DCYF Juvenile Rehabilitation Risk Assessment Validation Study

Juvenile Rehabilitation Risk Assessment Validation Study: RAR, RAI and RACF

As part of the May 2020 Integrated Treatment Model Assessment legislative report, it became clear that work was required to better understand the validity of some of the risk assessments currently being used in Juvenile Rehabilitation (JR). For further details, please see report Recommendation 2: Create an Accountability Structure for Risk and Needs Assessment, where it is stated, "The fact that JR continues to use risk assessment tools that have not been validated is a major concern (p. 12)." This report serves as a technical appendix to the report **Integrated Treatment Model report** that was produced in accordance with Engrossed Substitute Senate Bill 6168 Section 225(3).

Juvenile Rehabilitation uses a series of risk assessments that influence a variety of the most important decisions we make, including parole eligibility, release from residential care, community facility eligibility and treatment or program eligibility. Below are four main risk assessments and the decisions they inform.

Assessment Name	Decisions It Informs
Risk Assessment – Recidivism (RAR)	Release from incarceration within the sentencing range; parole eligibility; community facility eligibility.
Risk Assessment – Institution (RAI)	Institution security level.
Risk Assessment – Community Facility (RACF)	Community facility (CF) eligibility.
Integrated Treatment Assessment (ITA)	Currently used to determine Aggression Replacement Training (ART) eligibility. Other programs are exploring how to determine eligibility using this assessment.

The Office of Innovation, Alignment and Accountability (OIAA) contracted with two external researchers, Dr. Alexander Holsinger and Dr. Kristi Holsinger, to conduct a statistical validation study of the RAR, RAI and RACF. Their full analysis and reports are attached. Below is a brief summary, authored by OIAA, of the findings from the analysis, and some recommendations put

September 28, 2020
Office of Innovation, Alignment and Accountability | Approved for distribution by OIAA Director

forward by both the external researchers as well as some specific recommendations from OIAA researchers.

Main Findings from Statistical Validation Analysis

It is important to first note that the analysis is based on an examination of statistical validity (i.e., whether the assessments and items have an empirical relationship with the outcome of interest). There are other types of validity that can and should be considered during the process of assessing risk assessments. The authors made a determination on the performance of each assessment by gender, using a scale of "acceptable," "mixed results" and "poor."

> "'Acceptable' indicates that based on these initial analyses the scale performed at a level that is on a par with other instruments of similar ilk that have been deemed effective, and does so without a substantial number of problems or areas in need of improvement. A rating of "mixed results" means that the analysis may have revealed some signs of effectiveness, but there were issues and/or areas in need of improvement that should be addressed going forward. A rating of "poor" means that the analysis revealed the scale to be ineffective and/or to possess a number of things needing improvement. Likewise a rating of "poor" likely indicates that the instrument's use should be discontinued until further analyses and/or improvements can be made, or a more effective replacement tool can be found, tested and implemented."

The tables below indicate the performance of the RAR, RAI and RACF based on the results of the tests of statistical validity related to the outcomes of interest (recidivism, room confinement/isolation and a return from a CF).

Summary of RAR Validation Study				
Assessment	Gender	Any Recidivism	Felony Recidivism	Violent Felony Recidivism
First RAR	Female	POOR	POOR	POOR
First RAR	Male	ACCEPTABLE	ACCEPTABLE	MIXED RESULTS/POOR
Last RAR	Female	POOR	POOR	POOR
Last RAR	Male	ACCEPTABLE	MIXED RESULTS	MIXED RESULTS/POOR

Summary of RAI Validation Study			
Assessment	Gender	One or More Isolation Events	Number of Isolation Events
First RAI	Female	POOR	POOR
First RAI	Male	ACCEPTABLE	ACCEPTABLE
Last RAI	Female	POOR	POOR
Last RAI	Male	MIXED RESULTS	ACCEPTABLE

Summary of RACF Validation Study		
Assessment	Gender	Return From A CF
RACF	Female	POOR
RACF	Male	MIXED RESULTS/POOR

Findings Related to Gender

The authors **rated all three assessments as poor for females in JR.** "The vast majority of the analyses … appear to indicate that the RAR and the RAI in their current form do not hold an adequate amount of statistical validity for the female youth in JR ... The effectiveness of the [RACF] when assessing female youth is poor, with low sample size likely affecting the results."

"The RAR and RAI instruments appear to hold more validity for male youth. Statistically significant findings were commonplace [for males] across several different outcome measures … Further analyses may be warranted to fine-tune the instruments in order to determine whether different weighting schemes would improve the instrument, or elimination (or replacement) of certain items might improve prediction … If the RACF is kept in place, there is a great deal of room for improvement. Overall several of the items do not reveal a statistical relationship with the outcome. Of those that do reveal a significant relationship, differentiation in the rates of outcome tends to be inadequate in several instances."

Findings Related to Race

Other reviews of JR risk assessments have indicated racial disparity in the performance of the assessment tools.[1] For this work the authors determined that when looking that the RAR and RAI, "… male youth appear to have more disparity between white and non-white juveniles regarding false positives classification.[2] Based on the totality of the results, it appears possible that non-white youth are being over-classified in some instances by the instrument (meaning they are placed in a higher risk category than they might actuarially belong in). Additional analysis and data elements may be necessary in order to draw firmer conclusions." In terms of the RACF, "… there does not appear to be any difference in white and non-white boys regarding the number of false positive classifications. There was more difference between white and non-white girls, however … very low case counts greatly influenced these particular analyses … Finally, a series of logistic regression models (not shown) were calculated using sex, race and

[1] Internal memo on January 28, 2019, titled, "Using the RAR to determine the CERD."

