



Research and Evaluation

Validating the Probability of Escalation from County Probation to the Oregon Youth Authority Model

Research conducted by:
OYA Research and Evaluation

The mission of the
Oregon Youth
Authority
is to protect the
public and reduce
crime by holding
youth offenders
accountable and
providing
opportunities for
reformation
in safe environments

May 2023

Introduction

In 2016, under the direction of the Oregon Juvenile Justice Director's Association (OJJDA), the Research and Evaluation Unit of the Oregon Youth Authority (OYA) developed a model to predict the probability that a youth placed on County Probation (CP) would escalate to OYA. The model was implemented in Juvenile Justice Information System (JJIS) and County Juvenile Departments were trained by OYA implementation staff on how to use the estimates. The Theoretical notion was that youth with higher likelihood to escalate would be given more services. In addition, the models indicated what the important factors were that predicted escalation and suggested what types of services might be provided. As part of research units' protocol, periodic validation of the model constructs produced by the unit are necessary to assure that they continue to predict accurately and are used as intended. The purpose of the study is to assess the predictive validity of the overall escalation model. Further evaluation of the use of the tool will be conducted as a follow up to this analysis.

Research Questions

Does the overall escalation model predict escalation from County probation to OYA?

- Does the measure predict escalation for both males and females?
- Does the measure predict escalation for all OYA Race/ethnicity groups?

Method

Assessment Model

The original model was based on information from Juvenile Crime Prevention (JCP) assessment (JCP assessments only were included if the assessment occurred any time prior to placement on CP), OYA Recidivism Risk Assessment for a Violent crime (ORRA-V), and age. (see Appendix A for model parameters). The model calculated probability of escalation from County Probation to OYA (PECO) scores for each youth on CP that met inclusion criteria present below.

Data and Participants

Data for these analyses were extracted from JJIS via currently published reports in the form of excel extracts. Two Business Information Systems (BIS) reports were included: (a) Escalation Universe; and (b) ORRA Event Universe.

Participants for the current validation study included all youth with dispositions to CP from 1/1/2014 through 12/31/2017 (n=6,018). Youth were excluded if they did not have a PECO score. PECO scores were calculated from the logistic regression equation in Appendix A. Youth were only included if they were under 18 years of age, had a JCP, ORRA-V, and birthdate in the Juvenile Justice Information System (JJIS). Finally, youth were randomly selected to ensure that youth with multiple dispositions were only included one time. After exclusions, the sample for

this validation included 4,762 youth. A demographic breakdown of the sample is provided in Appendix B.

Diagnostics and Analysis

Demographics are presented in Appendix B. Descriptive statistics of PECO scores were applied by sex, race, and age and included frequencies, means, standard deviations, minimums, and maximums (see Appendix C)—a histogram of the frequency distribution also is provided for the overall sample of PECO scores. In addition, distributions of PECO scores were presented in Boxplots for the overall sample and by sex, race/ethnicity, and age (see Appendix D). Predictive analytic diagnostics included calibration (i.e., comparing actual and expected escalation rates) by demographics, risk level, risk level and race/ethnicity, and risk level and sex (Appendix E); confusion matrixes estimate the overall accuracy, precision, sensitivity, and specificity of the model (Appendix F); and Chi-square (Appendix G) and Area under the Receiver Operator Curve (AUC; Appendix H) estimate the extent to which PECO scores discriminates high risk from low risk youth.

Results

Demographics

The demographic makeup of the validation sample (Appendix B) indicates that the sample is close to what we would expect from a random sample drawn from the population of youth on county probation. Nearly 75%, male; about 60% white; and ages ranging from 11 to 17 (youth in the sample that were over age 17 were excluded from the sample). Males are typically overrepresented in juvenile criminal justice systems so 75% male is no surprise. Although not unexpected, at 9%, African American youth were the only minority that was seriously overrepresented—Oregon African American youth were 4% of Oregon’s 10-17 age group (Puzzanchera, Sladky, & Kang, 2021). And age groups were not unusual, the number of youths in each age group gets larger up to age 16 and drops slightly at 17.

Descriptive Statistics

Descriptive statistics are presented in Appendix C. An overall mean of 21.9% (PECO scores are presented in percentages for most the document; however, some of the Appendix may include PECO scores in whole numbers) is what we expect from PECO scores—almost identical to the 21.8% overall mean in the development sample. Scores ranged from near 0 to 97% with a standard deviation of 18%. Males (24%) had a much higher likelihood they would escalate than females (15%). Hispanics had the highest likelihood of any Race/Ethnicity group, and Other/Unknow had the lowest. Age at disposition PECO scores were expected with the highest scores with the youngest groups—recall age is negatively associated with escalation in the equation (see Appendix A).

The histogram in Appendix C plots the frequency distribution of PECO scores. There is nothing unusual about the distribution. The number of youths at each level rises swiftly, peaking at about 4%, and begin to decline steadily to almost 60%. Over 200 youth score above 60% and about 8% of the youth fall above 50%. Half of the scores fall below the mean at 22%, so the rest of the scores fall between 22% and 97%.

