. While the retrospective analysis provided significant evidence to support MJPS accuracy, a prospective MJPS pilot test was recommended to determine MJPS capacity to replace the Pre-screen BOT as a risk assessment for Montana Youth Court Services.

Current Study

This report details findings from a prospective pilot test of the MJPS. Findings are detailed in three sections: (1) MJPS Pilot Test; (2) The Juvenile Probation Survey; and (3) Predictive Validity. The

first section describes the MJPS administration process and demographics of youth screened in six pilot sites during a seven-month period. This section also examines MJPS risk factors and the frequency with which they occur among the screened sample, overall risk scores, face and concurrent validity, officer overrides, and MJPS administration errors. The second section presents results from the Juvenile Probation Survey. With the survey, the CRG investigated officer perceptions of the Pre-screen BOT, the Full-screen BOT, and the MJPS. Of particular interest in this section is the confidence levels of officers using the risk assessments and officer recommended strategies for improving recidivism risk prediction. The third section of this report examines MJPS accuracy predicting recidivism and how it compares to the Pre-screen BOT. This section details various analyses to gauge predictive validity including logistic regression analysis and receiver operating characteristic (ROC) analysis. The final section of this report examines limitations associated with this inquiry and future research recommendations.

MONTANA JUVENILE PROBATION SCREENER PILOT TEST

To begin, this section outlines the process of piloting the Montana Juvenile Probation Screener (MJPS) and the sample of youth screened. Data collected on the screened sample is examined including risk factors frequencies and risk score distributions. Next, evidence for face and concurrent validity are presented. These findings are based on how probation officers' professional judgement, youth dispositions, and Pre-screen BOT risk scores align with MJPS prediction results. Finally, common override justifications are examined along with a discussion of errors that occurred while administering the MJPS.

Methods

Early in 2018, six Youth Court Services districts agreed to pilot the MJPS. The sites included Lewis and Clark (District 1), Beaverhead (District 5), Flathead (District 11), Yellowstone (District 13), Lincoln (District 19), and Ravalli (District 21) counties. Before the pilot began, all sites were provided with the Montana Back on Track Report (McKay, Hollist, and Ranger, 2017) detailing information on MJPS creation; guidelines on how the screening and data collection process should be conducted along with pilot site expectations; a risk factor definition and interview guide explaining how risk factors should be chronicled in a standardized manner; and a training vignette demonstrating the MJPS administration process. All of the above-itemized materials are included in the appendix of this report (Appendix A-E).[1] Sites were asked to review the materials and to familiarize themselves with the MJPS before its launch.

On July 2, 2018, pilot sites began administering the MJPS during the youth intake process. Completed screeners were scanned by administrators and emailed to the University of Montana Criminology Research Group (CRG). The MJPS pilot test concluded on Jan. 31, 2019, when all pilot district stopped administering the MJPS.

Data. Two data sources were used for this analysis: data from the completed MJPS and data retrieved from the Juvenile Court Assessment and Tracking System (JCATS), which is Montana's juvenile justice data repository.

All data collected from MJPS administration were compiled into an SPSS database. Fifteen variables were collected from each MJPS screener (an MJPS is included in Appendix D). A JCATS identification number unique to each individual youth was used along with the intake date to locate and link youth with additional information from the JCATS system. Two separate datasets were queried from JCATS after the pilot test. CRG first obtained data on all intakes that occurred in the respective pilot districts. Demographics, current offenses, and prior offenses were included in the first dataset. These data were used to determine the rate at which youth were screened and examine how individuals were selected to be screened. Next, a dataset containing all Pre-screen Back on Track (BOT) risk scores was queried for each pilot site. These data were used to compare the

[1] The Montana Back on Track (2017) report is not located in the appendix but can be found at the following link: https://courts.mt.gov/portals/189/dcourt/yth_court/docs/UMreport2017.pdf

MJPS risk scores to the Pre-screen BOT risk scores. The three data files were merged into one file matching JCATS numbers and intake dates.[2]

Youth Selected for Screening. During the MJPS pilot test, 1,408 youth were cited in the six pilot districts. From that total, 444 youth (32%) were administered the MJPS. Pilot districts utilized the screener at different rates. One district completed the MJPS with 82% of intakes, another with 63% of intakes. The district with the lowest administration rate completed the MJPS on 18% of intakes, the second lowest administration rate was 23%.

Youth cited for a felony were unlikely to receive the MJPS regardless of which district the offense occurred in. During the pilot test, a total of 142 youth were cited with a felony in the six districts. Only 13 of those young people—or 9% of all felony citations—were administered the MJPS. Further, 35% of all individuals cited for a misdemeanor were administered the screener; 34% of all youth cited for a status, technical, or city ordinance were screened.

White youth were more likely to be screened; 35% of white youth received the screener, while 20% of non-white youth were screened. American Indians had the lowest administration rate of all race categories. During the pilot test, 113 American Indian youth were cited for an offense and 19 (17%) received the screener. The non-random selection of youth administered the MJPS may have ramifications when attempting to generalize these findings to the larger population. The implications are most significant for non-white youth and juvenile felony offenders.

Analytic Strategy. The results section below explores three areas of analysis: risk factors, risk score, and validity. To begin, risk factor frequencies were examined. Ideally, risk factors should not be occurring with a very high or a very low frequency and should not differ drastically between different subsamples (e.g., race and gender). This analysis examined distribution among final risk scores and how risk-score distributions differed between race and gender. Finally, two validity types were examined: face validity and concurrent validity. In this instance, face validity measured how well the screener appeared to be working based on professional opinions from probation officers in the field. Concurrent validity was evaluated by comparing MJPS risk prediction compared to predictions from the Pre-screen BOT, which has already been validated as an accurate recidivism predictor.

Results

Table 1.1 presents demographics from youth receiving the MJPS. More of the 444 youth screened were male, representing 63.6% (272) of the sample. White individuals comprised 89.5% (383) of the sample and American Indians represented 4.4% (19). The remaining youth identified as Native Hawaiian or Pacific Islander (.7%; 3), African American (1.9%; 8), and Hispanic or Latino (2.5%; 15).

[2]Intake date, MJPS administration date, and Pre-screen BOT administration date were often not identical. In most cases, these dates were within days from each other. Pre-screen BOT data was not included if the date was 30 or more days away from the MJPS date. Of the total youth administered the MJPS, 172 were unable to be matched with a Pre-screen BOT and 16 youth administered the MJPS did not have JCATS numbers that matched any youth in the JCATS.

Table 1.1: Demographics of Screened Youth (*N*=444)

Demographic		%
Gender	Female	36.4%
	Male	63.6%
Race/Ethnicity	White	89.5%
	American Indian	4.4%
	Native Hawaiian or Pacific Islander	0.7%
	African American	1.9%
	Hispanic/Latino	3.5%
Offense	Status/Technical	23.2%
	Misdemeanor	73.8%
	Felony	2.9%
District	1	21.2%
	5	2.3%
	11	20.9%
	13	27.7%
	19	10.6%
	21	17.3%
Age	11 and Under	4.7%
	12 to 14	25.8%
	15 to 17	69.1%
	18+	0.5%

Most youth were cited with a misdemeanor (73.8%; 315), followed by status/technical citations (23.2%; 99), and then felony citations (2.9%; 13). Each district piloting the MJPS varies in population size and therefore the number of youth served. Among youth receiving the MJPS, most of the sample fell between the ages of 15 to 17 (69.1%; 297) and 12 to 14 years old (25.8%; 111). Few youth were 11 years old or younger (4.7%; 94) and fewer were 18 or older (.5%; 2).

Risk Factors. Figure 1.1 displays the percent of youth with risk factors chronicled by the MJPS. The most common risk factor is *current anti-social friends*. More than 50% of all youth in the sample were deemed to have anti-social friends. The next most common risk factor was *does not problem solve* with 35%. Approximately one quarter of the youth were deemed to *believe in fighting*, had *more than one misdemeanor referral* in their past, and a *history of running away*. Fewer youth had *no positive adult role model* (16.8%), a *first offense under 13* (11.7%), or *one or more felony referrals* (5.2%).

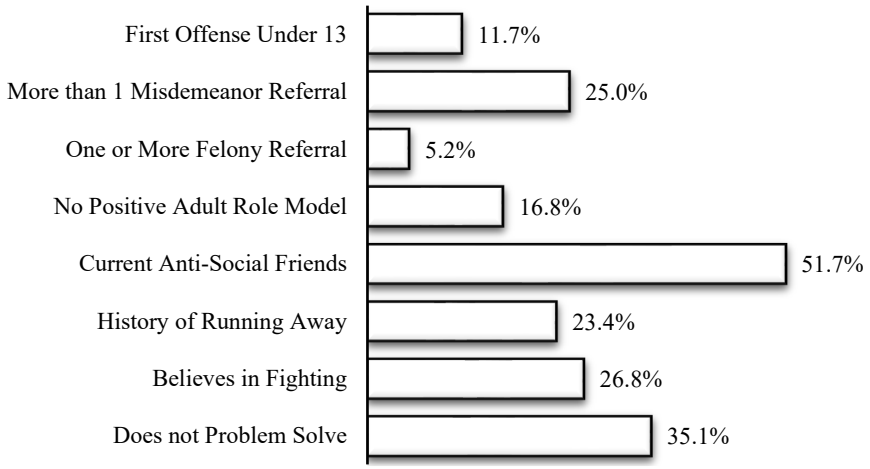


Figure 1.1: Risk Factor Frequency: Total Sample (*N*=444*)*

Figure 1.2 examines the different frequencies of risk factors between gender. Overall, males and females had a similar frequency of risk factors. Males were more likely than females to have a *first offense under the age of 13*, *more than one prior misdemeanor*, *one or more felony referrals*, *no positive adult role model*, and be *unable to problem solve*. Females were slightly more likely to have *anti-social friends*, a *history of running away*, and a *belief in fighting*.

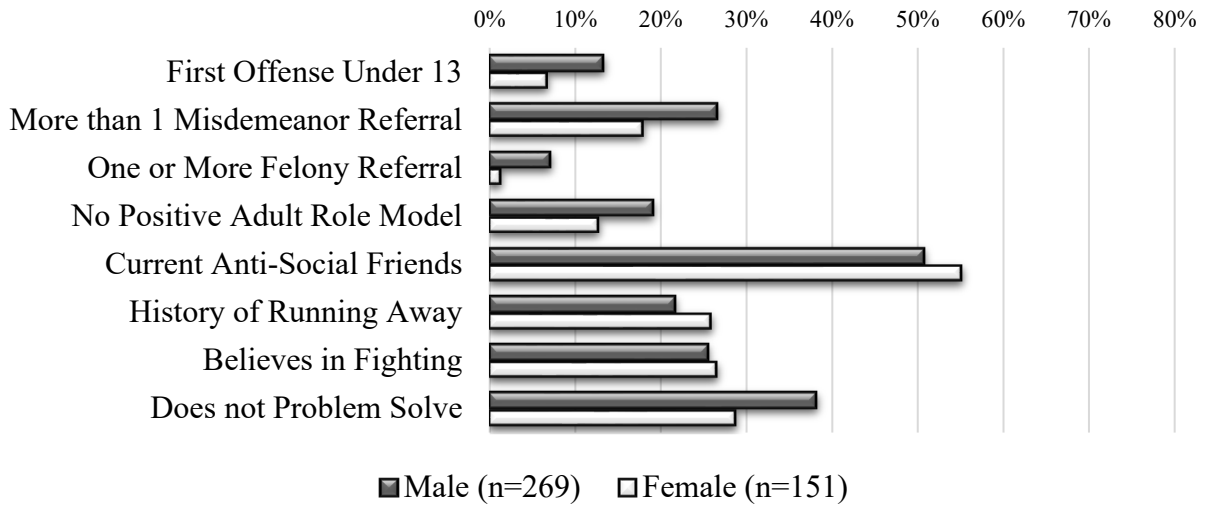


Figure 1.2: Risk Factor Frequency: Gender

Figure 1.3 breaks the frequency of risk factors down by white and non-white youth. Similar to gender, white and non-white youth appear comparable in their composition of risk factors. However, two risk factors present distinct differences. First, almost 70% of non-white youth were reported to have *anti-social friends*, compared to 50% of white youth. Secondly, non-white youth reported a *belief in fighting* 62% of the time, while white youth reported a *belief in fighting* 21% of the time. If not for those two factors, white and non-white youth appeared to have similar risk factor frequencies. The differences, which may be due to the limited sample size of non-white youth, warrant further monitoring.