[2] "False positive ratios were calculated by dividing the number of cases that were deemed false positives, by the sum of false positives and cases that were deemed true negatives. A case was considered a false positive if it had been classified via the instrument as being in the highest risk category, but did not recidivate. A case was considered a true negative if it had been classified via the instrument as being in the lowest risk category, but likewise did not recidivate. The method used is conservative in that only the extreme categorizations were used, and all subgroups (by race and sex) had the same method applied."

RACF score (or RACF category) as predictors, and the binary outcome (sent back to institution yes/no) as the dependent variable. In all modeling, sex was not significant (meaning neither boys nor girls were significantly more likely than the other to be sent back to an institution). Race however was statistically significant in all the models, indicating that non-white youth were statistically more likely to be returned to an institution from a community facility compared to white youth."

Next Steps and Recommendations

The RAR, RAI and RACF do not have adequate statistical validity, particularly for females and potentially for youth of color, to continue to rely on them for placement and release decisions. Based on the findings from the validation studies of the RAR, RAI and RACF, OIAA recommends that JR immediately stop using these assessments for females. Next, we recommend that JR establish a process to either re-tool the existing instruments or explore using variables from the Integrated Treatment Assessment (ITA) to replace the RAR, RAI and RACF scores. More specifically, JR should explore different sets of weights on the ITA to predict the outcomes that are most important (recidivism, behavior in the institution and success at a CF). The ITA contains most of the relevant information needed for risk prediction and this would result in JR needing to maintain only one instrument, instead of multiple.

The findings from the attached RAR, RAI and RACF validation study, confirm the pressing need for DCYF and JR to create an accountability structure for risk assessments moving forward.

Attachments

1. Testing the Statistical Validity of the RAR and the RAI, by Dr. Alexander M. Holsinger and Dr. Kristi Holsinger
2. Testing the Statistical Validity of the RACF, by Dr. Alexander M. Holsinger and Dr. Kristi Holsinger

Testing the statistical validity of the RAR and the RAI

Submitted to Washington State Juvenile Rehabilitation

By

Alexander M. Holsinger, Ph.D.

Kristi Holsinger, Ph.D.

June 30, 2020

Testing the statistical validity of the RAR and the RAI

Introduction

This technical report utilizes data obtained from the State of Washington's Juvenile Rehabilitation (JR) agency that allowed for tests of the statistical validity of the Risk Assessment-Recidivism (hereafter RAR), and the Risk Assessment-Institutional (hereafter RAI). A test of the Risk Assessment-Community Facility (hereafter RACF) is forthcoming pending the receipt of additional data elements. Both the RAR and the RAI are used in a number of different capacities in order to determine the likelihood of recidivism and institutional misconduct, respectively. As a result, scores on the instruments have the potential to influence placement and case processing.

Many of the analyses presented below do present the results of tests designed to determine whether a statistically significant relationship exists between individual items in an instrument and an outcome, as well as entire instruments themselves and an outcome. To say that a relationship is statistically significant means that the results obtained were unlikely to have occurred by chance, or random influences, or 'accident' alone. In other words, the observed results likely have very real meaning and decisions can be made with confidence based on the results. For example, assume the rates of felony recidivism are being compared between two groups of juveniles – those who exhibit chronic truancy (truancy Yes = 57% felony recidivism) and those who do not exhibit chronic truancy (truancy No = 25% felony recidivism). Also assume that the test statistic associated with that difference (most commonly chi-square) is revealed to be statistically significant. That means that the difference observed between the two groups is so large, that it cannot reasonably be attributed to chance factors alone – the relationship has substantive meaning. When comparing rates of recidivism across two or more groups, it is possible to observe *a* difference, but if the test statistic associated with the comparison is *not* statistically significant, that means that while a difference was observed the size of the difference can be attributed to random differences in the population that occur by chance. In other words, regardless of the difference observed, there is no substantive relationship between the groups being compared, and the outcome. It is also important to bear in mind that all statistical analyses are influenced by several conditions such as sample size, the base rates of the behaviors being

predicted, and how rigorous the test is, for example. Regardless, when a statistical analysis reveals a statistically significant relationship between an item (e.g., truancy Yes/No) and an outcome (recidivism Yes/No), that item is said to possess statistical validity. Likewise, when an entire instrument (e.g., a scale, or a score, or risk categories based on scores) reveals a statistically significant relationship with an outcome like recidivism, the entire scale or instrument is often referred to as possessing some degree of statistical validity.

As mentioned above the declaration of statistical validity is tied to the results of statistical analyses. There are other types of validity that are useful to bear in mind when developing and examining risk/need assessment tools. For example, face validity differs from statistical validity in that face validity is whether an item in an assessment (or an assessment as a whole) is subjectively measuring what it is supposed to measure. Face validity in some respects depends on appearances, or logic. For example, assume a workgroup of criminal justice stakeholders was interested in identifying items worth considering for inclusion in a yet-to-be-developed instrument that will help predict the likelihood of someone receiving a DUI conviction. The workgroup's initial brainstorming would likely be based on face validity, which means they would be coming up with items that (even to the layperson) would appear to, or "should" have, a relationship with the outcome. So it would make sense if the workgroup's initial thought processes revealed items like employment (yes/no), or how many times per week or month someone visits a bar, or how many alcoholic drinks they report having per day or week, or whether they have had a DUI charge before (and/or how many DUI charges), or whether someone reports attending church of any type, or not. All of these aforementioned items have what researchers refer to as face or logical validity, and all of them may (or may not) reveal an actual relationship with the desired outcome (in this case, likelihood of a DUI conviction).

Another form of validity that applies to existing risk/need assessment tools is measurement validity. Measurement validity refers to whether or not an item is measuring what it is purported to be measuring. Items that hold measurement validity are more likely to maintain the expected relationship with the assigned outcome, as opposed to items that do not have measurement validity. Take for example the last item listed in the hypothetical example, church attendance. The intent behind the suggestion of that item might be that if someone attends a

church of some sort, they are more likely to respect conventional norms and thereby are less likely to commit a DUI. Suppose however that in reality those who attend church are also more likely to be affluent. The item ("church attendance") is not really measuring adherence to convention, but is rather measuring socioeconomic status. As a result, the data collection based on this example would lack accuracy since the item does not measure what it purports to measure, and in turn could affect statistical validity and substantive meaning.