Boxplot graphs (see Appendix D) exhibit the spread and locality of data in a simple form that provides a clear view of how scores are distributed. The overall sample is presented in the first chart in Appendix D: The lowest scores fall just above 0%; the box (interquartile range) indicates that half of the scores fall just below 10% and just above 30%—and 75% of the scores falls below about 31%; the median line (in the center of the box) indicates that half the scores were above 17% and half the scores were below 17; outliers (extreme scores) start at the top line just above 65% and continue up to 97%.

In addition, boxplots provide a means for comparing the distributions across demographics. The second chart in Appendix D indicates that the distribution of male scores are more spread out and noticeably higher than female scores; the third chart indicates that most of the distribution of scores Race/Ethnicity category don't differ nearly as much as sex, except for the Other/Unknow category which has less spread and lower scores than the other categories; the last chart in Appendix D indicates that although the spread of scores is similar across age categories, scores do get lower as age goes up.

Calibration

Appendix E presents the calibrations for PECO scores in terms of percent differences between actual and expected escalation rates by demographic variables and risk level.¹ In the first table, the overall escalation rate was 8% higher than expected—the actual rate was 24% and the expected rate was 22%. By sex, females had a higher than expected rate (almost 20% higher) than males (5% higher). By race/ethnicity, African Americans (28% higher) and Native Americans (27% higher) had much higher expected rates than Hispanics (7% higher) and Whites (5% higher). By age at disposition, both the youngest (4% lower) and oldest groups (30% lower) than all other age groups. By risk level, the high-risk group² escalated 12 less than expected and the low-risk group escalated 14% higher than expected.

The second table in Appendix E present sex and race/ethnicity by risk level. Essentially, except for Native Americans, this follows the same pattern as the overall risk-level—higher risk groups had lower than expected escalation rates and lower risk groups had higher than expected escalation rates. Native Americans had higher than expected rates for both lower and higher

¹ For the most part, Other/Unknown Race/Ethnicity category was ignored in the report because there was no way of knowing what Race/Ethnicity the youth were; however, they are presented in the tables.

² The high-risk group was defined as a probability of escalating above 50%.

risk youth. Other/Unknowns and Asians should be interpreted with caution as only 7 Asian and 5 Other/Unknown youth were in the high-risk groups.

Confusion matrix are designed to measure model performance, or, in this case, how well do PECO scores classify risk levels? Four metrics were computed: Overall accuracy, precision, sensitivity, and specificity. Each of these metrics provide an estimate of validity that indicates the strength of the extent to which PECO scores predicts the likelihood of escalation. The cut-point for predicting escalation was a PECO scores above 50%. Appendix F presents matrix and a diagram that illustrates how the metrics were calculated.

Another flaw in the development of this model is the fact that the JCP at the time had large amounts of missing data. This was because the response options at the time included an option (more information needed) that could not be measured. Any data that was scored “more information needed” was considered missing and those data were removed from the sample. Also, if there was any missing data on any youth they were removed from the dataset. Only about 70% of the entire sample was used because about 30% was missing. For the past 6 years, this response option has not been available so this problem will no longer exist in the future.

The reduction in AUC from its development sample (.77) to this validation sample (.73) also is a concern. Although AUC=.73 is within industry standard (Silver & Banks, 1998; Silver, Smith & Banks, 2000), there is still reason for alarm. Future development with data that doesn't have large amounts of missing might help solve this problem. Adding a tracking period might also help. And, making sure that the data is divided into learn and test sample so the final model can be tested on a new set of data. Finally, using bootstrap sample to develop the model also would help to improve/stabilize the AUC.

Conclusions/Recommendations

Again, this study provided plenty of evidence to conclude that PECO consistently predicted escalation from county probation to OYA; across both sex and race/ethnicity. However, several recommendations are warranted for a more complete evaluation of the tool.

First, following Messick's (1989a, 1989b) suggested validating guidelines, making sure the tool is being interpreted as intended will be important in a more comprehensive review of PECO's validity. Analysis about what users think the score means will be important in determining if they understand what the purpose of the tool is. This might require some face-to-face interviews/focus groups to gather the information necessary for the analysis.

Second, the interviews/focus groups might also be designed to determine if the tool has value. For example, does it improve decision making or is it just wasting time?

Third, this recommended investigation will also need to gather information that indicates whether the tool is relevant or not. For example, does it inform decisions about the youth they are trying to serve?

Fourth, the analysis needs to determine what the social consequences are for both staff and youth. For example, have there been any reductions in escalation where the tool is being implemented appropriately?

Fifth, prior to adhering to Messick's guidelines, other concerns indicate possible revisions of PECO to improve stability and performance: (a) testing different tracking periods could improve the AUC and make the age variable more interpretable (as constructed, age may be a proxy for days of opportunity); (b) because there was so much missing data in the development of the tool, building the new tool with the JCP revision will considerably improve PECO's performance; (c) bootstrapping samples for the development of the model also will improve the performance; and (d) separating the sample into train and test samples will improve the stability of the AUC.