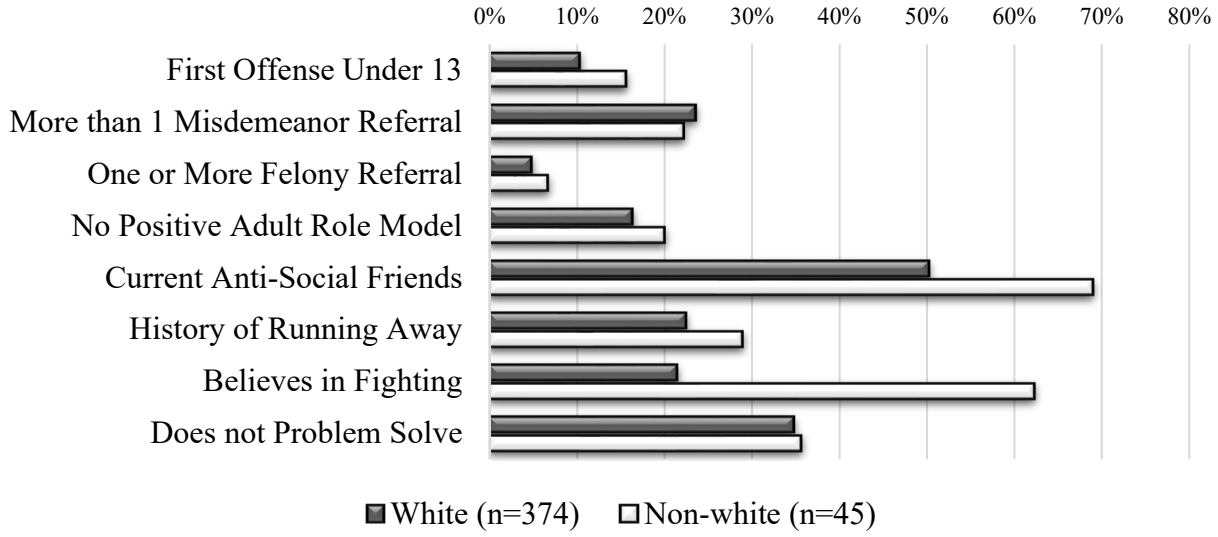


Figure 1.3: Risk Factor Frequency: Race

Total Risk Score. MJPS risk scores range from 0 to 8. A score of 0 is indicative of no risk factors and a score of 8 reflects an individual flagged for all risk factors. Figure 1.4 presents the overall risk score distribution for the pilot sample. Low risk scores range from 0 to 2, moderate risk falls in the 3-to-4 interval, and between 5 and 8 constitutes high risk. Approximately 68% (297) of all youth screened were predicted to have low recidivism risk within one year. Twenty-five percent (110) were predicted to be moderate risk, and fewer (6%; 27) high risk. No youth reached the maximum MJPS risk score of 8 points.

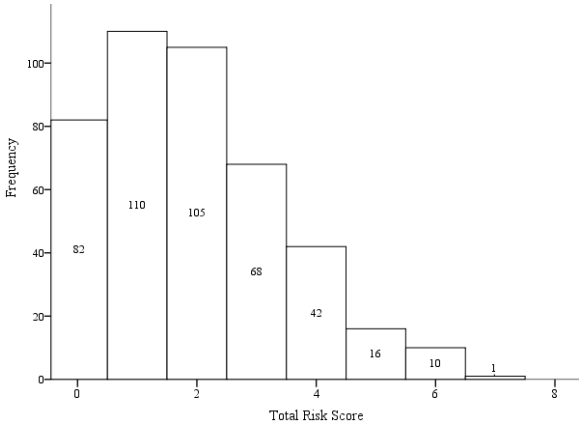


Figure 1.4: Distribution of MJPS Risk Scores (*N*=434)

Risk score distribution for race and gender are presented in Figure 1.6. [3] Interestingly, all race and gender subsamples are shown to have the same median risk score of 2. Non-white males had the greatest risk score variation compared to all other categories. White females had the lowest risk score distribution and non-white females and white males appeared to have identical risk score distributions.

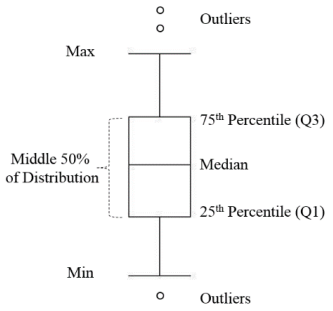


Figure 1.5: Interpreting Boxplots

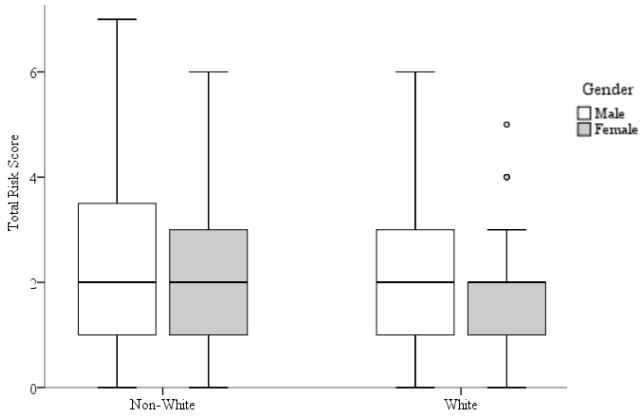


Figure 1.6: Distribution of Risk Scores by Race and Gender

[3] Boxplots (also called box and whisker plots) are created to display the distribution of risk scores between race and gender. Boxplots display the full range of variation (from minimum to maximum values), the likely range of variation (the inter-quartile range - IQR), and the typical or central value (median). The diagram (Figure 1.5) demonstrates the elements of the boxplot. The center box, from which the boxplot gets its name, is a display of the middle 50% of the distribution with the bottom of the box depicting the 25th percentile value, the line within the box is presenting the median value, and the box ends at the 75th percentile. Boxplots are a useful tool for examining the distribution variation and typical values. These displays are also useful for comparing distributions between different categories (e.g., race or gender) as shown in the analysis.

To further explore the difference in risk scores between race and gender Figure 1.7 displays the percent of individuals placed in each risk-level category (low, moderate, and high) based on race and gender. Risk-level categories between males and females were similar. In general, males were classified as more moderate and high risk than females and females were shown to have more low risk than males. Differences between white and non-white youth were more distinct. Non-white youth were more likely than white youth to be deemed moderate risk (31.8% compared to 24.8%) and significantly more likely to be classified as high-risk (13.6% to 4.8%).

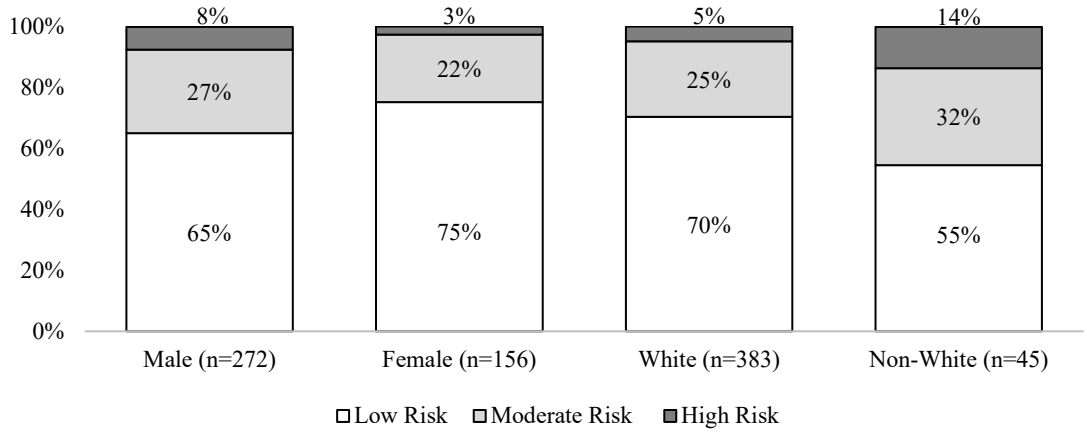


Figure 1.7: Risk Level Category by Race and Gender

Finally, Figure 1.8 presents mean risk scores by pilot district. The overall mean score is 1.93. Districts 5, 13, 1, and 19 are all close to the overall mean risk score. Differences can be seen between the district with the highest mean risk score and the district with the lowest. District 21 has the lowest mean MJPS risk score with 1.45 and District 11 has the highest mean risk score at 2.41. Natural variation among districts is expected. If these differences persist, however, or the differences grow, further investigation may be warranted to determine why risk is deemed higher or lower in one district over others.

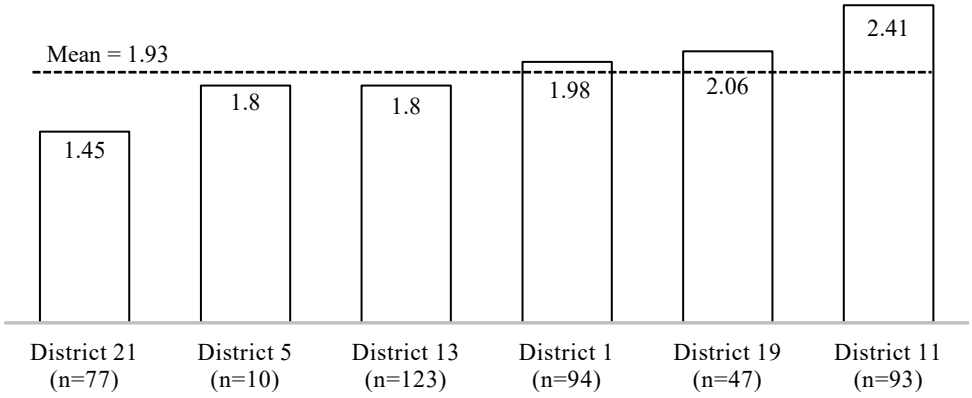


Figure 1.8: Average MJPS Risk Score by District

Face and Concurrent Validity. Face validity involves determining whether the instrument appears to reasonably measure variables of interest (see Babbie, 2015). The question asked when evaluating MJPS face validity was whether professionals using the instrument believed that it correctly predicted recidivism. The MJPS was constructed to collect this information from officers

after each administration of the instrument. After delivering the screener, officers were prompted to select what risk level they believe the youth occupied. Officers could agree with the MJPS risk level, disagree by selecting a lower risk level (override down), or disagree by selecting a higher risk level (override up). Figure 1.9 presents the frequency of agreements and disagreements for all districts.

Overall, officers agreed with the MJPS derived risk level 92.1% (400) of the time. All districts indicated a similar amount of agreement, ranging from 88% to 100%. The findings demonstrate a high degree of face validity where officers indicated they believe the tool accurately places the youth in the correct risk level.

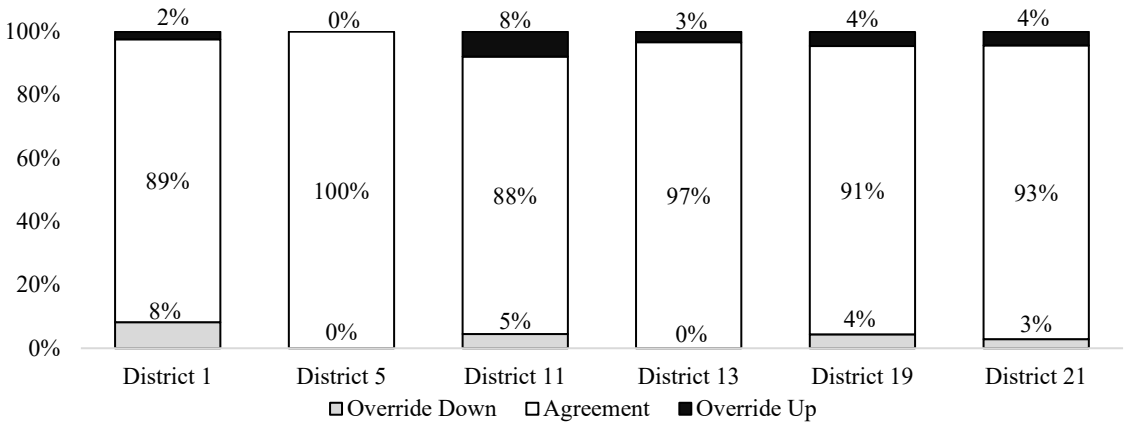


Figure 1.9: Agreement in Risk Level by District

An additional way to measure validity is to examine how comparable MJPS risk levels are to an already validated tool. This is typically referred to as concurrent validity, which is a type of criterion validity (see Litwin, 1995). During the pilot test, 265 of the 444 youth receiving the MJPS received both the MJPS and a Pre-screen BOT. Figure 1.10 presents the percent of youth found to have the same risk level on the MJPS and BOT (agreement), the percent of youth where the Pre-screen BOT indicated a higher risk level, and the percent of youth where the MJPS indicated a higher risk level.