Most of the analyses are presented in a disaggregated fashion regarding sex, meaning separate analyses were conducted for girls and boys. In addition each instrument was subjected to an item-by-item analysis, as well as statistical tests of the entire scale as a whole. In other words, the relationship between each individual item and the relevant outcome variables was tested, as was the scale's predictive validity as a whole. Each analysis (meaning each time an instrument and its items are subjected to testing) concludes with a rating as to the effectiveness of the tool overall, using the categories of "Acceptable," "Mixed results," and "Poor." A rating of "Acceptable" does not necessarily mean the scale performed perfectly (there is no perfect risk scale resulting in perfect prediction, with no room for improvement). Rather, "acceptable" indicates that based on these initial analyses the scale performed at a level that is on a par with other instruments of similar ilk that have been deemed effective, and does so without a substantial number of problems or areas in need of improvement. A rating of "Mixed results" means that the analysis may have revealed some signs of effectiveness, but there were issues and/or areas in need of improvement that should be addressed going forward. A rating of "Poor" means that the analysis revealed the scale to be ineffective and/or to possess a number of things needing improvement. Likewise a rating of "poor" likely indicates that the instrument's use should be discontinued until further analyses and/or improvements can be made, or a more effective replacement tool can be found, tested, and implemented.

As noted above several outcomes served as the success/failure criteria for most of the analyses. The vast majority of the outcome variables are binary in nature, and involved an 18 month follow-up period for each case. Specifically, "any conviction" (misdemeanor or felony level conviction were recidivists, all other cases were successes), "felony conviction" (cases that recidivated were those who were convicted of a felony level offense, making those who only

received a misdemeanor conviction during the follow-up period a success along with all others), and "violent felony conviction" (cases that recidivated were those who were convicted of a violent felony level offense, with successes comprised of all others including those that received a misdemeanor conviction during the follow-up period, as well as those that received a non-violent felony conviction) were utilized. Outcomes also included room isolation (5 or more hours) in response to an institutional infraction. One outcome variable was linear in nature, and was operationalized as total number of room isolations. Separate analyses were conducted using the first assessment that was conducted, as well as the last (both were provided, though additional assessments using the same instrument could have occurred between the first and last).

The sample of cases was comprised of unique individuals who had served their first placement with JR, and who were released to the community during the years spanning 2010 to 2017.

Results

First RAR, Female juveniles, any conviction

Table 1 contains the results for the first analysis testing the relationship between the first RAR and any conviction (18 month follow-up), for females. The zero-order correlation between each item and any conviction was calculated, and crosstabulation/chi-square analysis was used to calculate the rates of recidivism for each category that was part of each variable. When considering and evaluating the relationship between each item and the outcome variable, it is necessary to examine whether or not the relationship is statistically significant. A lack of statistical significance indicates that there is no relationship between the variable and the relevant outcome. Statistical significance can be determined by examining both the zero-order correlation results as well as the chi-square analysis results (significance conclusions are more often than not in agreement between the two analyses). In the event a relationship between an item and the outcome variable is statistically significant, it is also necessary to determine whether the direction (+/-) is in the 'right' or expected direction. Since the RAR and RAI are additive scales, meaning higher scores are meant to indicate a higher probability of recidivism, the relationship between each item and outcome should be positive (in other words, the category of the variable that contributes points to the overall score should be affiliated with a higher probability of

Table 1.
First RAR, Female – Outcome: Any conviction, 18 month follow-up period[1]

Item	r	chi-square	% recid.
Violent offender (n = 292)	.086 (n.s.)	2.150 (n.s.)	
No			52%
Yes			43%
Drug offender (n = 292)	.070 (n.s.)	1.415 (n.s.)	
No			48%
Yes			60%
Suicide[2] (n = 292)	-.036 (n.s.)	0.376 (n.s.)	
None/4			49%
1/2/3			43%
Gang member (n = 292)	.104 (n.s.)	3.156 (n.s.)	
No			46%
Yes			58%
Comm Fac. Incid. (n = 292)	.038 (n.s.)	2.124 (n.s.)	
None			29%
Never/not scored			49%
One or more[3]			0%
Prior admissions (n = 292)	.086 (n.s.)	2.165 (n.s.)	
None			47%
Not yet scored[4]			---
One or more			58%
Mental health needs (n = 292)	.149*	6.295*	
No			42%
Not yet scored[5]			---
Yes			57%
Chemical/alcohol use (n = 283)	.028 (n.s.)	0.226 (n.s.)	
No			44%
Not yet scored[6]			---
Yes			49%

[1] n.s. indicates no relationship
* indicates statistically significant relationship at p < .05
** indicates statistically significant relationship at p < .01
*** indicates statistically significant relationship at p < .001
[2] 8 cases had values of "2" which were recoded to "20" (factor present) for this analysis.
[3] Only 2 cases met the criteria.
[4] No cases had "not yet scored" as a value.
[5] Cases assessed as "not yet scored" were not included in the analyses.
[6] Cases assessed as "not yet scored" were not included in the analyses.

Table 1. (cont.)
First RAR, Female – Outcome: Any conviction, 18 month follow-up period

Item	r	chi-square	% recid.
Age at release (n = 292)	.048[7] (n.s.)	1.318 (n.s.)	
Over 16			45%
15 or 16			52%
Under 15			48%
Savy points (n = 292)	-.089[8] (n.s.)	3.804 (n.s.)	
Vulnerable			60%
Neither			46%
Aggressive[9]			57%

Summary statistics for entire scale

Risk category	% recid.	
Level 1 (20 – 62)	39%	(27/69)
Level 2 (65 – 72)	50%	(9/18)
Level 3 (75 – 100)	51%	(74/144)
Level 4 (105 – 110)	40%	(12/30)
Level 5 (112 – 142)	65%	(20/31)

Chi-square = 6.971 (n.s.)