Although these recommendations will require some resources from both the research and implementation units to complete, they will be very useful improving the tools usefulness and predictably.

References

- Anastasi, A. (1986). Evolving concepts of test validation. *Annual Reviews of Psychology*, 37, 1-15.
- Berk, R., Heidari, H., Jabbari, S., Kearns, M., & Roth A. (2021). Fairness in Criminal Justice Risk Assessments: The State of the Art. *Sociological Methods and Research*, 50(1), 3-44.
- Chouldechova, A. (2017). "Fair Prediction with Disparate Impact: A Study of Bias in Recidivism Prediction Instruments." arXiv:1703.00056v1 [stat.AP].
- Messick, S. (1989a). Meaning and values in test validation: The science and ethics of assessment. *Educational Researcher*, 18(2), 5–11.
- Messick, S. (1989b). Validity. In R. Linn (Ed.), *Educational measurement* (3rd ed., pp. 13–103). Washington, DC: American Council on Education/Macmillan
- Puzzanchera, C., Sladky, A. and Kang, W. (2021). "Easy Access to Juvenile Populations: 1990-2020." Online. Available: <https://www.ojjdp.gov/ojstatbb/ezapop/>
- Silver, E., Smith, W.R., & Banks, S. (2000). Constructing actuarial devices for predicting recidivism. *Criminal Justice and Behavior*, 27(6), 733-764.
- Silver, E. & Banks, S. (1998). Calibrating the potency of violence risk classification models: The Dispersion Index for Risk (DIFR). Paper presented at the 1998 American Society of Criminology Conference, Washington, DC.

Appendix A: Model Parameters

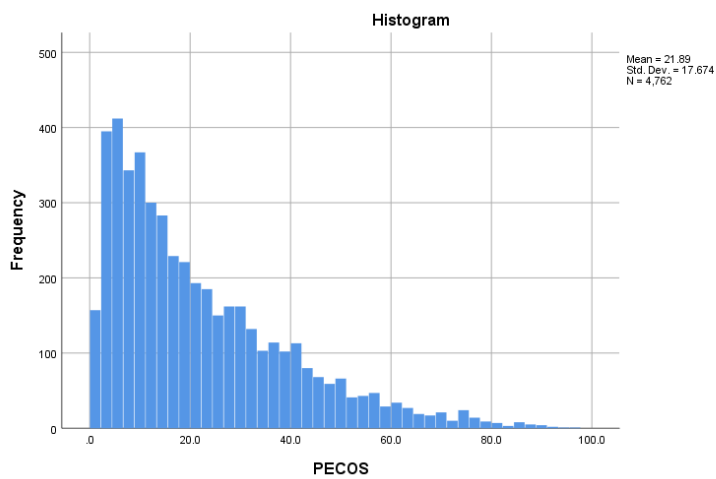
Logistic Regression: Parameter Estimates for Modeling Escalation - Main Effect Variables Listed by Order of Entry (N = 8,454; AUC=.77)				
PREDICTOR VARIABLES*	VALUES	PARAMETER ESTIMATE	P	ODDS RATIO
Main Effects				
ORRA Violent Risk Score	Decimal between 0 & 1	0.105	.000	1.110
JCP_7.1_Anti_social_thinking_attitudes_values_beliefs	No = 0, Yes = 1	0.398	.000	1.489
Committed for a Sex Offense	No = 0, Yes = 1	1.562	.000	4.767
JCP_4.7_Recent_runaway	No = 0, Yes = 1	0.689	.000	1.992
Age at Disposition	Age	-0.560	.000	.571
JCP_4.3_Three_or_more_referrals	No = 0, Yes = 1	0.936	.000	2.549
Severity	min=-5, max=19	-0.281	.001	.755
JCP_2.1_School_attachment	No = 0, Yes = 1	-0.448	.000	.639
JCP_6.1_Substance_use_beyond_experimental_use	No = 0, Yes = 1	0.308	.000	1.360
JCP_4.1_Chronic_aggressive_disruptive_behavior	No = 0, Yes = 1	0.250	.001	1.284
JCP_3.5_Substance_abusing_friends	No = 0, Yes = 1	0.383	.000	1.467
JCP_5.10_Has_close_positive_supportive_relationship	No = 0, Yes = 1	-0.276	.001	.759
Interactions				
ORRA Violent Risk Score X JCP_4.3_Three_or_more_referrals		-0.051	.000	.950
JCP_2.1_School_attachment X JCP_4.1_Chronic_aggressive_disruptive_behavior		0.269	.035	1.309
Severity X Age at Disposition		0.023	.000	1.023
Intercept	Constant	4.625	.000	101.972

