Iowa Delinquency Assessment (IDA)

Validation, Assessment, & Recommendations

Technical Report

Washington State University

DRAFT for JJAC– Not for release or redistribution

Zachary Hamilton, Ph.D.
Alex Kigerl, Ph.D.
Melissa, Kowalski, MA

Department of Criminal Justice and Criminology
Washington State Institute for Criminal Justice (WSICJ)

May 15, 2018

Introduction

The utilization of assessment for risk and needs of youth has been at the core of juvenile justice reform and operational practice nationally, and in Iowa, for nearly two decades. Launched in 2007, and adapted from the Washington State assessment, the Iowa Delinquency Assessment (IDA) and its use has historically been considered a leading effort in juvenile risk/needs assessment development. The tool itself consists of two parts, a short form (SF), with just over two dozen criminal and social history items, is used to establish a youth's risk level (low moderate, or high). This allows low-risk youth to be identified/selected for diversion. Moderate and high-risk youth are then also administered the long form (LF), where IDA domains establish levels of need in several areas, including education, employment, free time, relationships, family, alcohol/drug use, mental health, attitudes/behaviors, aggression, and skills.

In 2012, Barnoski completed an evaluation of the IDA and provided several recommendations. First, findings indicated that the SF possessed predictive validity but that prediction could be improved with revised risk scoring. Specifically, it was recommended that Iowa:

- Implement revised scoring (utilized IA specific data);
- Create a 4th, 'high violent' risk category;
- Determine if items from SF or LF can be removed due to lack of: prevalence, relationship to recidivism, case management utility; and
- Add items from the LF that are helpful in predicting recidivism.

Furthermore, based on recent developments with the assessment's use in other states (i.e. Washington, Delaware, and Florida), additional advancements in assessment were identified for consideration. In particular, the use of 1) statistical algorithms to select and weight assessment items that are predictive for the Iowa population, 2) development of *gender specific* assessments, and 3) the creation of multiple assessment models (i.e. felony, violent, property, and drug) to predict a youth's most likely type of recidivism were considered to improve predictive accuracy and greater utility of the tools' reports. Finally, as recent media concerns have suggested the potential for assessments to perpetuate racial and ethnic biases of the juvenile justice system, there was a need to assess the biases within the current scoring of the tool and any potential changes with recommended updates.

In 2016 Hamilton, Kigerl and Kowalski began an examination of the IDA. The examination outlined three major aims. First, create an updated version of the SF, exploring its potential though the three advancements, mentioned previously. Next, this process was repeated for those youth that had completed the LF, making use of the larger pool of assessment items.

Methods

To complete the study aims, we worked with the Division of Criminal and Justice Planning, assembling samples of youth who had completed the SF, LF, and recidivism information. Recidivism was defined as a new charge occurring within 12 months of the initial assessment. A total sample of 13,412 eligible youth were gathered, consisting of 10,571 males and 2,841 females. To analyze the predictive validity of the current and updated versions of the IDA, we used the industry standard Receiver Operating Characters (ROC) analysis, which produces a statistic, the Area Under the Curve (AUC), which evaluates youths' risk assessment scores. Essentially, the AUC rates the accuracy that youth with higher scores are observed to recidivate with greater frequency than those with lower scores and that those with lower scores do not recidivate with a greater frequency than those with higher risk scores. First, we assessed the predictive validity of the current risk assessment. The AUC ranges from 0.5 to 1.0 and can be viewed as a percentage of accuracy. Industry-standards have outlined ranges of AUC scores that indicate magnitude (effect sizes) of prediction strength and are indicated in Table 1.

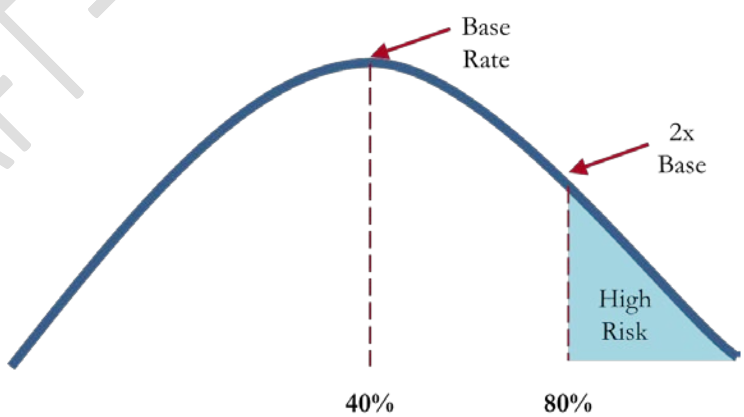
Table 1. AUC industry-standard effect size ranges (see Rice & Harris, 2005)