Zero-order correlation: r = .080 (n.s.)
AUC-ROC = .553
Alpha = .069[10]

[7] Bivariate correlation calculated using a recoded variable with values 0, 1, 2 to correspond with ascending categories.
[8] Bivariate correlation calculated using a recoded variable with values 0, 1, 2 to correspond with ascending categories.
[9] Only 7 cases were assessed as "aggressive."
[10] Cronbach's alpha calculated using recoded categorical variables.

recidivism). Finally, in the event a relationship between an item and the outcome variable is statistically significant and in the appropriate direction, it is beneficial to examine the rates of recidivism for each of the categories that make up the variable. Ideally an item that is part of a risk scale will have categories that substantially differentiate between those likely to be 'successful' and those likely to be 'failures.' Put another way, the categories of an item should have substantially different rates of failure, with the categories providing points (or more points in the event of an item with three or more possible categories) having substantially higher rates of failure.

Following the item-by-item analysis are the summary statistics for the entire scale. Specifically, the risk categories were used to determine if indeed rates of recidivism increase with each level of risk indicated (in addition a summary chi-square analysis was used). Much like the item-analysis referred to above, ideally, different categories will have substantially different rates of recidivism, and those rates will increase cleanly in a stair-step fashion. In other words, rates of failure should be lowest for the lowest risk categories and increase incrementally with each increase in risk level. Also presented as summary statistics for the entire scale is the zero-order correlation between the linear scale and the outcome measure, in addition to the area-under-the-curve/receiver-operator-characteristics (AUC-ROC analysis) which is commonly used in risk assessment research in order to test how well a scale differentiates between successes and failures. Zero-order correlations that are positive (indicating as the score increases, so does the likelihood of failure) and at or above .200 are largely considered acceptable (with larger values indicating better prediction). In lay-terms, an AUC-ROC value less than .500 indicates a scale that is less effective than flipping a coin, while a value of .500 is as effective as flipping a coin, and values greater than .500 are better than 50/50 odds at predicting success/failure, with larger values indicating better prediction. While there is some debate over what an "acceptable" AUC-ROC value is, generally values at or above .600 are considered acceptable, with values at or above .700 indicating a "good" or "effective" risk assessment scale when it comes to differentiating between successes and failures. Finally, the table concludes with a Cronbach's Alpha statistic, which is a measure of how well several items are measuring the same construct, commonly

referred to as internal consistency. Values of at least .700 for Alpha are considered to indicate an acceptable level of internal consistency for a psychometric scale.

In the case of the first RAR being used to predict any conviction for female juveniles, the results do not support the scale as an effective tool. Only one of the items evinced a significant relationship with the outcome (mental health needs). The singular significant relationship was in the appropriate direction, and the response categories (No/Yes) do appear to substantially differentiate between the likelihood of failure. Unsurprisingly the scale as a whole did not display a statistical relationship with the outcome, as shown via the rates of recidivism for each of the levels of risk, as well as the zero-order correlation. In short, there does not appear to be a relationship between the RAR female and any conviction as an outcome variable (likewise, the Alpha score was exceedingly low revealing very little if any internal consistency). It should be noted that the female juvenile sample was small relative to the male sample, though under the assumption the scale possesses some validity, more should have been revealed. In addition, approximately half of the sample was scored as "level 3" which may indicate the scale lacks sensitivity in differentiating between levels of risk.

Rating: Poor.

First RAR, male juveniles, any conviction

Table 2 contains analyses testing the relationship between the first RAR and any conviction, for male juveniles. Overall the results appear to lend support for the effectiveness of the tool. It should be noted that one (but only one) item (length of time/type of placement – see "Placement" in Table 2) did not display a statistically significant relationship with the outcome. In addition, two more items ("Age at release" and "Prior commitment") displayed rather weak relationships with the outcome (both were < .100, but both were significant), and in the case of "Age at release" the rates of recidivism were lower for the weightiest category, relative to the category below it.

Aside from the few issues noted above, the scale as a whole performed quite well when predicting any new conviction. The statistically significant correlation at .351 was strong, and the AUC-ROC value was just under .700 (Alpha score was acceptable as well). In addition the risk categories do appear to substantially differentiate between rates of recidivism, however it should

Table 2.
First RAR, Male – Outcome: Any conviction, 18 month follow-up period[11]

Item	r	chi-square	% recid.
Violent offender (n = 2430)	.222**	119.487***	
Yes			45%
No			67%
Gang member (n = 2430)	.157**	60.192**	
No			50%
Yes			66%
Placement (n = 2430)	.036[12] (n.s.)	5.192 (n.s.)	
Comm 90+			40%
Comm < 90			56%
Inst. 90 in comm			50%
Inst. no comm[13]			56%
Inst. Comm < 90			80%[14]
Prior commitment (n = 2430)	.098**	23.350**	
None			53%
One or more			65%
History of assault (n = 2294[15])	.163**	61.188**[16]	
No			41%
Yes			61%
Age first adjudication (2359[17])	.123**[18]	37.244**	
16 +			43%
14-15			55%
13 -			61%

[11] n.s. indicates no relationship
* indicates statistically significant relationship at p < .05
** indicates statistically significant relationship at p < .01
*** indicates statistically significant relationship at p < .001
[12] Bivariate correlation calculated using a categorical recoded variable.
[13] 97% of the sample scored in this category.
[14] This percentage was based on 4 cases.
[15] The original variable had cases coded as "2" which were not part of the original assessment.
[16] Chi-square and percentages were calculated using a 2 category variable even though the original data contained 3 categories (0, 2, 3, as opposed to 0, 3).
[17] The original data contained an additional category not included on the assessment (weight of 7, very few cases).
[18] Bivariate correlation calculated using three-category variable corresponding with original weights.