Appendix B: Validation Sample Demographics

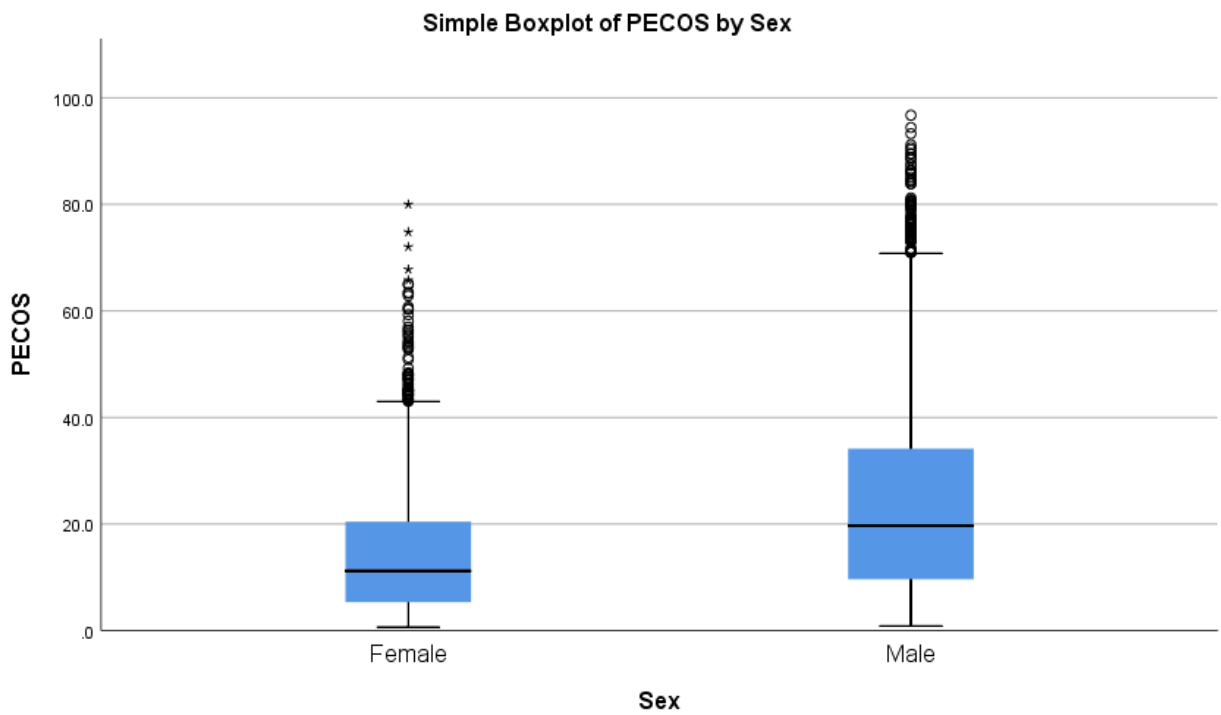
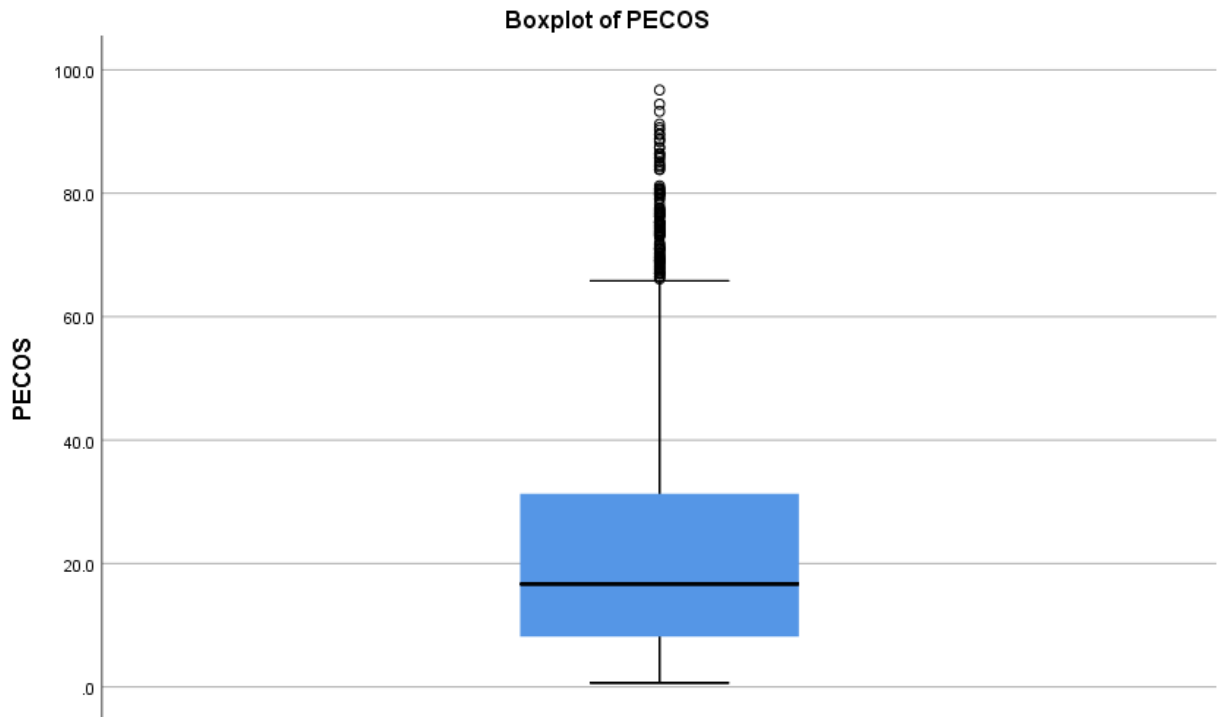
Sex	#	%
Female	1220	25.6%
Male	3542	74.4%
Race/Ethnicity	#	%
African American	423	8.9%
Asian	84	1.8%
Hispanic	959	20.1%
Native American	191	4.0%
Other/Unknown	209	4.4%
White	2896	60.8%
Age at Disposition	#	%
11	8	0.2%
12	144	3.0%
13	375	7.9%
14	727	15.3%
15	1066	22.4%
16	1293	27.2%
17	1149	24.1%
Total	4762	100.0%