As a reference, during the MJPS retrospective test, (McKay, Hollist, and Ranger, 2017) agreement between the MJPS and the Pre-screen BOT occurred approximately 67% of the time. Agreement between the MJPS and the Pre-screen increased during the pilot test to 80% (221). The Pre-screen indicated that certain youth should be classified at higher risk than the MJPS 6.8% (18) of the time, and the MJPS indicated that certain youth should be higher risk than the Pre-screen BOT predicted 13.6% (36) of the time.

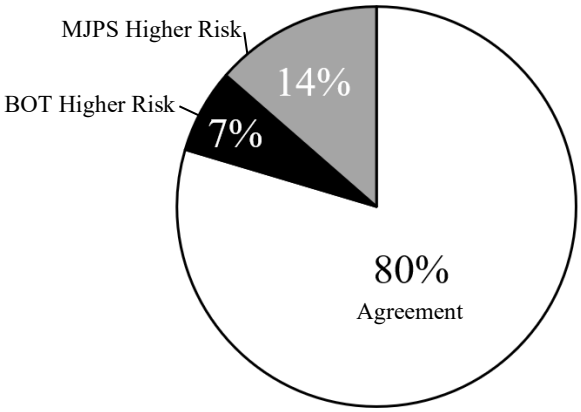


Figure 1.10: Agreement Between Pre-screen BOT and the MJPS Risk Level

Table 1.2 presents the bivariate correlations between the MJPS risk score, the Pre-screen BOT risk score, and probation officer professional opinion of risk. This analysis only includes data on youth who received both an MJPS and a Pre-screen (*N*=255). The MJPS risk score and professional opinion show a strong positive correlation with a Pearson's r of .872. Interestingly, the Pre-screen and professional opinion demonstrate a lower correlation with a Pearson's r of .711. Agreement occurred between the Pre-screen BOT risk score and professional opinion 83% of the time. Evidence presented here demonstrates probation officer are more confident in the MJPS compared to predictions from the Pre-screen. Risk scores from the MJPS and the Pre-screen had a moderate to strong positive correlation with a Pearson's r of .627 which is comparable to the correlation found during the retrospective test of the MJPS at .707 (McKay, Hollist, and Ranger, 2017).

Table 1.2 Bivariate Correlations Between MJPS, BOT, and Officers (*N*=255)

Risk Level Determined by	MJPS	Officer Judgement	BOT
MJPS	1	.887**	.627**
Officer Judgement	.887**	1	.711**
Pre-Screen BOT	.627**	.711**	1

Disposition. The MJPS was designed to separate low-risk youth who can be released with minimal supervision from medium and high-risk youth who may need increased supervision. At the beginning of this inquiry, it was expected that low-risk youth, as determined by the MJPS, would be placed in the least restrictive services, and, as risk level increased, so too would the severity of disposition. The pattern presented in Figure 1.11 appears to show the MJPS accomplished this objective. The most common disposition was "diversion," the least restrictive, followed by "informal probation," then "formal probation," the most restrictive. A handful of cases that were "dismissed" are not included in the figure below. As illustrated in Figure 1.11, 80% of low-risk youth were diverted, 65% of medium risk youth were diverted and only 33% of high-risk youth were diverted. The inverse pattern of diversion is shown for informal and formal probation. Only 15% of low-risk youth were placed on informal probation compared to 29% of medium risk and 41% of high risk. Additionally, no low-risk youth were placed on formal probation, 1% of medium risk youth and 15% of high-risk youth were placed on formal probation.

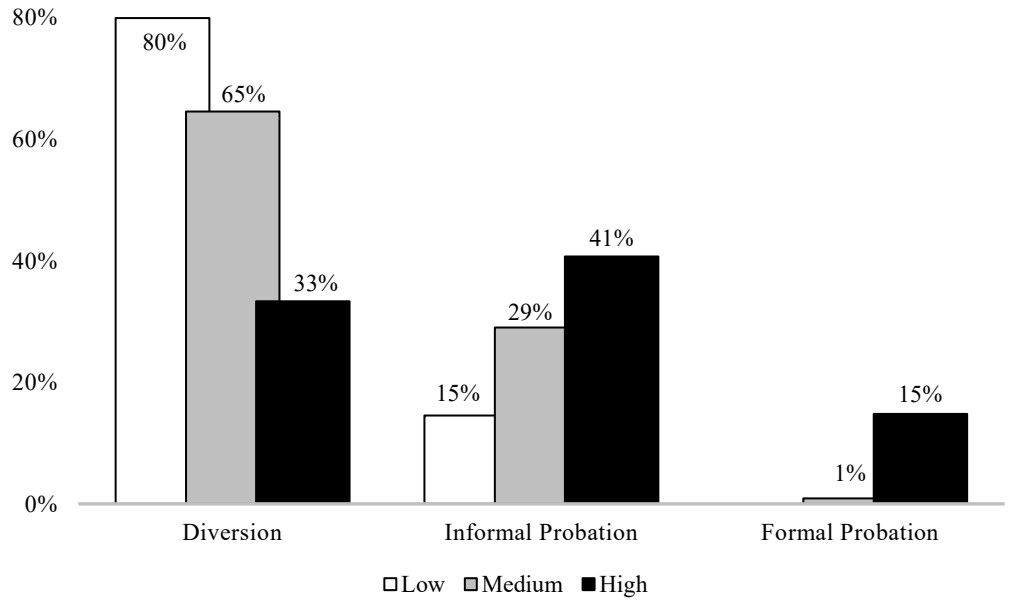


Figure 1.11: Risk Level by Disposition

Administration Errors. Ninety percent (400) of the screeners were administered without error. This is a high number for a risk assessment administered completely by hand. The remaining 10% (44) of the screeners had at least one error. The most common error found was the omission of practitioner judgement, which accounted for 33% of all errors. Omissions occurred when the probation officer did not indicate what risk level they believed the youth belonged to. Approximately one quarter (26%) of errors were due to the officer not listing the youth's disposition. Fourteen percent of the errors were due to the officer not answering one or more of the screener's risk factors. It was not possible to discern if officers left risk factor fields blank because they intended it to reflect a "0" score, or if they did not ask about the risk factor. Screeners with omitted risk factors were not included in this analysis. Four screeners included incorrect JCATS numbers. Other less common errors included no override justification, miscalculated or wrong risk score, and a non-recognized rater ID. In one instance, there were different outcomes and a discrepancy in indicated risk scores for a youth screened twice within a 48-hour period. Errors were minimal overall, however, and few hindered analyses. MJPS automation, which would coincide with prompts forcing responses to omitted questions, would greatly minimize potential for future errors.

Override Justifications. When practitioner judgement of risk did not agree with the MJPS risk assessment, officers were prompted to briefly explain why the risk level determined by the MJPS should be lower or higher. Overrides occurred for approximately 8% of the youth screened.

For half of the overrides (15 youth), officers stated that the risk should be lower than indicated by the screener. Officer explanations for those overrides fell into two categories. In the first category, officers noted that they performed an override to account for the fact that the MJPS did not measure protective factors. In the second category, officers noted that the MJPS did not measure mitigating factors. The following protective factors were itemized as rationale for override justifications: receiving mental health services, educational investment, and demonstrated growth and improvement over time. The remaining override down justifications fell into the category of

mitigating factors. Examples of mitigating factors cited included a lack of full participation in the offense for which the youth was cited, or succumbing to pressure from anti-social friends; a first offense; length of time since the young person was last in trouble; mental health issues; and bad judgement, rather than a criminal demeanor.

Overrides up, when officers believed the youth should have been deemed higher risk than indicated by the MJPS, occurred with 18 of the screened youth. Similar to overrides down, override up justifications fall into one of two categories. In the first category, it was noted that the youth had additional risk factors not included in the MJPS. In the second category, officers articulated an override due to aggravating factors not measured by the MJPS. Additional risk factors included family problems (e.g., family history of incarceration, family substance use, or unstable living conditions); mental health challenges (e.g., high ACEs score, or history of trauma and or victimization); poor school performance; and homelessness. Aggravating factors that officers used as override up justifications included an absence of learning or changing from the previous offense; escalating substance abuse; a previous offense not yet in the system, or unreported history of crime; a particularly severe offense (e.g., sex offense); highly oppositional behavior; a criminal mind; and previous placements. Section 3 includes a further analysis of to what extent the override decisions assisted in recidivism prediction accuracy.

Discussion

Several pieces of evidence presented here demonstrate the MJPS performed well during the pilot test. Possibly the most important piece of evidence is an extremely high degree of face validity. Officers agreed with the MJPS-indicated risk levels assigned to screened youth almost every time the instrument was administered (92%). Officer agreement was greater for the MJPS than the Pre-screen BOT an indication officers were more confident in the new screener. Additionally, the Pre-screen BOT agreed with the MJPS risk level the majority of the time demonstrating a measurement of concurrent validity. Face and concurrent validity are important for gaining buy-in. Officers will not use a risk assessment with any fidelity if they do not trust it. The MJPS was also able to predict disposition severity for screened youth. The overall goal of the MJPS is to locate as many low risk youth as possible who do not need restrictive dispositions. Evidence to support this goal was presented with the majority of screened youth scoring in the low-risk score range and this level of risk was most associated with a disposition of diversion, the least restrictive services. Overall, the risk score distribution and risk factors frequency looked good. Risk scores tended to be low and risk factors frequencies were diverse and not overly occurring. Areas of possible concern arose, however, including low MJPS administration rates for non-white and felony offences and differences were located among risk factors and risk scores between white and non-white youth. Differences between race may be reflective of the low sample size of non-white youth but will continue to be an important inquiry moving forward.

To provide context for the findings discussed in this section, a survey was distributed to all juvenile probation officers in Montana. The survey asked about officers' professional opinion related to the MJPS, the Pre-screen BOT and also the Full-screen BOT. Findings from the survey are presented in the following section.

JUVENILE PROBATION SURVEY

The Montana Juvenile Probation Survey invited Youth Court Services officers administering risk assessments to voice their professional opinions about instruments currently in use. Inclusion of the probation officer perspective constitutes an important step to increase buy-in for utilization of any risk assessment instrument. This section details data collection methodology for the Juvenile Probation Survey, along with results and analyses. It also explores improvements that should be made to the Pre-screen BOT, how data is collected for the risk assessments, officer confidence levels with the Montana Juvenile Probation Screener (MJPS), and how it compares to the Pre-screen Back on Track (BOT).

Methods

On Oct. 11, 2018, the Criminology Research Group (CRG) distributed the Montana Juvenile Probation Survey using Qualtrics Survey Software. Sent to all Youth Court Services probation officers across the state's 22 judicial districts, the online survey asked a variety of questions about the Full-screen BOT, the Pre-screen BOT, and the MJPS. Twenty-one of 22 districts had at least one officer take part in the survey. The CRG collected responses after the survey closed on Nov. 1, 2018 and analyzed the results in SPSS. While all juvenile probation officers across Montana were given the survey and asked to respond to questions about the Pre-screen and Full-screen BOT, only officers from MJPS pilot districts were given questions related to the MJPS.

Data. In total, 80 probation officers completed the Montana Juvenile Probation Survey. Of those respondents, 20 individuals from the six districts participating in the MJPS pilot test completed the survey. Table 2.1 displays the distribution of officers who responded to the survey from the six districts participating in the pilot test.

Table 2.1: Respondents to Survey from Pilot Districts (*N*=20*)*

Pilot District	Percent
District 1	25%
District 5	10%
District 11	25%
District 13	10%
District 19	10%
District 21	20%
Total	**100%**

Results

The MJPS was designed with the intention that it would replace the Pre-screen BOT as a brief and accurate recidivism risk predictor. If the MJPS were to replace the Pre-screen, it should remedy Pre-screen limitations. To examine to what extent the MJPS could be an improvement, all officers across Montana were asked to share their ideas on strategies for improving the Pre-screen BOT. Of the 80 officers who filled out the survey, 25 (31%) responded to that open-ended question. Summaries of officer-recommended strategies are included below.