Table 2. (cont.)
First RAR, Male – Outcome: Any conviction, 18 month follow-up period

Item	r	chi-square	% recid.
Chemical/Alc. Use (n = 2430)	.240**[19]	140.489**	
None			33%
Unknown/not avail			46%
Impairment			62%
Prior adjudications (n = 2430)	.313**[20]	241.700**	
None			33%
One or two			56%
Three or more			68%
Compliance prior (n = 2430)	.163***[21]	66.017**	
High level			49%
Mod/not scored			63%
Non/minimal			71%
Sex offender (n = 2430)	.257**	160.296**	
Yes			30%
No			62%
Age at release (n = 2428[22])	.054**[23]	12.807**	
Over 16			52%
15 or 16			60%
Under 15			56%
Compliance w/in JRA (n = 2430)	.134**[24]	43.496**	
High			46%
Mod/not scored			58%
Non/minimal			69%

[19] Bivariate correlation calculated using three-category variable corresponding with original weights.
[20] Bivariate correlation calculated using three-category variable corresponding with original weights.
[21] Bivariate correlation calculated using three-category variable corresponding with original weights.
[22] Original data contained 3 cases scored/weighted as "1" which did not correspond with instrument. These cases were not included in the analyses.
[23] Bivariate correlation calculated using three-category variable corresponding with original weights.
[24] Bivariate correlation calculated using three-category variable corresponding with original weights.

Table 2. (cont.)
First RAR, Male – Outcome: Any conviction, 18 month follow-up period

Summary statistics for entire scale[25]

Risk category	% recid.
Level 1 (Low-36)	27%
Level 2 (37-56)	55%
Level 3 (57-69)	67%
Level 4 (70-high)	71%

Chi-square = 292.527**

Zero-order correlation: r = .351**
AUC-ROC = .698
Alpha = .618

[25] Risk categories were based on quartiles.

be noted that in the absence of original risk categorizations natural quartiles were utilized. As a result, the relatively small difference between Levels 3 and 4 for example was not ideal.

Rating: Acceptable

Last RAR, Female juveniles, any conviction

Table 3 contains the results for the analyses examining the relationship between the last RAR and any conviction for female juveniles. Only three of the 10 items revealed a statistically significant relationship with the outcome variable. Of the non-significant relationships, the rate of recidivism actually being lesser for the weightier categories (in terms of points assigned) was a common occurrence. In addition, the overall risk categories (based on natural quartiles) did not display substantial differentiation nor consistency as the risk level increased. For example, Level 3 had a higher rate of recidivism than did Level 4. In addition, overall the differentiation between Level 1 (39%) and Level 4 (52%) was narrow. While the correlation was statistically significant, it was of modest strength at best, and the AUC-ROC value was below what is considered acceptable (as was the Alpha score). Regardless of the statistically significant correlation for the scale as a whole, in light of the limitations noted above, the scale does not appear to be effective.

Rating: Poor.

Last RAR, Male juveniles, any conviction

Like the first RAR for male juveniles, the last RAR revealed the instrument to be at least somewhat effective in assessing the likelihood of future delinquent/criminal behavior (see Table 4). All individual items held a statistically significant relationship with the outcome variable, and the relationships were in the appropriate direction. In addition, for the vast majority of the items, the categories that made up the predictor item revealed satisfactory differentiation between rates of recidivism. Two exceptions to these results include “Placement” and “Age at release” which, while statistically significant, were weak relative to the other items in the scale as well as convention (e.g., zero-order correlations below .100). In addition, the rates of recidivism did not ascend consistently along with the weight and assigned points of the categories for those items.

As a whole the RAR performed well in these analyses. The risk levels revealed consistently increasing rates of recidivism that ascended accordingly, and likewise revealed some substantial differentiation. It should be pointed out that the rates of recidivism between levels 3 and 4, and

Table 3.
Last RAR, Female – Outcome: Any conviction, 18 month follow-up period[26]

Item	r	chi-square	% recid.
Violent offender (n = 346)	.107*	3.913*	
No			52%
Yes			41%
Drug offender (n = 346)	.024 (n.s.)	1.415 (n.s.)	
No			47%
Yes			52%
Suicide (n = 346)	-.051 (n.s.)	0.906 (n.s.)	
None/4			49%
1/2/3			42%
Gang member (n = 346)	.108*	4.016*	
No			45%
Yes			58%
Comm Fac. Incid. (n = 346)	.024[27] (n.s.)	.429 (n.s.)	
None			45%
Never/not scored			49%
One or more			47%
Prior admissions (n = 346)	.084[28] (n.s.)	2.455 (n.s.)	
None			46%
Not yet scored[29]			---
One or more			58%
Mental health needs (n = 342)	.177*	10.731*	
No			39%
Not yet scored[30]			---
Yes			57%

[26] n.s. indicates no relationship
* indicates statistically significant relationship at p < .05
** indicates statistically significant relationship at p < .01
*** indicates statistically significant relationship at p < .001
[27] Bivariate correlation calculated using a recoded categorized variable corresponding with original weights.
[28] Bivariate correlation calculated using a two-category dummied variable based on original scoring and weighting, without "not yet scored."
[29] No cases had "not yet scored" as a value.
[30] Cases assessed as "not yet scored" were not included in the analyses.