<.55	Negligible
>.55	Small
>.63	Moderate
>.71	Strong

To create current IDA scoring, the criminal and social history scores, gathered from the SF, were summed. To create the updated SF and LF models, items were included into multiple statistical algorithms, including LASSO, LARS, Elastic Net, Boosted, and Stepwise. This 'batch algorithm' process was developed by WSICJ researchers to ensure that predictive items are identified. All selected items were then placed into a final (Boosted regression) algorithm, where item weights (i.e. values) are created. An industry-standard M-fold validation process was then used to compute model AUCs. This process was completed multiple times, predicting recidivism outcomes for 'any' (felony or misdemeanor), felony, violent, property, and drug recidivism. In addition, all models were created separately, once for males and again for females. The culmination of these processes produced 10 SF and 10 LF models.

Next risk level categories (RLCs) were compared between the current IDA and the updated SF and LF models. While still considered preliminary, the common method of applying cut points, or scoring thresholds, was implemented. Specifically, the base rate of recidivism for each outcome was calculated. The base rate represents the average/rate of recidivism for each outcome for the sample of youth. For instance, the base rate for 'any' recidivism is roughly 40% for youth assessed with the IDA. For each model, youth's scores provide a point estimate for the probability of recidivating. For each model, high-risk cut points were set at twice the base rate; where male and female scores possessing an 80% (or greater) likelihood of recidivating were identified as 'high risk'.

Figure 1. Illustration of RLC cut point placement process



This process was replicated for each of the felony, violent, property, and drug models. Finally, half the base rate for the 'any' recidivism model was the method used to create a the low-risk cut point; where anyone not identified as high or low-risk are classified as moderate-risk.

Results

With regard to the updated SF, our findings indicated that a more accurate prediction can be created by selecting and weighting items, by gender, using local, Iowa data. Specifically, when using the described methods, two items were found to not be predictive within all models - misdemeanor and felony sex offenses. While all other items were predictive in at least one model type, these two items could potentially be removed from the SF. When comparing the AUC values, the updated SF models provided an average improvement in AUC of 5% (and more specifically, 4% for male and 6% for females). While seemingly a modest improvement, keep in mind that the scale only ranges from 50-100%; therefore a 5% improvement is substantial. In total, using Iowa specific data to make adjustments on the SF, with a considerably smaller pool of items and responses, still made substantial improvements.

Table 2. AUC Values Comparing Current to Updated SF scoring

Model	Male	Female
Current IDA	*0.58*	*0.55*
Any	*0.63*	*0.58*
Felony	0.60	0.63
Violent	0.67	0.68
Property	0.61	0.60
Drug	0.59	0.56

Next we set the cut points for the updated SF. Again, we used the 'twice the base rate' method for high risk, half the base rate for low-risk, and all others were identified as moderate-risk. With five models, we identified high risk as anyone identified as high-risk in one or more models. Table 3 provides the proportion of the population identified in each category and each category's associated recidivism percentage.

It should be stressed that the updated SF cut points are still considered preliminary but, despite this caution, there are some truly promising results. First, when examining males, one can observe a greater proportion of males identified as high-risk in the updated SF and this group is indicated to have a greater proportion of recidivism. This is paired with a relatively equal proportion of low-risk males, however the updated SF provide a *decreased* rate of recidivism. This indicates a greater accuracy of the cut-point placement procedure and the underlying selection/weighted scoring.

For females, the biggest improvement is the reduction of high-risk and the increase in the low risk proportion. In fact, the proportion of females shifted to low-risk nearly doubled. The resulting recidivism proportions indicate a slight increase in the high-risk group recidivating, with only a slight up-tick in the low-risk recidivism rate. Overall, these findings indicate encouraging findings, where not only are the AUC values demonstrating increased accuracy but the updated SF cut points indicating improved prediction and a reduction in the proportion of over-classified female youth.