Appendix C: PECO Descriptive Statistics by Demographics

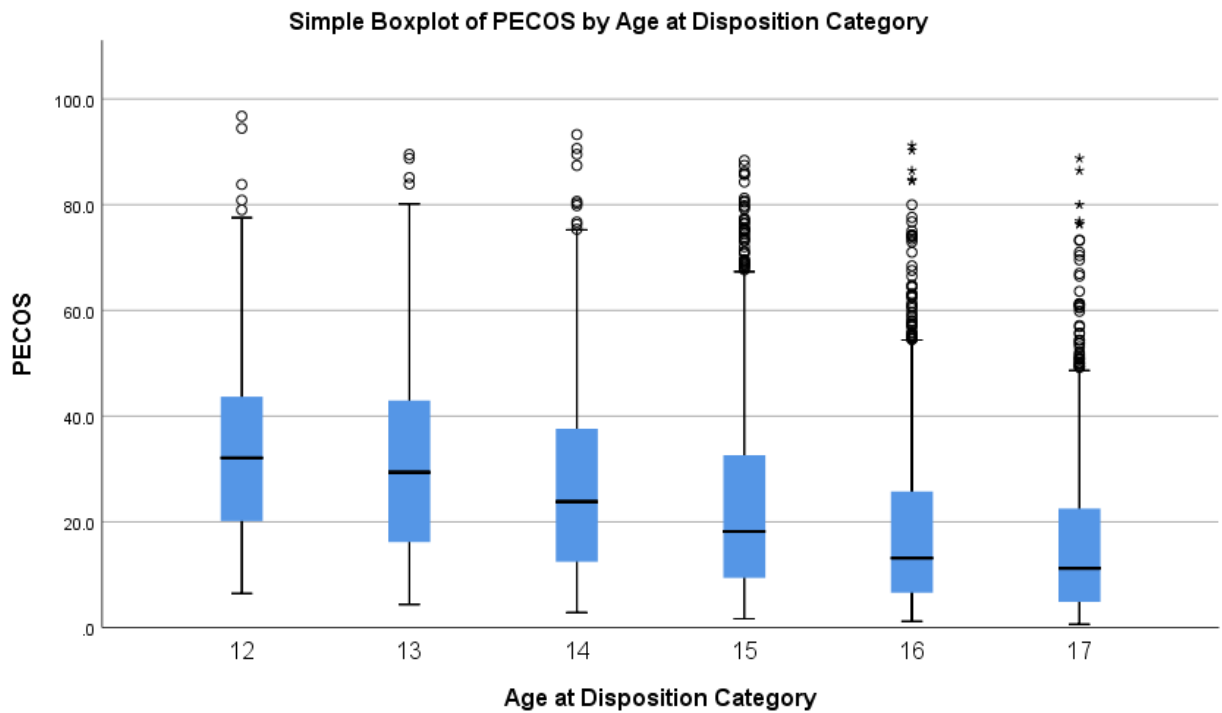
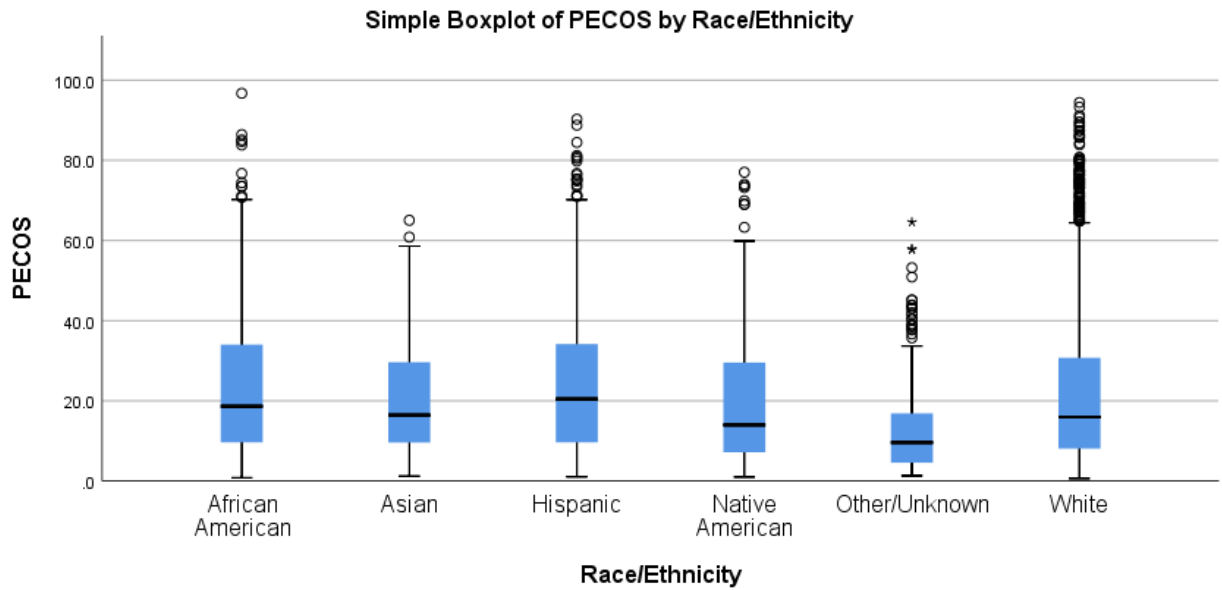
All	N	Minimum	Maximum	Mean	Std. Deviation
	4762	0.6%	96.7%	21.9%	17.7%
Sex					
Female	1220	0.6%	80.0%	15.2%	13.4%
Male	3542	0.9%	96.7%	24.2%	18.4%
Race/Ethnicity					
African American	423	0.8%	96.7%	23.6%	18.2%
Asian	84	1.2%	65.1%	21.2%	15.9%
Hispanic	959	1.0%	90.3%	24.2%	17.9%
Native American	191	1.0%	77.0%	21.0%	18.7%
Other/Unknown	209	1.3%	64.6%	13.9%	13.0%
White	2896	0.6%	94.4%	21.5%	17.6%
Age at Disposition*					
12	152	6.5%	96.7%	33.7%	18.6%
13	375	4.4%	89.5%	31.6%	18.7%
14	727	2.9%	93.3%	27.2%	17.8%
15	1066	1.7%	88.4%	23.6%	18.5%
16	1293	1.2%	91.2%	18.6%	15.9%
17	1149	0.6%	88.8%	16.0%	14.7%
* Age 12 included 8 youth 11 years old					