Table 3. (cont.)
Last RAR, Female – Outcome: Any conviction, 18 month follow-up period

Item	r	chi-square	% recid.
Chemical/alcohol use (n = 342)	.022 (n.s.)	0.158 (n.s.)	
No			45%
Not yet scored[31]			---
Yes			48%
Age at release (n = 346)	.094[32] (n.s.)	5.148 (n.s.)	
Over 16			42%
15 or 16			54%
Under 15			48%
Savy points (n = 346)	-.012[33] (n.s.)	(n.s.)	
Vulnerable			52%
Neither			46%
Aggressive[34]			80%

Summary statistics for entire scale[35]

Risk category	% recid.	
Level 1 (Low – 69)	39%	(29/75)
Level 2 (70 – 84)	43%	(37/87)
Level 3 (85 – 99)	56%	(45/81)
Level 4 (100 – High)	52%	(54/103)

Chi-square = 6.312 (n.s.)

Zero-order correlation: r = .138*
AUC-ROC = .582
Alpha = .204[36]

[31] Cases assessed as "not yet scored" were not included in the analyses.
[32] Bivariate correlation calculated using a recoded variable with values 0, 1, 2 to correspond with ascending categories.
[33] Bivariate correlation calculated using a recoded variable with values 0, 1, 2 to correspond with ascending categories.
[34] Only 5 cases were assessed as "aggressive."
[35] Risk categories were based on quartiles.
[36] Cronbach's alpha calculated using recoded categorical variables.

Table 4.
Last RAR, Male – Outcome: Any conviction, 18 month follow-up period[37]

Item	r	chi-square	% recid.
Violent offender (n = 2808)	.215**	129.887***	
Yes			45%
No			66%
Gang member (n = 2808)	.144**	58.046***	
No			49%
Yes			64%
Placement (n = 2808)	.052*[38]	24.733***	
Comm 90+			46%
Comm < 90			60%
Inst. 90 in comm			61%
Inst. no comm[39]			55%
Inst. Comm < 90			64%
Prior commitment (n = 2808)	.102**	29.441**	
None			52%
One or more			65%
History of assault (n = 2681[40])	.150**	60.386**[41]	
No			40%
Yes			59%
Age first adjudication (2808)	.129**[42]	47.362**	
16 +			43%
14-15			53%
13 -			61%
Chemical/Alc. Use (n = 2808)	.224**[43]	141.585**	
None			34%
Unknown/not avail			45%
Impairment			61%

[37] n.s. indicates no relationship
* indicates statistically significant relationship at p < .05
** indicates statistically significant relationship at p < .01
*** indicates statistically significant relationship at p < .001
[38] Bivariate correlation calculated using a categorical recoded variable.
[39] 62% of the sample scored in this category.
[40] The original variable had cases coded as "2" which were not part of the original assessment.
[41] Chi-square and percentages were calculated using a 2 category variable even though the original data contained 3 categories (0, 2, 3, as opposed to 0, 3).
[42] Bivariate correlation calculated using three-category variable corresponding with original weights.
[43] Bivariate correlation calculated using three-category variable corresponding with original weights.

Table 4. (cont.)
Last RAR, Male – Outcome: Any conviction, 18 month follow-up period

Item	r	chi-square	% recid.
Prior adjudications (n = 2808)	.314**[44]	279.828**	
None			32%
One or two			54%
Three or more			68%
Compliance prior (n = 2808)	.158***[45]	69.993**	
High level			49%
Mod/not scored			60%
Non/minimal			71%
Sex offender (n = 2808)	.259**	188.207**	
Yes			29%
No			61%
Age at release (n = 2807[46])	.060**[47]	15.730**	
Over 16			52%
15 or 16			60%
Under 15			57%
Compliance w/in JRA (n = 2808)	.138**[48]	53.523**	
High			45%
Mod/not scored			57%
Non/minimal			69%

Summary statistics for entire scale[49]

Risk category	% recid.
Level 1 (Low-13)	17%
Level 2 (14-47)	37%
Level 3 (48-53)	60%
Level 4 (54-70)	66%
Level 5 (71-high)	73%

Chi-square = 302.376***
Zero-order correlation: r = .352**
AUC-ROC = .702
Alpha = .568

[44] Bivariate correlation calculated using three-category variable corresponding with original weights.
[45] Bivariate correlation calculated using three-category variable corresponding with original weights.
[46] Original data contained 2 cases scored/weighted as "1" which did not correspond with instrument. These cases were not included in the analyses.
[47] Bivariate correlation calculated using three-category variable corresponding with original weights.
[48] Bivariate correlation calculated using three-category variable corresponding with original weights.
[49] Risk categories were based on quartiles.

4 and 5 were not as wide as might be desired. However, the zero-order correlation (.352) was of a magnitude that would be considered 'strong' and the AUC-ROC was above .700 which is at least effective (though the Alpha score was lower than desired, due to "Placement" and "Age at release"; the scale would perform even better if those items were eliminated and/or replaced with variables that reveal a stronger relationship with the outcome and/or a better fit in terms of internal consistency).

Rating: Acceptable

First RAI, Female juveniles, any conviction

Table 5 contains the results examining the relationship between the first RAI and any conviction (misdemeanor or felony) for female youth. These results should be approached with some caution as the RAI is being used to predict traditional recidivism, which is beyond the scale's intended purpose. Nonetheless, since the instrument does contain criminologically relevant variables tests were conducted in order to determine if the instrument does have validity regarding traditional recidivism (as opposed to institutional misconduct, which is the RAI's intended purpose).

None of the variables, save two, revealed a statistical relationship with recidivism. The two exceptions were "Prior adjudications" and "Sex offender" and the relationships were in the appropriate direction. Further, while the zero-order correlation testing the relationship between the scale as a whole and recidivism was statistically significant and positive, it was not strong, and moreover the risk level categories revealed inconsistent rates of recidivism. Similarly the AUC-ROC value was very low, and the Alpha score was below par as well.