Four officers said there should be more training on the Pre-screen, so districts all use it in the same manner. Those officers said that increased training and consistency of use would make the Pre-

screen a more effective tool. Two officers questioned to what extent there was empirical evidence to demonstrate Pre-screen effectiveness. For instance, one officer stated, "continue study to determine efficacy." Other officers stated that they simply do not see the utility of the Pre-screen at all. For example, one officer wrote, "I don't understand the purpose of the BOT… The true value is to statisticians, budget people, and politicians … It has near zero effect on my case planning with youth." One officer stated that the tool would be more effective if it "spit out a case plan or recommendation."

Three officers stated that Pre-screen questions are overly invasive. As one officer said, for example, "I don't feel like, when you are just getting to know a kid, you can ask some of the questions on this, because they are very personal." Another stated, "It is very uncomfortable for parents/youth to disclose all (of) that kind of information."

Officers had contrasting views on the length of the tool and the information collected. A few officers stated that the Pre-screen is too long. "It should be shorter with questions that are easier to answer," one officer said. Others stated the Pre-screen should collect more information. As another officer noted, "More questions pertaining to the youth to get a bigger picture."

Five officers who responded to the open-ended question about how to improve the Pre-screen BOT said they believed it worked well for their respective districts. Some of those respondents said that while they like the Pre-screen, they only utilize the Full-BOT, and this works well for them.

Additional discussion about the Montana Juvenile Probation Survey related to MJPS administration follows. Table 2.2 documents the ways in which MJPS information was collected, as reported through the Juvenile Probation Survey from those districts that piloted the MJPS. All 20 respondents stated that they interviewed the youth to answer MJPS questions. All but two respondents reported that they also interviewed the parent or guardian. Nearly half of respondents reported that they used the Juvenile Court and Assessment Tracking System (JCATS) data to inform MJPS inputs. Two officers said they talked with school officials or teachers and two referred to law enforcement records.

Table 2.2: How MJPS Information was Collected (*N*=20)

Mode	Percent
Interview with Youth	100%
Interview with Parent/Guardians	90%
Information From JCATS	45%
Talk with School Officials/Teachers	10%
Law Enforcement Notes	10%

A primary goal associated with MJPS introduction is reducing the time required to administer the screener in comparison to the Pre-screen BOT. Table 2.3 presents Juvenile Probation Survey results related to the length of time required to administer the Pre-screen BOT relative to the MJPS. Samples sizes are different between these questions because only those officers who belong to pilot districts were ask to respond to the length of time for the MJPS. More than half (53%) of the sample described the MJPS as taking less than 10 minutes to complete, while only one-quarter (24.3%) of respondents said the Pre-screen BOT took less than 10 minutes to administer. Nearly the entire sample (96%) agreed that the MJPS took fewer than 30 minutes to complete, one

respondent stated that it took between 51 and 60 minutes to complete. Most (83%) respondents said the Pre-screen BOT took fewer than 30 minutes to administer, but several respondents (17%) said that the Pre-screen took between 31 and 61+ minutes. Overall, most officers stated the MJPS reduced administration time from that required to deliver the Pre-screen BOT.

Table 2.3: Time to Complete Pre-screen Versus MJPS

Time to Complete Screener	MJPS	Pre-Screen
1 to 10 minutes	53%	24.3%
11 to 20 minutes	32% (96%)	32.9% (83%)
21 to 30 minutes	11%	25.7%
31 to 40 minutes	0%	2.9%
41 to 50 minutes	0%	5.7%
51 to 60 minutes	5%	7.1%
61+ Minutes	0%	1.4%
Sample Size	19	70

Nearly half (45%) of the administrators from MJPS pilot sites responded *yes* to utilizing the MJPS to determine risk to re-offend, as illustrated in Table 2.4. An additional 20% said they *sometimes* used the screener and 30% said *no,* they did not use the MJPS to determine recidivism risk. Officers were asked if they utilized the Pre-screen BOT to determine re-offense risk. Of the 77 total respondents to this question, 60% answered *yes*, 25% stated *sometimes*, and 15% said *no*.

Table 2.4: Utilize MJPS and Pre-screen to Determine Risk to Re-Offend

Response	MJPS	Pre-Screen
Yes	45%	60%
No	30%	15%
Sometimes	20%	25%
Sample Size	19	77

Officers were asked to what extent they agreed the Full-screen BOT, the Pre-screen BOT, and the MJPS informed case management decisions. Answers to this question are displayed in Table 2.5. Respondents overall stated that they *somewhat agree* or *strongly agree* each of the assessments informed case management decisions. Respondents indicated the most agreement for application of the Full-BOT to inform case management decisions, at 78%; 75% said the MJPS informed such decisions, while 68% said the Pre-screen BOT informed case management.

Table 2.5: Tool Informs Case Management Decisions

Response	MJPS	Pre-Screen	Full Screen
Strongly agree	5%	6%	16%
Agree	40% (75%)	23% (68%)	22% (78%)
Somewhat agree	30%	39%	40%
Somewhat disagree	15%	13%	8%
Disagree	5%	13%	10%
Strongly disagree	5%	7%	4%
Sample Size	20	70	73

Table 2.6 displays officer confidence levels in the capability of the Pre-screen BOT and the MJPS to predict recidivism risk. While no officer from the pilot districts stated they have *a great deal* of confidence in the MJPS's ability to predict risk, 90% of the sample said their level of confidence was *a lot* or *a moderate amount*. Ten percent reported they had *a little* confidence in the MJPS, no officer reported *none at all*. When all officers across Montana were asked about confidence in the Pre-screen BOT's ability to predict recidivism risk, 4% of respondents answered *a great deal*, 70% answered *a lot* or *a moderate amount*. Nearly a quarter (23%) of officers surveyed said they have *a little* confidence in the Pre-screen BOT, 3% answered *none at all* related to the Pre-screen.

Table 2.6: Confidence in Ability to Predict Risk

Response	MJPS		Pre-Screen	
A Great Deal	0%		4%	
A Lot	35%	90%	24%	74%
A Moderate Amount	55%		46%	
A Little	10%		23%	
None At All	0%		3%	
Sample Size	20		70	

Pilot district officers were prompted to answer the question, "To what extent do you agree with the following statement: Overall, the Montana Juvenile Probation Risk Screener is an improvement over the Pre-screen BOT." Table 2.7 presents answers provided to this question. Three-quarters of the officers (74%) said they *strongly agree*, *agree*, or *somewhat agree* with this statement. Twenty six percent stated that they *somewhat disagree*, *disagree*, or *strongly disagree*. Officers were then asked to explain why they either agreed or disagreed that the MJPS was an improvement. Those responses are summarized below.

Table 2.7: Screener is an Improvement Over Pre-screen BOT (*N=20)*

Response	Percent	
Strongly Agree	21%	
Agree	32%	74%
Somewhat Agree	21%	
Somewhat Disagree	11%	
Disagree	5%	
Strongly Disagree	11%	

Eight of the 14 officers who agreed the MJPS was an improvement over the Pre-screen explained why they believed that was the case. Attributes cited included brevity, directness, and that the MJPS seems to predict the same outcomes as the Pre-screen BOT. All five officers who stated that it was not an improvement noted that the MJPS collects only limited information. Several officers who said the MJPS was not an improvement indicated that they actually like the tool, but that it lacked information necessary to accurately determine youth services.

The Juvenile Probation Survey asked officers if any MJPS risk factors were confusing or challenging to collect and why this was the case. For those responding to that question, four of the seven risk factors were noted as confusing or challenging to collect: (1) *positive adult role model*,

(2) *history of running away*, (3) *belief in fighting*, and (4) *problem solving*. *Positive adult role model* and *problem solving* were selected with the highest frequency as difficult factors to identify. Officers said those questions are subjective and at times ambiguous. They stated further that unclear guidelines were provided to collect relevant information. *Belief in fighting* was flagged by two officers as ambiguous, but they offered no additional explanation. One officer questioned if running away for under 24 hours counted towards the *history of running away* risk factor.

The Juvenile Probation Survey asked officers what could be done to improve the MJPS. A wide range of responses came from officers who answered the open-ended question. One officer stated the MJPS may be helpful as a quick risk assessment, but that it is not an adequate substitute for the BOT. That comment raises an important point—the MJPS was designed only to serve as a quick risk assessment. The MJPS should not be conceptualized as a substitute for the Full-BOT. Rather, the MJPS should work in conjunction with the Full-BOT, with youth who score high on the MJPS also receiving the Full-BOT. Additional comments about possible MJPS improvements included a suggestion from one officer about 'scrapping it,' another commenter said the MJPS works great and is easy to use.

Discussion

The Montana Juvenile Probation Survey was administered statewide with only one judicial district failing to respond. Survey results related to MJPS administration demonstrate probation officer confidence in the instrument and indicate that officers perceive it as an improvement from the Pre-screen. Survey findings, however, indicate that probation officer buy-in and screener fidelity continues to challenge the utility and applicability of all risk and needs assessments.

Adoption of the MJPS by Youth Court Services would eliminate some of the problems and concerns with the Pre-screen BOT raised by probation officers. First, the MJPS addresses the concern that there is a lack of research support behind the assessment. There is a significant and growing body of empirical evidence to support MJPS utilization in Montana. The MJPS has been demonstrated through a retrospective test to be an accurate recidivism predictor (see McKay, Hollist, and Ranger, 2017), and this analysis further examines accuracy through prospective testing. Secondly, some officers said the Pre-screen BOT is too time consuming. The MJPS is an improvement in this aspect as well. Some Pre-screen questions were described as overly invasive. Alternatively, there are no distinctly personal questions on the MJPS. Moreover, because the MJPS is significantly shorter than the Pre-Screen BOT, training all Youth Court Services officers to use the tool with the same universal delivery methodology will be a more obtainable goal than prior assessments. Overall, the Criminology Research Group believes MJPS brevity, when coupled with universal training, will increase instrument fidelity.

The following section presents information on MJPS accuracy predicting recidivism. Demonstrating such predictive validity constitutes an important step toward gaining probation officer confidence and thus to continue growing buy-in from those slated to use the tool. It is also important to examine how well the MJPS predicts recidivism among different samples, such as across race and gender, to ensure that all youth are treated fairly, and that prediction accuracy is not lost when utilizing the tool with different samples.

PREDICTIVE VALIDITY

On Feb. 10, 2020, one year after the Montana Juvenile Probation Screener (MJPS) pilot test concluded, the Criminology Research Group (CRG) received from the Montana Office of the Court Administrator (OCA) Youth Court Services, recidivism data on youth administered the MJPS. This final section examines MJPS predictive validity. In the first section, face and concurrent validity were analyzed by examining agreement between risk levels from the MJPS, officer's professional opinion, and the Pre-screen BOT. Predictive validity, alternatively, refers to a risk assessment's ability to accurately predict an outcome of interest. The MJPS was created to predict recidivism risk. Several measurements of predictive validity are presented here and when possible, MJPS results are compared to those from the Pre-screen BOT. Additionally, we examine probation officer's ability to predict recidivism and how the MJPS and professional discretion can work together in to maximize prediction accuracy.

Methods

Data. Recidivism is defined here as a new offense leading to an intake during the one-year period following MJPS administration. To ensure a recidivism could be reviewed for a full year for all youth administered the MJPS, information on all youth intakes in the six MJPS pilot districts between July 3, 2018 and Feb. 1, 2020 were queried from the Juvenile Court Assessment and Tracking System (JCATS). Recidivating offenses committed outside of the one-year period after MJPS administration were not included in this analysis. For youth with multiple recidivating offenses within the one-year period, the single most serious recidivating offense was selected and used for the analysis. The CRG and OCA decided on three categories of recidivism: (1) *any recidivism* which includes technical violations, city ordinance, status, misdemeanor, or felony offenses; (2) *misdemeanor or felony recidivism;* and (3) *felony recidivism*. These broad outcomes are not mutually exclusive but are useful for demonstrating how well the MJPS predicts recidivism for all offense types and whether offense severity impacts prediction accuracy. All recidivating categories are dichotomously coded, "1" indicates the presence of a recidivating offense and "0" for those youth who did not have a recidivating offense.

Youth were matched to recidivating offenses using their JCATS number, a unique identifier assigned to each young person. If the youth's JCATS number was associated with a Youth Court Services intake after MJPS administration, the new intake was flagged as a recidivating offense.

The term "base rate" refers to the percentage of youth who committed a recidivating offense during the one-year risk period. The base rate for each of the three recidivism categories is itemized below. The final sample size used for the analyses in this section is 422 youth. Recidivism data was unable to be matched or errors occurred during the administration process for 22 youth, these cases were removed before the analysis.