Appendix D: Distribution of Probability of Escalation from County Probation to OYA Scores



Appendix D Continued: Distribution of Probability of Escalation from County Probation to OYA Scores



Appendix E: Calibration—Actual vs Expected PECO

	Actual	Expected	% Difference	N
Total	23.6%	21.9%	7.7%	4762
Sex				
Female	18.2%	15.2%	19.6%	1220
Male	25.4%	24.2%	5.1%	3542
Race/Ethnicity				
African American	30.3%	23.6%	28.3%	423
Asian	25.0%	21.2%	18.0%	84
Hispanic	26.0%	24.2%	7.3%	959
Native American	26.7%	21.0%	27.4%	191
Other/Unknown	9.1%	13.9%	-34.4%	209
White	22.6%	21.5%	4.9%	2896
Age at Disposition Category				
12	32.2%	33.7%	-4.2%	152
13	38.1%	31.6%	20.8%	375
14	33.8%	27.2%	24.6%	727
15	27.1%	23.6%	15.0%	1066
16	20.6%	18.6%	11.0%	1293
17	11.1%	16.0%	-30.2%	1149
Risk Level *				
High	55.4%	63.2%	-12.2%	395
Low	20.7%	18.2%	13.9%	4367

* The high risk cut point was 50% probability that a youth would escalate

Appendix E Continued: Calibration—Actual vs Expected PECO

Sex	Risk Level	Actual	Expected	% Difference	N
Female	High	53.3%	59.7%	-10.7%	30
	Low	17.3%	14.1%	22.8%	1190
Male	High	55.6%	63.5%	-12.3%	365
	Low	21.9%	19.7%	11.5%	3177
Race/Ethnicity					
African American	High	60.5%	65.1%	-7.0%	38
	Low	27.3%	19.5%	39.9%	385
Asian	High	57.1%	57.5%	-0.6%	7
	Low	22.1%	17.9%	23.5%	77
Hispanic	High	57.6%	62.0%	-7.1%	92
	Low	22.6%	20.2%	12.0%	867
Native American	High	68.2%	60.9%	12.0%	22
	Low	21.3%	15.8%	35.1%	169
Other/Unknown	High	0.0%	56.9%	-100.0%	5
	Low	9.3%	12.8%	-27.2%	204
White	High	53.7%	63.9%	-15.9%	231
	Low	19.9%	17.9%	11.3%	2665

Appendix F: Confusion Matrix

Confusion Matrix: All			
Confusion Matrix: All Observed/Actual	Predicted/Expected		Total
	No	Yes	
No	3464	176	3640
Yes	903	219	1122
Total	4367	395	4762

Overall Accuracy	77%
Precision	20%
Sensitivity	55%
Specificity	79%

Calculations

Observed/Actual	Predicted/Expected	
	No	Yes
No	tn	fn
Yes	fp	tp

Overall Accuracy: $(tp+tn)/(tp+fp+tn+fn)$
Precision: $tp/(tp+fp)$
Sensitivity: $tp/(tp+fn)$
Specificity: $tn/(tn+fp)$