Rating: Poor

First RAI, Male juveniles, any conviction

The first RAI was more effective for male youth when predicting any conviction (see Table 6). The majority of the items in the scale revealed a statistically significant relationship with the outcome variable (in the appropriate direction) with some particularly noteworthy exceptions. "Prior commitments" and "Manifest justice" revealed significant and positive correlations, but they were both below .100. "Maximum sentence" also revealed a significant and weak relationship, though the correlation was in the wrong direction which likewise resulted in rates

Table 5.
First RAI, Female juveniles selected – Outcome: Any conviction, 18 month follow-up period[50]

Item	r	chi-square	% recid.
Prior assault behavior (n = 325)	.043[51] (n.s.)	3.963 (n.s.)	
None			49%
Unknown			29%
Prior asslt beh.			50%
Impulsive (n = 325)	.086[52] (n.s.)	7.184 (n.s.)	
Generally no			39%
Unknown			50%
Occasional impulsive			28%
Frequent impulsive			57%
Chemical/Alcohol use (n = 325)	.041[53] (n.s.)	.543 (n.s.)	
None/no impairment			44%
Unknown			46%
Impairment			50%
Prior adjudications (n = 325)	.244**	19.398**	
None			32%
One or two			46%
Three or more			59%
Comply w/facility (n = 318[54])	.071[55] (n.s.)	2.564 (n.s.)	
High compliance			46%
Moderate compliance			56%
No/min. compliance			52%
History escape (n = 318[56])	.033 (n.s.)	.357 (n.s.)	
None			48%
Some			51%

[50] n.s. indicates no relationship
* indicates statistically significant relationship at p < .05
** indicates statistically significant relationship at p < .01
*** indicates statistically significant relationship at p < .001
[51] Bivariate correlation calculated using a recoded variable that reflected the categorical weights.
[52] Bivariate correlation calculated using a recoded variable that reflected the categorical weights.
[53] Bivariate correlation calculated using a recoded variable that reflected the categorical weights.
[54] The original variable contained 7 cases that were coded as "7." Those cases were not included in the analyses.
[55] Bivariate correlation calculated using a recoded variable that reflected the categorical weights.
[56] The original variable contained 7 cases that were coded as "1." Those cases were not included in the analyses.

Table 5. (cont.)
First RAI, Female juveniles selected – Outcome: Any conviction, 18 month follow-up period

Item	r	chi-square	% recid.
Prior commitments (n = 325)	.093 (n.s.)	3.066 (n.s.)	
None			47%
One			60%
Two or more			60%[57]
Sex offender (n = 325)	.136*	5.986*	
No			50%
Yes			15%
Manifest injustice (n = 325)	-.036 (n.s.)	.432 (n.s.)	
No			50%
Yes			47%
Maximum sentence (n = 325)	-.033 (n.s.)	.344 (n.s.)	
< one year			50%
One year +			46%
Admit off. Asslt/robbery (n = 325)	-.002 (n.s.)	.001 (n.s.)	
No			49%
Yes			49%

Summary statistics for entire scale

Risk category	% recid.	
Level 1 (very low 0-19)	31%	(8/26)
Level 2 (low 20-50)	50%	(114/226)
Level 3 (moderate 51-58)	39%	(12/31)
Level 4 (high 59-73)	55%	(16/29)
Level 5 (very high 74-high)	62%	(8/13)

Chi-square = 6.203 (n.s.)

Zero-order correlation: r = .120*
AUC-ROC = .565
Alpha = .437[58]

[57] Percentage based on a total of 5 cases.
[58] Cronbach's alpha calculated using recoded categorical variables.

Table 6.
First RAI, Male juveniles selected – Outcome: Any conviction, 18 month follow-up period[59]

Item	r	chi-square	% recid.
Prior assault behavior (n = 2679)	.161**[60]	73.986**	
None			41%
Unknown			43%
Prior assaultive			60%
Impulsive (n = 2679)	.159**[61]	89.571**	
Generally no			44%
Unknown			56%
Occasional impulsive			44%
Frequent impulsive			68%
Chemical/Alcohol use (n = 2679)	.234**[62]	147.052**	
None/no impairment			33%
Unknown			45%
Impairment			62%
Prior adjudications (n = 2679)	.304**	249.276**	
None			35%
One or two			55%
Three or more			68%
Comply w/facility (n =2645[63])	.164**[64]	72.139**	
High compliance			49%
Moderate compliance			62%
No/min. compliance			71%
History escape (n = 2645[65])	.165**	71.592**	
None			51%
Some			70%
Prior commitments (n = 2679)	.094**	26.226**	
None			52%
One			64%
Two or more			65%

[59] n.s. indicates no relationship
* indicates statistically significant relationship at p < .05
** indicates statistically significant relationship at p < .01
*** indicates statistically significant relationship at p < .001
[60] Bivariate correlation calculated using a recoded variable that reflected the categorical weights.
[61] Bivariate correlation calculated using a recoded variable that reflected the categorical weights.
[62] Bivariate correlation calculated using a recoded variable that reflected the categorical weights.
[63] The original variable contained 34 cases that were coded as "7." Those cases were not included in the analyses.
[64] Bivariate correlation calculated using a recoded variable that reflected the categorical weights.
[65] The original variable contained 34 cases that were coded as "1." Those cases were not included in the analyses.

Table 6. (cont.)
First RAI, Male juveniles selected – Outcome: Any conviction, 18 month follow-up period

Item	r	chi-square	% recid.
Sex offender (n = 2679)	.253**	171.168**	
No			61%
Yes			30%
Manifest injustice (n = 2679)	.065**	11.370**	
No			53%
Yes			60%
Maximum sentence (n = 2679)	-.062**	10.274**	
< one year			57%
One year +			51%
Admit off. Asslt/robbery (n = 2679)	-.026 (n.s.)	1.863 (n.s.)	
No			56%
Yes			53%

Summary statistics for entire scale

Risk category	% recid.
Level 1 (very low 0-19)	36%
Level 2 (low 20-50)	53%
Level 3 (moderate 51-58)	58%
Level 4 (high 59-73)	69%
Level 5 (very high 74-high)	74%

Chi-square = 104.100***

Zero-order correlation: r = .205***
AUC-ROC = .613
Alpha = .636[66]

[66] Cronbach's alpha calculated using recoded categorical variables.

of outcome being "backward" relative to the weight/scoring of the item's categories. "Admitting offense" did not reveal a relationship with the outcome.