- *any recidivism* 22%
- *misdemeanor or felony recidivism* 19%
- *felony recidivism* 5%.

As discussed in the first section of this report, certain groups of youth, such as white non-felony offenders, were more likely to be given the screener than others. Table 3.1 displays recidivism rates and how they potentially were impacted by the non-random selection of screened youth.

Overall, youth receiving the MJPS had lower recidivism rates than the general intake population among the pilot sites. The first type of recidivism shown in Table 3.1 is *any recidivism*. This category displayed the largest difference between the MJPS sample and the general population. For the full sample and each subsample under *any recidivism*, the recidivism rate was approximately 10 percentage points lower for youth administered the MJPS than the general population with the exception of non-white youth. The American Indian sample administered the MJPS had a slightly higher *any recidivism* rate at 47% compared to 45% for the general population of American Indians going through intake during the pilot test. Non-white/non-American Indian youth (Other) receiving the MJPS were significantly more likely to have *any recidivism* at 42% compared to the general population with 30%. For the full sample, MJPS screened youth have a *misdemeanor/felony recidivism* rate eight percentage points lower than the general population of youth. Subsamples show similar results, with the American Indian sample, once again, being an outlier. American Indians receiving the MJPS had a *misdemeanor/felony recidivism* rate five percentage points higher than the general population of American Indian youth. *Misdemeanor/felony* rates for non-white/non-American Indian youth administered the MJPS looked similar to their respective population. Finally, *felony recidivism* rates for screened youth and the population look similar for the full sample and all subsamples. According to the full sample, the *felony recidivism* rate is only one percentage point lower for the screened youth compared to the general population.

Females administered the MJPS were also found to be much less likely to have *any recidivating* offense or a *misdemeanor/felony recidivating* offense relative to the general population of females going through intake at the six pilot sites. The *felony recidivism rate* was equivalent for screened and general populations of females. Males were also less likely to have a recidivating offense if they were screened with the MJPS, but differences were not quite as extreme as found with females.

Table 3.1 Recidivism Rate: MJPS Youth Versus All Pilot Site Intakes

Sample:	Any Recidivism		Misd/Fel Recidivism		Felony Recidivism	
	MJPS	All Pilot Site Intakes	MJPS	All Pilot Site Intakes	MJPS	All Pilot Site Intakes
Full Sample	22%	32%	19%	27%	5%	6%
Gender						
Male	27%	32%	23%	28%	6%	7%
Female	17%	28%	14%	21%	3%	3%
Race						
White	20%	29%	18%	24%	5%	5%
American Indian	47%	45%	42%	37%	11%	10%
Other	42%	30%	27%	30%	8%	6%
Sample Size	422	1408	422	1408	422	1408

Across all recidivism types, and regardless of whether the youth was given the MJPS or not, American Indian youth were significantly more likely to have a recidivating offense compared to their white counterparts. Nearly half (45%) of American Indian youth going through intake at the pilot sites had *any recidivism* compared to only 29% of white youth. American Indians also had a much higher *misdemeanor/felony recidivism* rate compared to white youth (37% and 24% respectively) and double the *felony recidivism* rate of white youth (10% and 5% respectively).

Non-white/non-American Indian youth had slightly higher recidivism rates than white youth but not to the same extent as American Indian youth. Additionally, males consistently present higher recidivism rate than females.

Analytic Strategy. The first step in assessing MJPS predictive validity was to examine the percentage of youth within each risk score (0 to 7) with a recidivating offense. This simple analysis presents a meaningful output. If the instrument is an accurate recidivism predictor, recidivism rates will increase with each point increase in risk score. This same analysis was conducted using the designated risk thresholds of low, moderate, and high risk.

The second analysis incorporated logistic regression (see Baglivio and Jackowski, 2013; Baglivio, 2009; Conley, Spurzem, Marsh, and Hazlett, 2009). In this analysis, the overall risk score determined by the MJPS served as the independent variable, while the dependent variable was whether the youth committed a recidivating offense. Race and gender are included as independent variables in the models to hold their respective impacts on recidivism constant. Logistic regression results demonstrate whether an increase in MJPS risk score boosts the likelihood of recidivism and whether the increase is due to random chance alone.

The final analysis to assess predictive validity examined area under the curve (AUC) derived from receiver operating characteristic (ROC) analysis. ROC analysis plots the true positive rate against the false positive rate for each risk assessment point threshold. The AUC statistic measures this relationship ranging from 0.0 to 1.0. An AUC of 0.0 is indicative of perfect negative prediction and 1.0 is interpreted as perfect positive prediction. An AUC of 0.5 is no better than random chance. Rice and Harris (2005) developed guidelines based on AUCs ranges that serve as a standard in risk assessment literature. Table 3.2 presents these ranges.

Table 3.2 AUC Effect Size Ranges

Effect Size	AUC Range
Negligible	<0.56
Small	0.56 to 0.63
Moderate	0.64 to 0.70
Large	≥0.71

When applicable, MJPS analyses results were compared to Pre-screen BOT results. This provided insight into how well the MJPS would do if it were to replace the Pre-screen BOT and to what extent it compares to the Pre-screen BOT.

Results

Risk scores. Recidivism rates for each MJPS risk score are presented in Table 3.3. Each point increase in risk score on the MJPS should indicate an actual increase in recidivism risk. Thus, each point increase should be associated with an increase in the recidivism rate. In general, this pattern is found in Table 3.3.

The MJPS performs well locating low-risk youth. Youth scoring 0 on the MJPS are unlikely to commit any category of recidivating offense. Only 7% of youth scoring a 0 commit *any recidivism*, 6% committed a *misdemeanor/felony recidivism*, and 0% committed *felony recidivism*. There was a significant increase in recidivism for youth scoring a 1 on the MJPS. *Any recidivism* increases to

17%, *misdemeanor/felony* to 15%, and *felony recidivism* to 3%. Notably, when a youth scores a 1 or a 0 on the MJPS, they are less likely than a randomly selected young person in the general population to commit any type of recidivating offense. Youth with an MJPS score of 2, once again, have an increased recidivism rate for each category (*any recidivism* 24%, *misdemeanor/felony* 21%, and *Felony* 5%). Continuing the ideal pattern, youth scoring a 3 on the MJPS had an increase in *any recidivism* (33%) and *misdemeanor/felony recidivism* (29%). *Felony recidivism* for youth with an MJPS score of 3 occurred at a similar re-offense rate to those with a 2 score, at 4%. Youth placed at 4 or 5 had similar recidivism rates as those who scored a 3 for *any recidivism* and *misdemeanor/felony recidivism,* but there is a significant increase among those scoring 6 (*any recidivism* 60%, *misdemeanor/felony* 50%). *Felony recidivism* rates increased for youth scoring a 4 (9%), 5 (12%) and 6 (50%). Only one youth received an MJPS score of 7, and the tool correctly predicted recidivism under the *misdemeanor/felony recidivism* category. In general, it is apparent that there is a strong positive correlation between risk score and recidivism rate.

Table 3.3: Recidivism Percent by Risk Score (*N*=438)

MJPS Risk Score	*n*	Any Recidivism	Misdemeanor/Felony	Felony
0	82	7%	6%	0%
1	110	17%	15%	3%
2	106	24%	21%	5%
3	69	33%	29%	4%
4	43	30%	23%	9%
5	17	29%	29%	12%
6	10	60%	50%	50%
7	1	100%	100%	0%
Base Rate	438	22%	19%	5%

Table 3.4 continues to illustrate this line of inquiry by examining the recidivism rate for each MJPS-recommended risk category (low, moderate, high) and risk categories from the Pre-screen BOT. Data included in Table 3.4 includes information only from youth who were given both the MJPS and the Pre-screen BOT, which enabled direct comparisons to be made.

As illustrated in Table 3.4, both the MJPS and Pre-screen performed well identifying low-risk youth. Only 17% of low-risk youth receiving both the MJPS and the Pre-screen committed *any recidivism*, 14% committed a *misdemeanor/felony recidivism*, and 3% of low-risk youth committed *felony recidivism*. Moderate risk on both tools is shown to increase the percent of recidivism compared to low risk. However, youth receiving moderate scores on the Pre-screen BOT had significantly higher rates for *any* and *misdemeanor/felony recidivism* compared to those receiving moderate or high scores on the MJPS. Recidivism rates in the moderate MJPS risk score category increased to 34% for *any recidivism*, 26% for *misdemeanor/felony*, and 10% for *felony,* while the Pre-screen increased to 46% for *any*, 41% for *misdemeanor/felony*, and 14% for *felony*. The MJPS and Pre-screen also performed differently with high-risk youth. For each recidivism type, youth scoring high on the MJPS were shown to be more likely to commit a recidivating offense. Scoring high on the MJPS increases *any recidivism* to 42%, *misdemeanor/felony* to 33%, and *felony recidivism* to 25%. Interestingly, recidivism rates for the Pre-screen actually decrease from moderate to high risk for *any recidivism* and *misdemeanor/felony recidivism*, and slightly

increase for *felony recidivism*. High-risk youth, as designated by the Pre-screen, had a 31% *any recidivism* rate, 25% *misdemeanor/felony*, and a 19% *felony recidivism* rate.[4] Overall, the MJPS performed preferably compared to the pre-screen showing a more consistent pattern of increased recidivism as risk level increases.

Table 3.4: Recidivism Rate by MJPS & Pre-screen BOT Risk Level (*N*=262*)*

	Any Recidivism		Misdemeanor/Felony		Felony	
Risk Level	MJPS	Pre-screen	MJPS	Pre-screen	MJPS	Pre-screen
Low	17%	17%	14%	14%	3%	3%
Moderate	34%	46%	26%	41%	10%	14%
High	42%	31%	33%	25%	25%	19%
Base Rate	22%		18%		6%	

Agreement. While noted as playing a major role in the success of a risk assessments (Andrews et al., 1990), officer discretion is often an overlooked feature in predictive validity analyses. Girard and Wormith (2004) found that adjusted risk scores based on officer discretion slightly increased the correlation of a new conviction from original risk scores. Results are more commonly found in the opposite direction however, where prediction accuracy decreases by adjusting risk scores to include professional discretion (e.g., Carns and Martin, 2011; McCafferty, 2017; Vaswani, and Merone, 2014; Wormith, Hogg, and Guzzo, 2012; 2015).

During the MJPS pilot test, officers agreed with the MJPS indicated risk 92% (400) of the time with only 5% (18) stating the indicated risk was too low and 4% (18) saying the indicated risk was too high. Table 3.5 displays the agreements and overrides based on *misdemeanor/felony recidivism* rates.

Interestingly, when the MJPS-indicated decision and the administrating officer agreed on risk, recidivism prediction was optimized. For those youth where agreement occurred, 13.5% of low-risk youth had a *misdemeanor/felony recidivism*, 30.3% of moderate risk youth had such an offense, and 43.5% of high-risk youth had a *misdemeanor/felony recidivism*. MJPS score overrides occurred infrequently. When overrides did occur, however, they were also shown to be beneficial for prediction accuracy. Particularly helpful was officer decisions to designate a youth low risk, despite an MJPS finding of moderate risk. Of the 14 youth in which this scenario occurred, not one had *misdemeanor/felony recidivism*. Additionally, when officers deemed a youth moderate risk, though the MJPS found that individual to be high risk, there was only a 25% recidivism rate, which is comparably low. Officers were also able to locate groups of higher risk youth than indicated by the MJPS. Youth placed in the category of override up from low to moderate risk had a 38.5% recidivism rate, and young people whose scores were moved from moderate to high had a 40% recidivism rate.

[4] A table similar to 3.4 was created using all available data (*N*=444) and is presented in Appendix F. Overall, patterns are the same for the full MJPS data with only minor differences in recidivism percentages. Also see Appendix G for a figure of these findings.