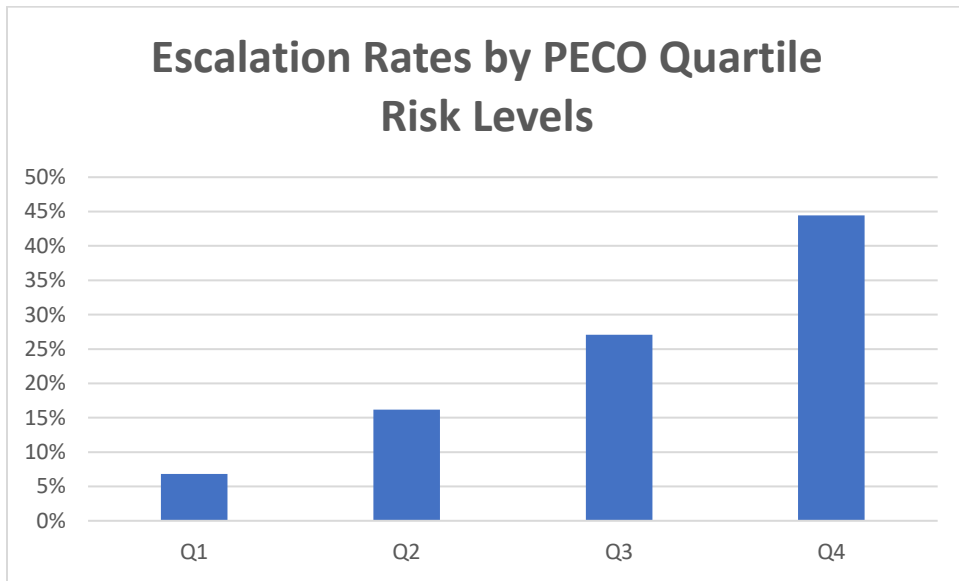
Confusion Matrix—Metrics by Sex & Race Ethnicity

Performance Metric	Performance Metrics								
	Overall	Sex		Race/Ethnicity					
	All	Female	Male	African American	Asian	Hispanic	Native American	Other Unknown	White
Overall Accuracy	77%	82%	76%	72%	76%	75%	77%	89%	78%
Precision	20%	7%	23%	18%	19%	21%	29%	0%	19%
Sensitivity	55%	53%	56%	61%	57%	58%	68%	0%	54%
Specificity	79%	83%	78%	73%	78%	77%	79%	91%	80%

Appendix F: Confusion Matrix—Sex & Race/Ethnicity

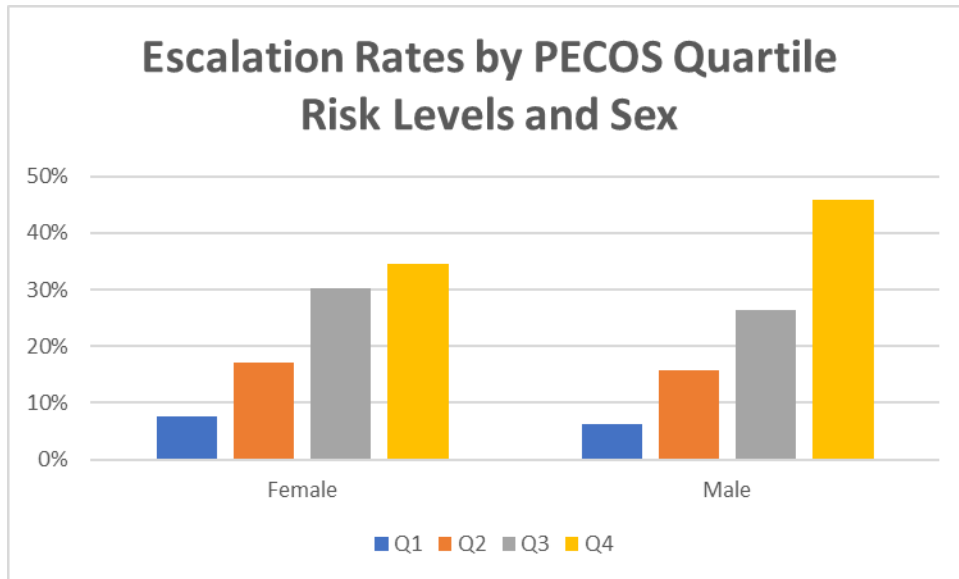
Female		Predicted/Expected		
Observed/Actual	No	Yes	Total	
No	984	14	998	
Yes	206	16	222	
Total	1190	30	1220	
Male		Predicted/Expected		
Observed/Actual	No	Yes	Total	
No	2480	162	2642	
Yes	697	203	900	
Total	3177	365	3542	
African American		Predicted/Expected		
Observed/Actual	No	Yes	Total	
No	280	15	295	
Yes	105	23	128	
Total	385	38	423	
Asian		Predicted/Expected		
Observed/Actual	No	Yes	Total	
No	60	3	63	
Yes	17	4	21	
Total	77	7	84	
Hispanic		Predicted/Expected		
Observed/Actual	No	Yes	Total	
No	671	39	710	
Yes	196	53	249	
Total	867	92	959	
Native American		Predicted/Expected		
Observed/Actual	No	Yes	Total	
No	133	7	140	
Yes	36	15	51	
Total	169	22	191	
Other/Unknown		Predicted/Expected		
Observed/Actual	No	Yes	Total	
No	185	5	190	
Yes	19	0	19	
Total	204	5	209	
White		Predicted/Expected		
Observed/Actual	No	Yes	Total	
No	2135	107	2242	
Yes	530	124	654	
Total	2665	231	2896	