Table 3. RLC classifications of Current IDA and updated SF

	Pop %	Recid %	Pop %	Recid %
Updated 3 Cat. SF	**Male**		**Female**	
High	37.0	56.3	17.2	45.4
Moderate	32.5	49.0	42.8	39.1
Low	30.5	33.5	32.4	29.7
Current IDA				
High	28.1	53.7	25.8	42.0
Moderate	41.2	49.7	56.5	36.8
Low	30.7	37.7	17.8	25.8

Following the analyses of the SF, we then explored the larger pool of items and responses of the LF. The same model development and cut point placement methods were utilized. These findings were even more encouraging. First, we compared the AUC values of the current IDA with that of the updated LF models. The findings are presented in Table 4.

As anticipated, the improvements in accuracy observed in the updated SF are further expanded with the updated LF. On average, the AUC values increased by 9% (and more specifically, have an average of 8.2% and 9.8% improvement for males and females, respectively). This is a substantial improvement, with the female violent model reaching the industry-standard level of 'strong' predictive accuracy.

Table 4. AUC Values Comparing Current to Updated LF scoring

Model	**Male**	**Female**
Current IDA	*0.58*	*0.55*
Any	*0.65*	*0.62*
Felony	0.65	0.64
Violent	0.69	0.71
Property	0.64	0.60
Drug	0.68	0.67

Next we set the cut points for the updated LF. Again, we used the 'twice the base rate' method for high risk, half the base rate for low-risk, and all others were identified as moderate-risk. With five models, we identified high risk as anyone identified as high-risk in one or more models. Table 5 provides the proportion of the population identified in each category and each category's associated recidivism percentage.

First, when examining the males, one can observe a reduced proportion of males identified as high-risk in the updated LF. This is paired with an increased proportion of low-risk males. What is notable, is that the greater proportion of low-risk males is found to commit a *reduced* rate of recidivism and the greater proportion of high-risk males is indicated to have a *higher* rate of recidivism. This indicates an even greater accuracy of the cut-point placement procedure and the underlying selection/weighted scoring.

For females, we observe a dramatically reduced proportion identified as high-risk and an increased proportion of low-risk youth in the updated LF. Again, it is notable that the greater proportion of low-risk females is found to commit a near equal rate of recidivism and the greater proportion of high-risk females is indicated to have a *higher* rate of recidivism. Once again, these findings indicate an even greater accuracy of the cut-point placement procedure and the underlying selection/weighted scoring.

Table 5. RLC classifications of Current IDA and updated LF

	Pop %	**Recid %**	**Pop %**	**Recid %**
Updated 3 Cat	**Male**		**Female**	
High	15.2	60.2	7.9	52.1
Moderate	44.6	53.8	51.1	37.9
Low	40.2	29.4	40.1	25.9
Current IDA				
High	28.1	53.7	25.8	42.0
Moderate	41.2	49.7	56.5	36.8
Low	30.7	37.7	17.8	25.8

Finally, based on the Barnoski's recommendations, we added a fourth, 'high-violent' category to our RLC analysis. This category was created by using the violent risk model and setting the cut point at twice the base rate of violent recidivism. The remaining high-risk category retains youth from the other models (i.e. high-risk for non-violent recidivism). Here we compared the updated LF with and RLC assignment that divided the high-risk category into violent and non-violent. In addition to examining 'any' recidivism, we also provided a rate of violent recidivism. Findings reveal that a fourth, high-violent category improves violent risk prediction. When examining the rate of violent recidivism, the fourth category identifies a small proportion of high-violent youth – roughly 5% of males and 3% of females. However, the violent recidivism rate for the fourth category is greater and there is a substantial increase in recidivism observed for male youth, where a 36% violent recidivism rate was identified for high-violent males as compared to only 25% for the three-category model.

Table 6. Three versus Four Category RLC

	Pop %	Recid %	Vio Recid%	Pop %	Recid %	Vio Recid%
Updated 3 Cat	**Male**			**Female**		
High	15.2	60.2	25.2	7.9	52.1	25.0
Moderate	44.6	53.8	18.3	51.1	37.9	13.3
Low	40.2	29.4	7.4	40.1	26.9	5.1
Updated 4 Cat						
High Violent	4.5	58.0	36.4	2.5	40.0	26.7
High Non-Violent	10.7	59.7	20.6	5.4	57.6	24.2
Moderate	44.6	53.8	18.3	51.1	37.9	13.3
Low	40.2	29.4	7.4	40.1	26.9	5.1

Conclusions

Since its development in Washington State, and its implementation in Iowa, the IDA has remained relatively unchanged. The items, original response weights, and utilization of the SF and LF were retained in their original form. As recommended by Barnoski (2004; 2010) and the creators of the Risk, Need, Responsivity model (Andrews & Bonta, 1995; 2010), best practices indicate that risk and needs assessments should be updated and adjusted to suit local agency needs. Specifically, in his last evaluation of the IDA, Barnoski suggested that less predictive items from the LF be removed and strongly predictive items weighted to improve model accuracy.