Table 3.5 Misdemeanor or Felony Recidivism by Agreement (*N*=422)

MJPS to Rater Judgement	*n*	No	Yes	Total
Agreement (91% of cases)				
Low	274	86.5%	13.5%	100.0%
Moderate	89	69.7%	30.3%	100.0%
High	23	56.5%	43.5%	100.0%
Override Up (5% of cases)				
Low to Moderate	13	61.5%	38.5%	100.0%
Moderate to High	5	60.0%	40.0%	100.0%
Override Down (4% of cases)				
Moderate to Low	14	100.0%	0.0%	100.0%
High to Moderate	4	75.0%	25.0%	100.0%

To explore agreement and recidivism further, Table 3.6 presents the actual MJPS risk score (0 to 7) and the percent of youth with a recidivating offense for each category. The agreement section in Table 3.6 displays similar findings from those illustrated in Table 3.3: recidivism rates increase for each step up in MJPS risk score, except for the moderate category, where a score of 3 has a greater recidivism rate (37%), than a score of 4 (22%) and 5 (33%). This table helps to pinpoint scores that were overridden, and when the override worked versus when it did not. When youth scored a 0 or a 1 and the officer believed the youth was higher risk than low risk, none of the youth committed a recidivating offense. However, when youth scored a 2—the highest placement in the low-risk category—and the officer believed the youth was higher risk, 46% of the young people had a recidivating offense. Scores on the edge of risk thresholds appear to be vulnerable to diverging recidivism rates. Scores exemplifying this are shown with asterisks in Table 3.6. A score of 5—the lowest high-risk slot—illustrates this scenario. When youth scored a 5, but officers believed the youth was lower risk, the recidivism rate was relatively low at 25%. While these results present interesting insights, they should be interpreted cautiously, as the small sample sizes analyzed here make recidivism rates unstable.

Table 3.6 Recidivism Rate by MJPS Risk Score and Agreement Category

MJPS Score	Agreement			Override Up		Override Down	
	Low	moderate	High	Low to Moderate	Moderate to High	Moderate to Low	High to Moderate
0	5%	-	-	0%	-	-	-
1	16%	-	-	0%	-	-	-
2	18%	-	-	46%*	-	-	-
3	-	37%	-	-	33%	0%	-
4	-	22%	-	-	50%*	0%	-
5	-	-	33%*	-	-	-	25%*
6	-	-	50%	-	-	-	-
7	-	-	100%	-	-	-	-

Table 3.5 and 3.6 above demonstrate how a risk assessment and professional judgment can work together to provide the most accurate prediction. Risk assessments have consistently been demonstrated to produce the most reliable, consistent, and unbiased judgment of youth risk (Hilton, Harris and Rice 2006; Bishop and Trout 2002; Wilcox, Beech, Markall, and Blacker 2009) but professional discretion can be beneficial and should never be removed from important decisions. These tables also provide evidence that while the three MJPS risk threshold categories define risk level relatively well, overall risk score (0 to 8) is a more dynamic and accurate

recidivism predictor. Based on these findings, it is recommended that officers examine the actual risk score (0 to 8) instead of relying completely on the three risk categories (low, moderate, high). Further, the MJPS-indicated decision when applied alongside professional judgement is particularly effective predicting recidivism for scores on the verge of risk thresholds (2 and 5) and the moderate scores (3 and 4).

Logistic Regression. Table 3.7 presents the results of logistic regression, which analyzes the effect that the total MJPS risk score (0 to 7) has on predicting the dichotomous dependent variable of recidivism while holding the effect of gender and race constant. Statistical significance (*) indicates the findings from logistic regression are not likely the result of random chance.

Logistic regression results show the MJPS risk score is a significant recidivism predictor for each category, holding race and gender constant. For each point increase in MJPS risk score, there is a 43% increase in the likelihood of *any recidivism*, a 40% increase in the likelihood of *misdemeanor/felony recidivism*, and an 89% increase in the likelihood of *felony recidivism*. Additionally, results indicate that being male is a statistically significant predictor for two of the recidivism categories. Male youth are 81% more likely to commit a*ny recidivism* and 77% more likely to commit a *misdemeanor or felony recidivating* offense compared to female youth holding race and MJPS risk score constant. Biological sex, however, is not a significant predictor of *felony recidivism*. Moreover, white youth, relative to non-white youth, are 64% less likely to commit *any recidivism* holding gender and MJPS risk score constant. Race is not a significant predictor for *misdemeanor/felony recidivism* or *felony recidivism* alone.

Table 3.7 Logistic Regression Predicting Recidivism (*N*=422)

	Any Exp(b)	Misdemeanor or Felony Exp(b)	Felony Exp(b)
Independent Variables			
MJPS Risk Score (0 to 7)	1.43 ***	1.40 ***	1.89 ***
Male (0,1)	1.81 *	1.77 *	1.65
White (0,1)	0.36 **	0.52	0.85
Model Fit			
Nagelkerke R Square	0.13	0.10	0.17

Note: *p≤.05, **p ≤.01, ***p ≤.001

Receiver Operating Characteristic (ROC). Four tables were created based on area under the curve (AUC) values derived from ROC analysis to analyze how well the MJPS performed predicting recidivism compared to the Pre-screen BOT.

Table 3.8 presents the results for the full sample. The full MJPS sample consisted of 422 youth and the full Pre-screen BOT sample consisted of 272 youth. For each recidivism category, the MJPS outperformed the Pre-screen BOT. The MJPS had an AUC of 0.66 for *any recidivism,* placing it in the moderate effect size compared to the Pre-screen BOT with 0.60, placing it in the small effect size. The MJPS predicted *misdemeanor/felony recidivism* similarly with a 0.65 AUC, compared to a 0.61 for the Pre-screen BOT. Finally, the MJPS predicted *felony recidivism* with a large effect size of 0.76 compared to a moderate effect size for the Pre-screen BOT with 0.67.

Table 3.8 Full Sample Accuracy Results (AUCs)

Type of Recidivism	MJPS (*N*=422)	Pre-Screen (*N*= 272)	AUC Improvement
Any	0.66	0.60	0.06
Misdemeanor or Felony	0.65	0.61	0.04
Felony	0.76	0.67	0.09

MJPS findings presented here reflect those from the screener's retrospective study. Table 3.9 presents the full sample accuracy results utilizing retrospective data from between 2010 and 2015 that was partially used to create the MJPS. The retrospective sample was much larger, with data from 1,395 youth used to analyze predictive accuracy of both the MJPS and the Pre-screen BOT. As in the pilot test, the MJPS in the retrospective study outperformed the Pre-screen BOT for each recidivism category. Importantly, there was not a significant drop in AUCs between the sample used to create the MJPS and the 2018 pilot test data. Pilot test data performed comparably when predicting *any recidivism* with an AUC of 0.66 compared to 0.69 for 2010 to 2015 data. The largest drop in AUCs between the 2018 pilot test data and the earlier dataset was for the prediction of *misdemeanor/felony recidivism*. The pilot test had an AUC of 0.65, compared to an AUC of 0.71 for the 2010 to 2015 data. Finally, pilot test results demonstrated that the MJPS performed better at predicting *felony recidivism* than the MJPS did with 2010 to 2015 data, with an AUC of 0.76 compared to 0.70.

Table 3.9 Full Sample Accuracy Results 2010-2015 Data (AUCs)

Type of Recidivism	MJPS (*N*=1395)	Pre-Screen (*N*= 1395)	AUC Improvement
Any	0.69	0.64	0.05
Misdemeanor or Felony	0.71	0.63	0.08
Felony	0.70	0.67	0.03

A drop in accuracy is expected between a retrospective study and a prospective one. This phenomenon is referred to as "shrinkage" (see Copas, 1983). According to Gottfredson and Snyder (2005 p. 27) "A smaller amount of shrinkage might give greater confidence that the validity of the prediction method, and the classification procedure derived from it, will hold up on repeated applications." Original data will inherently have a better model fit. Sample size will also affect AUC results. It is promising, however, that the shrinkage was not significant between the pilot test and the 2010 to 2015 data. This provided evidence that MJPS prediction accuracy is relatively stable between the MJPS retrospective test and the prospective analysis.

The following analyses break the full sample down into subsamples, including sex and race. Each subsample is comprised of small sample sizes, making accuracy measurements more variable. The following analyses, therefore, must be cautiously evaluated. Certain subsamples had so few cases that they had to be removed from analysis.

Table 3.10 presents AUC results for the male and female only samples for both the MJPS and the Pre-screen BOT. Sample sizes are displayed in the bottom of the table. Similar to the results above, the MJPS outperformed the Pre-screen BOT for both the male and female samples. The MJPS predicted female *felony recidivism* with the greatest accuracy (0.79) within the large effect size range. The MJPS predicted *any recidivism* and *misdemeanor/felony recidivism* for females with a

moderate effect size of 0.70, falling close to the threshold of 0.71 for a large effect size. Similar to the female sample, the MJPS predicted male *felony recidivism* with the greatest accuracy. The MJPS predicted *felony recidivism* for the male sample with an AUC of 0.76 within the large effect size range. *Misdemeanor/felony recidivism* and *any recidivism* for the male only sample are both in the moderate effect size range (0.64 and 0.65 respectively).

Table 3.10: Sex Accuracy Results (AUCs)

Type of Recidivism	MJPS	Pre-Screen	AUC Improvement
Any			
Male	0.65	0.60	0.06
Female	0.70	0.64	0.06
Misdemeanor/Felony			
Male	0.64	0.60	0.04
Female	0.70	0.63	0.06
Felony			
Male	0.76	0.68	0.08
Female	0.79	0.70	0.09
Sample Size			
Male	272	169	
Female	156	97	

Table 3.11 presents results of the race subsamples. Subsamples were broken down into white, American Indian, and non-white non-American Indian. Sample sizes are shown at the bottom of Table 3.11. As found in the previous analyses, AUC results indicated the MJPS outperformed the Pre-screen BOT on all subsamples of race and recidivism types. The MJPS predicted *any recidivism* for white youth within a moderate effect size, with a 0.67 AUC, compared to the Pre-screen's AUC of 0.64. MJPS predictions of *any recidivism* for American Indian youth were not as accurate, compared to white youth or non-white non-American Indian youth. For the American Indian sample, the MJPS slightly outperformed the Pre-screen with an AUC of 0.56, compared to 0.55 for the Pre-screen. An AUC of 0.56 is in the small effect-size range and 0.55 is in the negligible range. MJPS prediction accuracy for *any recidivism* for non-white non-American Indian youth was similar for white youth and significantly better than the Pre-screen BOT. The MJPS predicted *any recidivism* for non-white non-American Indian youth with an AUC of 0.66, within the moderate effect-size range, compared to 0.45 for the Pre-screen BOT. MJPS prediction results for *misdemeanor/felony recidivism* were almost identical to those from *any recidivism*. The MJPS predicted *misdemeanor/felony recidivism* for white youth with a 0.67, the Pre-screen garnered a 0.65. The MJPS demonstrated a significant drop in prediction accuracy for *misdemeanor/felony recidivism* with American Indian youth, with the MJPS showing an AUC of 0.52 compared to the Pre-screen BOT's 0.46. These AUC's demonstrate performance that is no better than chance. The MJPS predicted *misdemeanor/felony recidivism* for non-white non-American Indian youth with an AUC of 0.61, significantly outperforming the Pre-screen BOT with an AUC of 0.38. Finally, the MJPS predicted *felony recidivism* for white youth with a large effect size, an AUC of 0.76, compared to the moderate effect size demonstrated by the Pre-screen, with an AUC of 0.67. American Indian youth and non-white, non-American Indian AUCs for f*elony recidivism* are not presented due to the small sample sizes.

Results presented for the American Indian sample demonstrate low recidivism prediction accuracy. Because American Indian youth were underrepresented in the MJPS pilot test, however, indicating that their small sample was not randomly selected, results presented here likely do not reflect the MJPS's capacity to predict recidivism for this population. For this reason, analysis of MJPS prediction accuracy among these smaller subsamples should be ongoing.

Table 3.11 Race Accuracy Results (AUCs)

Type of Recidivism	MJPS	Pre-Screen	AUC Improvement
Any			
White	0.67	0.64	0.03
American Indian	0.56	0.55	0.01
Non-White, Non-AI	0.66	0.45	0.21
Misdemeanor/Felony			
White	0.67	0.65	0.02
American Indian	0.52	0.46	0.06
Non-White, Non-AI	0.61	0.38	0.23
Felony			
White	0.76	0.67	0.09
American Indian	N/A	N/A	-
Non-White, Non-AI	N/A	N/A	-
Sample Size			
White	389	232	
American Indian	19	15	
Non-White, Non-AI	13	19	

As discussed in the override analysis above, officer discretion in tandem with the MJPS demonstrated increased prediction performance. Based on this finding, MJPS administration protocol may be improved by allowing officers to slightly adjust the risk score if they wish override the total risk score.

To test the above hypothesis and adjust the risk score for officer discretion, one point was added to the youth's total MJPS risk score when there was an override up and one point was subtracted from the MJPS total score when there was an override down.[5]

- Officer Agrees with MJPS indicated decision = Original MJPS Risk Score
- Officer believes youth is higher risk than MJPS indicated = MJPS Risk Score + 1
- Office believes youth is lower risk than the MJPS indicated = MJPS Risk Score - 1

ROC analysis was then used to compare the accuracy of predicting *misdemeanor/felony recidivism* with the original MJPS risk score in contrast with the new score including a one-point increase or decrease. Results are illustrated in Table 3.12. For the full sample and all sub-samples, prediction accuracy increases when officer judgement is considered.