Appendix G: Chi-Square



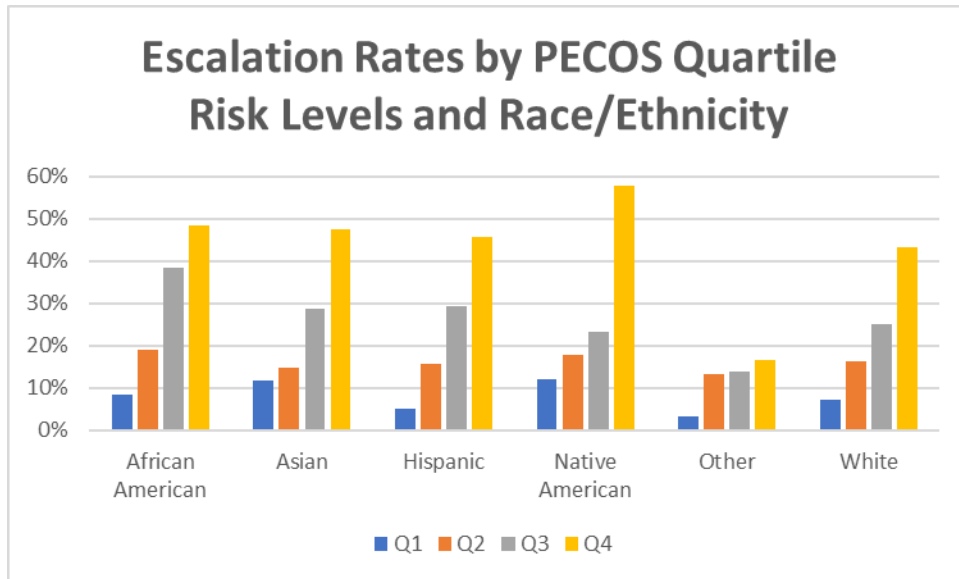
Chi-Square Tests			
	Value	df	p
Pearson Chi-Square	514.367 ^a	3	0.000
N of Valid Cases	4762		

Appendix G Continued: Chi-Square by Sex



Chi-Square Tests		Value	df	p
Sex				
Female	Pearson Chi-Square	85.427 ^a	3	0.000
	N of Valid Cases	1220		
Male	Pearson Chi-Square	407.614 ^b	3	0.000
	N of Valid Cases	3542		

Appendix G Continued: Chi-Square by Race/Ethnicity



Chi-Square Tests		Value	df	p
African American	Pearson Chi-Square	46.957 ^a	3	0.000
	N of Valid Cases	423		
Asian	Pearson Chi-Square	8.295 ^b	3	0.040
	N of Valid Cases	84		
Hispanic	Pearson Chi-Square	115.246 ^c	3	0.000
	N of Valid Cases	959		
Native American	Pearson Chi-Square	30.641 ^d	3	0.000
	N of Valid Cases	191		
Other/Unknown	Pearson Chi-Square	7.684 ^e	3	0.053
	N of Valid Cases	209		
White	Pearson Chi-Square	292.039 ^f	3	0.000
	N of Valid Cases	2896		

Appendix H: Area Under the Receiver Operator Curve (AUC)

	AUC	p	Confidence Level	
			Lower Bound	Upper Bound
Total	0.732	0.000	0.715	0.748
Sex				
Female	0.705	0.000	0.669	0.740
Male	0.735	0.000	0.716	0.753
Race/Ethnicity				
African American	0.716	0.000	0.664	0.767
Asian	0.738	0.001	0.616	0.861
Hispanic	0.745	0.000	0.711	0.778
Native American	0.748	0.000	0.668	0.827
Other/Unknown	0.682	0.009	0.573	0.791
White	0.724	0.000	0.703	0.746
Age at Disposition				
12	0.724	0.000	0.637	0.811
13	0.708	0.000	0.653	0.762
14	0.654	0.000	0.612	0.695
15	0.696	0.000	0.660	0.731
16	0.740	0.000	0.708	0.772
17	0.744	0.000	0.702	0.786

Joseph O'Leary
Director

Carlos Rodriguez
Deputy Director



Oregon Youth Authority
530 Center St. NE, Suite 200
Salem, OR 97301-3765

503-555-5555
oya.info@oya.state.or.us
www.oregon.gov/oya