The current study attempted to do much of what prior recommendations indicated as best practices. Specifically, we used advanced statistical algorithms to explore reduced, weighted, and improved SF and LF tools. Cut points were then established to compare and contrast updated models with the current scoring and RLCs of the IDA.

Findings revealed substantial improvements when comparing the current IDA scoring and RLCs to that of updated models. Furthermore, LF models provided greater improvements than that of updated SF. As indicated in similar developments in Washington, Delaware, and Florida, these improvements are a result of increased accuracy of statistical modeling, gender and outcome specific modeling, as well as the use of local data. While the models are still considered preliminary, current work with a Subject Matter Expert (SME) group is moving this work forward and tailoring it for Iowa youth needs and juvenile justice policy.

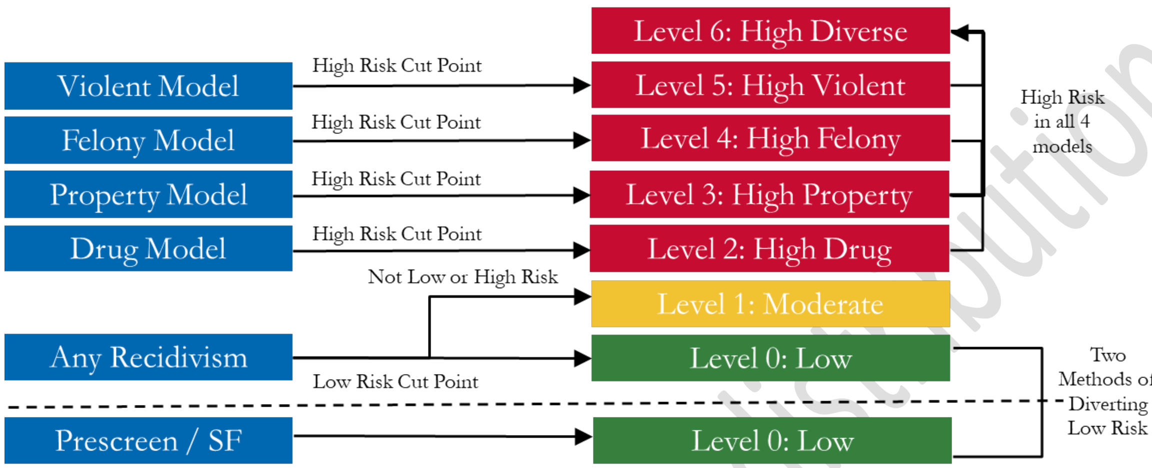
Recommendations

Based on the findings, several recommendations are proposed for future implementation of the current analyses.

1. The IDA can be reduced in length by removing items that are not predictive of recidivism, which will, in turn, reduce assessment labor.
2. IDA predictive accuracy can be improved with the proposed gender and outcome specific models.
3. Utilize the updated SF to identify low risk youth eligible for diversion.
4. Utilize the updated LF to identify a more accurate and outcome specific report of risk categorization.

As a way of further explaining the use of multiple tools and outcomes, Figure 2 provides an illustration of the proposed process. Essentially, while updated, the SF would be used as it is currently; where those identified as low-risk would be eligible for diversion. Those not identified as low-risk will be administered the LF, where youth's risk will be more accurately calibrated in to one of several models of risk.

Figure 2. Proposed IDA RLC Process



As is currently being implemented in both Washington and Delaware, a prioritization of risk levels is important to create a balance of both predictive accuracy and public safety. In particular, there is a hierarchy of outcome type as it pertains to public safety and societal costs. The proposed ranking allows case managers to identify a youth's most likely recidivism type. This type of model construction will further improve targeting of needs and adoption of responsive, Evidence-Based Practices to further reduce recidivism and youth involvement in the justice system.