[5] This analysis was completed after the pilot test of the MJPS to demonstrate one possible strategy for incorporating officer discretion. Probation officers had the ability to make overriding decisions but did not have the option to manipulate the MJPS score during the pilot test.

Table 3.12 Improving MJPS Accuracy Using Officer Judgment

Sample	MJPS	MJPS with Officer Judgement	AUC Improvement
Full (*N*=428)	0.65	0.68	0.03
Sex			
Male (*n*=272)	0.64	0.66	0.02
Female (*n*=156)	0.70	0.71	0.02
Race			
White (*n*=383)	0.67	0.69	0.02
American Indian (*n*=19)	0.52	0.55	0.03
Non-White, Non-AI (*n*=26)	0.61	0.61	0.01
Initial Offense			
Status/Tech (*n* = 99)	0.75	0.75	0.01
Misdemeanor (*n*=315)	0.63	0.66	0.03
Felony (*n*=13)	0.69	0.66	-0.03

The accuracy does not drastically change for any sample, but the exercise does demonstrate a possible strategy to incorporate officer discretion. As discussed in the first section of this report, officers believed that youth were lower risk if they had protective or mitigating factors that the MJPS does not ask about. Similarly, officers believed that youth were higher risk if the youth had additional risk factors or there was an aggravating circumstance not weighed by the MJPS. This new scoring technique could allow officers to adjust the final score if they detect a mitigating or aggravating factor not collected during MJPS administration and could further increase buy-in from officers in the field.

Discussion

Overall, the predictive validity results discussed here largely mirror those found during the retrospective analysis of the MJPS (see McKay et al., 2017). The MJPS has consistently demonstrated similar, if not, increased accuracy for predicting recidivism over the Pre-screen BOT. Higher risk scores are empirically shown to increase recidivism, MJPS risk scores are statistically significant predictors of recidivism according to logistic regression, and AUC results from ROC analysis are consistent for a tool with a moderate to large effect size at predicting recidivism. Shrinkage is low between the results found during the retrospective analysis and the current prospective analysis. Overall, the MJPS has maintained prediction accuracy while used in the field and the accuracy has been now been validated on two separate samples (retrospective and prospective). The following section includes a discussion on limitations to the analysis presented in this report and suggestions for future analysis.

CONCLUSION

This report details the first prospective validation study of the Montana Juvenile Probation Screener (MJPS). The MJPS was designed to serve as a brief and accurate risk assessment that would increase fidelity compared to more time-consuming instruments. Probation officers offered their recommendations to improve the Pre-screen BOT and the MJPS already addresses several shortcomings. For example, the screener is significantly shorter, with risk factors that are not overly intrusive. Brevity allows for easier training and understanding of the instrument, which, in turn, will increase inter-rater reliability. The MJPS was built upon data collected in Montana and has established itself as an accurate risk predictor during a retrospective test with comparable results during this prospective analysis. The Criminology Research Group believes these improvements, among others, will increase instrument fidelity, which can be a significant barrier for the successful administration of any assessment.

Six districts in Montana piloted the MJPS, administering it to 444 youth from July 2018 through the end of January 2019. Results documented here are promising. Face validity, for the MJPS was high. Juvenile probation officers agreed with the predicted risk level 92% of the time and reported—via survey—confidence in the screener's ability to predict recidivism. Officers documented a higher degree of confidence and agreed with the MJPS indicated risk level to a greater extent than they did for the current risk assessment utilized, the Pre-screen BOT. During the pilot test, the MJPS successfully located low-risk youth, most of whom were given the least restrictive disposition. Such identification and application of disposition satisfies a primary goal associated with MJPS introduction. Further, prediction validation analyses demonstrated that the MJPS is comparable and at times more accurate than the Pre-screen BOT when predicting recidivism. Overall, this study supports MJPS validity and supports MJPS use as a brief tool to determine recidivism risk. The findings, however, do have limitations and continued research is recommended.

Limitations. The low MJPS administration rate constitutes this study's greatest limitation. Of the total 1,408 youth going through intake at pilot sites during the pilot test period, 32% (444) were given the screener. While the sample size itself is sufficient, the non-representative nature of the sample may impact prediction accuracy analyses. Factors affecting the likelihood of whether a youth received the screener included what district the young person was in, offense severity, and race. American Indian youth were the most underrepresented in the sample, with only 17% receiving the screener, compared to 35% of white youth. The MJPS performed most poorly when predicting recidivism among American Indian youth. It is likely those results would differ if all or a significantly greater proportion of American Indian youth were sampled. In addition, recidivism rates for MJPS screened youth were not representative of the larger population. In general, youth administered the MJPS were less likely to have a recidivating offense. Low fidelity to a risk assessment is not a novel problem. In fact, the same issue arose in the Montana Pre-screen BOT validation analysis (see Mckay et al., 2015). That analysis and this study show a need to increase buy-in among probation officers to encourage risk assessment instrument consistency and fidelity.

It was beneficial for the accuracy analysis discussed earlier in this report to compare MJPS results to those from the Pre-screen BOT. The authors of this report were able to ask, "If the MJPS was to take the place of the Pre-screen BOT, would prediction accuracy remain the same, worsen, or improve?" Such comparisons were limited, however, by the fact that not all youth administered

the MJPS also received the Pre-screen BOT. Of the 444 young people administered the MJPS, more than one-third (35%; 156) were not given the Pre-screen BOT.

A final limitation—inherent to all non-randomize pilot studies—is the inability to generalize findings from the pilot sites to the larger population. The MJPS pilot test occurred in six of Montana's 22 judicial districts. Findings from this analysis, therefore, can only be generalized to districts participating in this study. If the MJPS is introduced to the remaining Montana Youth Court Services judicial districts, further validation analyses should occur to ensure findings are consistent across geography.

Future Research. The investigations detailed in this report opened several areas of inquiry for future research. First, new inquiries should examine ways to reduce limitations inherent to this study. For example, the use of risk assessments in Montana could be improved through an examination of strategies for increasing probation officer buy-in and instrument fidelity. Further work is needed to make the argument that the MJPS and BOT have value beyond the data being collected. The MJPS or Pre-screen should assist in difficult decisions to ensure equitable treatment. The Full-Screen BOT should be used to place youth in services that match need. Without a persuasive argument for instrument use, MJPS buy-in will continue to be low. Additionally, future work should examine why race and offense type affect the likelihood of a young person being administered a risk assessment. To maximize risk assessment effectiveness, all youth—regardless of location, gender, or race—should be given the assessment.

Information gathered from the Montana Juvenile Probation Survey showed that certain MJPS risk factors may be difficult to collect or ambiguously worded. Before implementing the MJPS state-wide, the risk factor definitions should be further refined to decrease subjectivity which will increase inter-rater reliability and confidence in administrating the screener. Moreover, automating the MJPS in JCATS would likely reduce errors discovered during the administration.

If the MJPS does replace the Pre-screen BOT, future inquiries should examine how the MJPS and Full-BOT can work together to maximize outcomes. It is currently recommended that youth who receive a moderate or high score on the Pre-screen BOT receive the Full-BOT. This could be a starting place for the MJPS, but thresholds should be re-examined to ensure youth receiving the Full-BOT are those with the most need.

Finally, this investigation uncovered a trend that warrants future inquiry. Justice-involved American Indian youth were almost twice as likely to have a recidivating offense compared to their white counterparts. Montana would benefit from future studies examining why American Indian youth are at an increased risk of recidivism and what can be done to reduce that disparity. Risk assessments are intended to reduce disparities between white and non-white youth through an objective and fair evaluation of recidivism potential. They also have the potential to assist in service decisions. Future MJPS analysis should ensure those objectives are being met.

REFERENCES

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

Babbie, E. R. (2015). *The practice of social research*. Nelson Education.

Baglivio, M. T. (2009). The assessment of risk to recidivate among a juvenile offending population. *Journal of Criminal Justice*, *37*(6), 596-607.

Baglivio, M. T., & Jackowski, K. (2013). Examining the validity of a juvenile offending risk assessment instrument across gender and race/ethnicity. *Youth Violence and Juvenile Justice*, *11*(1), 26-43.

Bishop, M. A., & Trout, J. D. (2002). 50 years of successful predictive modeling should be enough: Lessons for philosophy of science. Philosophy of Science, 69(S3), S197-S208.

Carns, T. W., & Martin, S. (2011). Does the YLS/CMI help to predict recidivism? An assessment of the Division of Juvenile Justice's use of the Youth Level of Services/Case Management Inventory. Alaska Judicial Council and the Institute for Social and Economic Research.

Conley, T. B., Spurzem, K., Marsh, E., & Hazlett, J. Juvenile offenders in Montana.

Copas, J. B. (1983). Regression, prediction and shrinkage. *Journal of the Royal Statistical Society: Series B (Methodological)*, 45(3), 311-335.

Girard, L., & Wormith, J. S. (2004). The predictive validity of the Level of Service Inventory-Ontario Revision on general and violent recidivism among various offender groups. Criminal justice and behavior, 31(2), 150-181.

Gottfredson, D. M., & Snyder, H. N. (2005). The Mathematics of Risk Classification: Changing Data into Valid Instruments for Juvenile Courts. NCJ 209158. Office of Juvenile Justice and Delinquency Prevention.

Hilton, N. Z., Harris, G. T., & Rice, M. E. (2006). Sixty-six years of research on the clinical versus actuarial prediction of violence. The Counseling Psychologist, 34(3), 400-409.

Litwin, M. S. (1995). How to measure survey reliability and validity (Vol. 7). Sage publications.

McCafferty, J. T. (2017). Professional discretion and the predictive validity of a juvenile risk assessment instrument: Exploring the overlooked principle of effective correctional classification. Youth Violence and Juvenile Justice, 15(2), 103-118.

McKay, P., Hollist, D., Bunch, J., Acton, D., Tillman, T., & Harris, C. (2015). The Montana Back on Track Risk Assessment Instrument: An Assessment and Validation Study.

McKay, P., Hollist, D., Ranger., M. (2017). Montana Back on Track (BOT): Developing a Risk Assessment Screener and Utilizing the BOT to Assist in Service Placement.

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

Vaswani, N., & Merone, L. (2014). Are there risks with risk assessment? A study of the predictive accuracy of the Youth Level of Service–Case Management Inventory with young offenders in Scotland. The British Journal of Social Work, 44(8), 2163-2181.

Wilcox, D., Beech, A., Markall, H. F., & Blacker, J. (2009). Actuarial risk assessment and recidivism in a sample of UK intellectually disabled sexual offenders. Journal of Sexual Aggression, 15(1), 97-106.

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

Wormith, J. S., Hogg, S., & Guzzo, L. (2012). The predictive validity of a general risk/needs assessment inventory on sexual offender recidivism and an exploration of the professional override. Criminal Justice and Behavior, 39(12), 1511-1538.

APPENDIX A:

Montana Juvenile Probation Screener Guidelines

Screening Process

- Piloting of the screening tool is tentatively scheduled to last six months., commencing July, 2018.
- The screener shall be administered in the paper form provided to the pilot sites.
- During the data collection phase of the pilot project, the screener shall be administered on all youth subject to an intake/preliminary inquiry from the piloting district.
- The screener shall be administered in person by the assigned worker. It is not to be completed by the youth or legal guardian.
- Piloting districts are still required to comply with current OCA policies governing the administration of the BOT Assessment Instruments. During this phase of the pilot project, the screener shall not be used as the sole source for decision making and case planning.
- Remember the purpose of this phase of the project is to collect information/data to assist in determining the screeners practical application in the field. Your work will be used to vet the screener for reliability, validity, and fidelity

Collection Process

- Each person administering the screener will be assigned a *Rater ID Number*. Your Rater ID Number will be your Judicial District Number (01,05,11,13,19,21) combined with your User ID (your CU or CI number) **Ex. 054081**.
- Please note that the U of M Research Group will never report on individual officers or associate officer names with these IDs. This ID # along with the youths JCATS # and the Intake Date must be included on every screener administered.
- Once the Screener has been completed, a copy shall be provided to the Chief JPO for that piloting district.
- The CJPO shall be responsible for collection and transmission of the completed screener to the U of M Criminology Research Group. Chiefs, please ensure that the Rater ID, JCATS# and Intake Date are included on every screener before submitting them to the U of M CRG.
- Completed screeners are to be scanned as a PDF file and emailed to Patrick McKay - patrick.mckay@mso.umt.edu
- Completed screeners shall be submitted to Patrick bi-monthly (every two weeks)

Questions and Feedback

- Specific questions about the screener should be addressed to Patrick McKay with the U of M CRG.
- As the piloting sites, we not only encourage feedback but are expecting you to provide a constructive critic of the screener. Please provide input and feedback.

APPENDIX B:

Age at First Offense, Past Misdemeanor and Felony Offenses

PROBING CRIMINAL HISTORY Why don't we start by talking about your experience with the justice system, since that's the main reason we're meeting. Can you tell me about the different times you've had trouble with the law, starting with the very first time, and we'll map out what that looks like together?	 *Use SUMMARY*
RISK FACTORS:	**ITEM CHECKLIST**
1.Age at first offense: Age at the time of the offense for which the youth was referred to juvenile court for the first time on a no-traffic offense resulting in an admission, conviction, diversion, deferred adjudication, or deferred disposition. Does not include status offense. If this is the youth's first offense put 0. How to collect: JCATS Data Interview with youth and guardian. How to score on Risk Assessment: - 13 or older or first offense = 0 - Under 13 = 1	Age of first offense
2.Misdemeanor Referrals: Total number of referrals for which the most serious offense was a no-traffic misdemeanor that resulted in an admission, a conviction, diversion, deferred adjudication, or deferred disposition (regardless of whether the deferred final adjudication was successfully completed). Multiple offenses occurring within a 24-hour period are to be considered a single referral. Do not include status offenses. How to collect: JCATS data How to score on Risk Assessment: - One or zero total prior misdemeanors = 0 - More than one total misdemeanor = 1	# Misdemeanors
3. Felony Referrals: Total number of referrals for which the most serious offense was a felony that resulted in an admission, a conviction, diversion, deferred adjudication, or deferred disposition (regardless of whether the deferred final adjudication was successfully completed). Multiple offenses occurring within a 24-hour period are to be considered a single referral. (not including current offense) How to collect: JCATS data How to score on Risk Assessment: - Zero prior felonies = 0 - One or more felonies = 1	# Felonies

Positive Adult Role Models and Current Friends

LEAD-IN What adults, other than family members, have really been there for you or appeared to be supportive of you in the past? What was enjoyable about spending time with them?	**REMINDERS**
NETWORK PROBE I would like to get a better sense of what kind of friendships and the people you spend time with.	*SUMMARY*
RISK FACTORS **4. History of positive adult non-family relationships not connected to school or employment:** Included adults who are not teachers and not part of youth's family and can provide support and model pro-social behavior, such as religious leaders, club members, mentors, coaches etc. (a teacher can be counted in this category if they are a mentor, coach or group leader outside of the school). How to collect: Ask youth/family/school about extra-curricular activities, mentors, participation in religious groups. How to score on Risk Assessment: - Positive adult = 0 - No Positive adult = 1 **5. Current Friends/Companions youth actually spend time with** How to collect: Ask youth/family/school who their closest friends are. Search JCATs data for friends / did the current offense take place with friends? How to score on Risk Assessment: - No current anti-social friends = 0 - Current anti-social friends = 1	**ITEM CHECKLIST** **Prior relationships:** # Adult positive relationships # Anti-social friends

History of Running Away

PAST FAMILY / LEAD-IN / PROBES	REMINDERS
Tell me about the family that you have lived with for the longest period in your life. Who lives there? And who else? How do you all get along together? How many times have you left (runaway) or had to be removed? Have any of your family members have ever been in jail or prison?	
RISK FACTORS: **6**. **History of Running away or getting kicked out of the home:** Times the youth did not voluntarily return within 24 hours and incidents not reported by, or to, law enforcement. How to collect: JCATS data. Ask youth/family about running away or kicked out of home history. How to score on Risk Assessment: - No history of running away or getting kicked out of home = 0 - History of running away or getting kicked out of home = 1	**ITEM CHECKLIST** # Runaways

Belief in Fighting

PROBE LEAD-IN	**REMINDERS**
What does it look like when you manage your anger the way you want to and what's it look like when your anger is hard to manage? Tell me about a time when you have lost your temper. What kind of things might cause you to lose your temper? When have you had to resort to violence? In what kind of circumstances are you most likely to use violence? What is the worst violence you have seen or been involved in? When is hitting and or yelling at someone a good idea? How do you feel after you hit or yell at someone?	*MI may be unnecessary and unhelpful in this section*
RISK FACTORS **7. Belief in fighting and physical aggression to resolve a disagreement or conflict** How to collect: JCATS data- any history or current violent offenses. Ask youth/family/school about the youth's belief in fighting and physical aggression. How to score on Risk Assessment: - No evidence of fighting to resolve a disagreement or conflict = 0 - Evidence youth fights or believes in fighting to resolve conflict = 1	**ITEM CHECKLIST** Verbal Aggressive Beliefs Physical Aggressive Beliefs History of Violence

Problem Solving

PROBES	**REMINDERS**
I would like to talk to you about the kinds of skills we acquire throughout our life in order to become a more disciplined person; sometimes we refer to these as self-control skills. In your experience, what are some goal-setting and problem-solving skills? How are these skill sets related? What are some goals that you have set and achieved? Tell me more about that. What are some early signs that you are beginning to have a problem with some person, place or thing? When you experience strong emotions like anger, fear or depression – what kinds of things do you do to manage and work through these feelings? Where would you like to be in, say, four years? What would you like your life to look like and where would you want your self-control skills to be then?	*Use ELABORATION TECHNIQUES, e.g., "Tell me more about that," "'What else," "Give me an example of that," and follow-up with REFLECTIONS*
RISK FACTORS: **8. Problem solving:** Youth can/cannot identify problem behaviors and does not think of solutions or apply solutions to problem behavior. How to collect: Ask youth how their current offense could have been resolved in a manner that was not illegal. How to score on Risk Assessment: - Can problem solve = 0 - Cannot problem solve = 1	**ITEM CHECKLIST** Consequential thinking Goal setting Problem-setting Situational perception Social skills Handling conflict Dealing with feelings Dealing with Internal triggers Dealing with External triggers Impulse control Aggression control

APPENDIX C:

Training Vignette
Marcie

Intake Report JCATS:

Marcie's first involvement with the juvenile court occurred when she was thirteen years old. School officials caught her stealing items out of an unlocked car with two friends in the alley next to the junior high school. JCATS does not show any past offenses.

Officer Discussion with Marcie and her Parents/guardians:

Disciplinary records show that she regularly cut classes, often skipping school altogether, and twice had been involved in physical altercations with other students.

Marcie struggled with school; in particular, she had difficulty learning to read. Diagnosed with dyslexia and attention-deficit/hyperactivity disorder, she was on an individualized learning plan and was reading below a fifth-grade level at the time of her arrest.

Marcie was abandoned by an alcoholic mother who was fleeing an abusive relationship with Marcie's biological father. At that time, she entered the foster care system where she had previously run-away multiple times. She has been receiving medical care for stomach pains. The doctor diagnosed her with stomach ulcers, thought to be the result of stress and worry. These physical problems worsened, and she often missed school due to chronic stomach pain. Marcie's foster parents involved her in several extra-curricular activities where she excelled in swimming and has a great relationship with her coach.

Marcie admitted to the theft and showed no remorse for her actions. When asked what she could have done differently in the situation; she explained that she could have run faster from school officials.

APPENDIX D:

Montana Juvenile Probation Screener

Rater ID #: 0 0 0 0 Youth JCATS #: 1111 Intake Date: 06 / 12 / 2018

1. Age at First Offense *MM / DD / YYYY*

This is youth's first offense.....0
13 or older.....0 + 0
Under 13.....1

2. Misdemeanor Referrals

One or zero total prior referrals..... 0 + 0
More than 1 referral..... 1

3. Felony Referrals

No prior referrals..... 0 + 0
One or more..... 1

4. Positive Adult Role Model

Yes..... 0 + 0
No..... 1

5. Current Friends

No anti-social friends..... 0 + 1
Current anti-social friends..... 1

6. History of Running Away / Kicked out

No..... 0 + 1
Yes..... 1

7. Belief in Fighting

No..... 0 + 1
Yes..... 1

8. Problem Solving

Yes..... 0 + 1
No..... 1

Total Score..... = 4

Screener Indicated Risk:	0-2 Low Risk	3-4 Medium Risk	5+ High Risk
Probability of Recidivism:	*20%*	*45%*	*66%*

Practitioner's Judgement (*check*): ☐ Low Risk ☒ Medium Risk ☐ High Risk

If practitioner's judgement does not match indicated risk please briefly explain why risk should be higher/lower:

_____
_____
_____

Disposition:

☐ Formal Probation ☒ Informal Probation ☐ Diversion ☐ Dismissed

APPENDIX E:

Training Vignette:

Risk Assessment Explanation:

Youth's First Offense under 13: 0 → First involvement with juvenile court

Misdemeanor Referral: 0 → No past offenses JCATS

Felony Referral: 0 → No past offenses JCATS

Positive Role Model: 0 → Relationship with coach

Current Friends: 1 → Caught with friends

History of Running Away: 1 → Ran away from foster care

Belief fighting: 1 → Involved twice in physical altercations

Problem Solving: 1 → Solving problem behavior by running from school officials

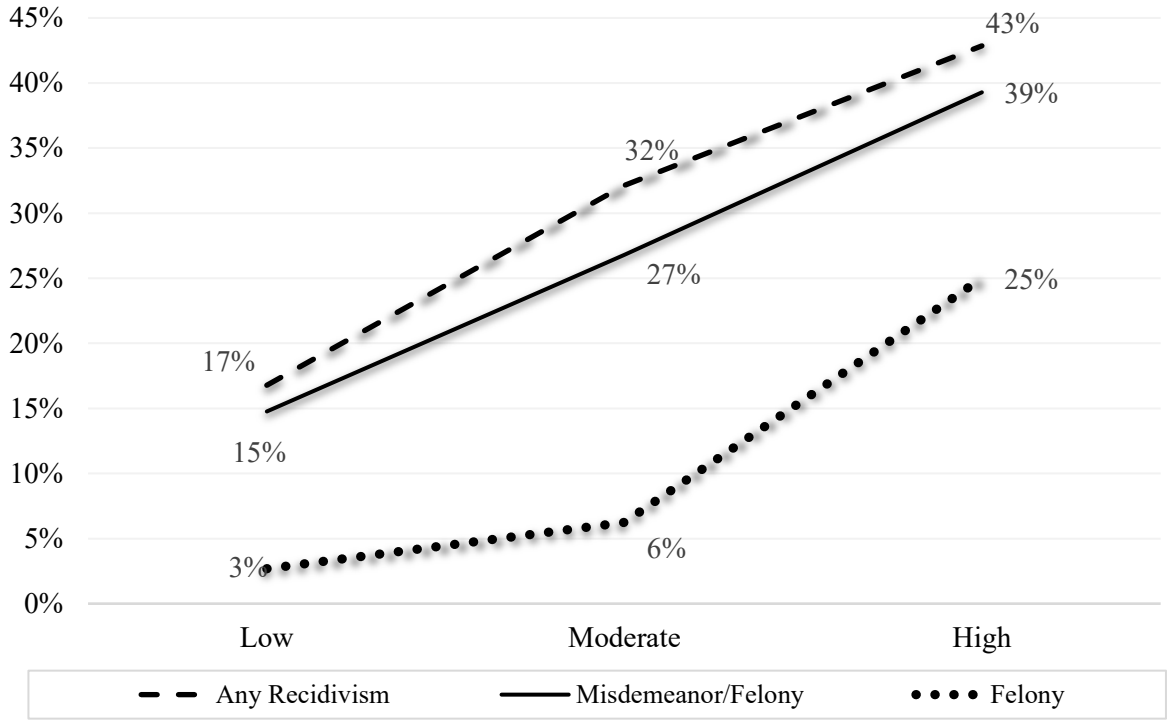
Total Score: 4 = Medium Risk

APPENDIX F:

Recidivism Rate by MJPS Risk Level: Full Sample (*N*=444)

Risk Level	*n*	Any Recidivism	Misdemeanor/Felony	Felony
Low	298	17%	15%	3%
Moderate	112	32%	27%	6%
High	28	43%	39%	25%
Recidivism Base Rate	438	22%	19%	5%

APPENDIX G:



Recidivism rates for low, moderate, and high risk on the MJPS.