Despite the deficiencies noted above, the scale as a whole did reveal a statistically significant relationship with the outcome variable (which, as noted above was not the outcome variable the instrument was intended for). Rates of recidivism did ascend with each increase in the level of risk accordingly (though differentiation was narrower than desired between levels 2 and 3, and 4 and 5), and the zero-order correlation was significant as well, and was likewise above .200. The AUC-ROC value was above the threshold of acceptability, though the Alpha score was slightly below where it should be.

Rating: Mixed results

Last RAI, Female juveniles, any conviction

The last RAI revealed similar results as the first RAI for female youth when predicting any conviction. Of the 11 items only two revealed a statistical relationship with the outcome (results contained in Table 7). "Prior adjudications" and "Sex offender" both revealed significant and positive relationships with any conviction, with both correlations above .100, and each item's categories showed substantial differentiation in rates of recidivism. Overall however the scale did not perform well. The zero-order correlation was not significant, and the rates of recidivism did not ascend with categories of risk. Unsurprisingly the AUC-ROC was below what is considered acceptable as was the Alpha score (though in keeping with Tables 5 and 6, these results should be considered experimental since the scale is being applied to an unintended outcome variable).

Rating: Poor

Last RAI, Male juveniles, any conviction

Table 8 contains the results examining the last RAI for male juveniles when predicting any conviction. Of the 11 items on the instrument four showed weak and/or problematic results. Specifically the relationships for "Prior commitments" and "Manifest injustice" were significant and positive but weak (< .100), "Maximum sentence" was significant but negative (resulting in higher recidivism rates for the non-point category), and "Admitting offense" did not maintain a relationship with the outcome variable at all.

Table 7.
Last RAI, Female juveniles selected – Outcome: Any conviction, 18 month follow-up period[67]

Item	r	chi-square	% recid.
Prior assault behavior (n = 282)	-.017[68] (n.s.)	1.375 (n.s.)	
None			53%
Unknown			33%
Prior asslt beh.			48%
Impulsive (n = 282)	.043[69] (n.s.)	1.339 (n.s.)	
Generally no			47%
Unknown			47%
Occasional impulsive			33%
Frequent impulsive			54%
Chemical/Alcohol use (n = 282)	.028[70] (n.s.)	.290 (n.s.)	
None/no impairment			43%
Unknown			50%
Impairment			48%
Prior adjudications (n = 282)	.155**	6.828**	
None			37%
One or two			48%
Three or more			55%
Comply w/facility (n =282)	.077[71] (n.s.)	1.898 (n.s.)	
High compliance			46%
Moderate compliance			55%
No/min. compliance			56%
History escape (n =282)	-.040 (n.s.)	.447 (n.s.)	
None			49%
Some			45%
Prior commitments (n = 282)	.035[72] (n.s.)	2.469 (n.s.)	
None			47%
One			58%
Two or more			25%[73]

[67] n.s. indicates no relationship
* indicates statistically significant relationship at p < .05
** indicates statistically significant relationship at p < .01
*** indicates statistically significant relationship at p < .001
[68] Bivariate correlation calculated using a recoded variable that reflected the categorical weights.
[69] Bivariate correlation calculated using a recoded variable that reflected the categorical weights.
[70] Bivariate correlation calculated using a recoded variable that reflected the categorical weights.
[71] Bivariate correlation calculated using a recoded variable that reflected the categorical weights.
[72] Bivariate correlation calculated using a recoded variable that reflected the categorical weights.
[73] Percentage based on a total of 4 cases.

Table 7. (cont.)
Last RAI, Female juveniles selected – Outcome: Any conviction, 18 month follow-up period

Item	r	chi-square	% recid.
Sex offender (n = 282)	.120*	4.043*	
No			49%
Yes			18%[74]
Manifest injustice (n = 282)	.022 (n.s.)	.133 (n.s.)	
No			47%
Yes			49%
Maximum sentence (n = 282)	-.031 (n.s.)	.271 (n.s.)	
< one year			49%
One year +			45%
Admit off. Asslt/robbery (n = 282)	-.014 (n.s.)	.055 (n.s.)	
No			49%
Yes			47%

Summary statistics for entire scale

Risk category	% recid.	
Level 1 (very low 0-19)	37%	(10/27)
Level 2 (low 20-50)	49%	(104/212)
Level 3 (moderate 51-58)	41%	(7/17)
Level 4 (high 59-73)	55%	(11/20)
Level 5 (very high 74-high)	50%	(3/6)

Chi-square = 6.203 (n.s.)

Zero-order correlation: r = .078 (n.s.)
AUC-ROC = .544
Alpha = .407[75]

[74] Percentage based on 11 cases.
[75] Cronbach's alpha calculated using recoded categorical variables.

Table 8.
Last RAI, Male juveniles selected – Outcome: Any conviction, 18 month follow-up period[76]

Item	r	chi-square	% recid.
Prior assault behavior (n = 2153)	.153**[77]	52.212**	
None			41%
Unknown			57%
Prior assaultive			60%
Impulsive (n = 2153)	.145**[78]	51.612**	
Generally no			46%
Unknown			58%
Occasional impulsive			54%
Frequent impulsive			67%
Chemical/Alcohol use (n = 2153)	.218**[79]	105.071**	
None/no impairment			33%
Unknown			55%
Impairment			62%
Prior adjudications (n = 2153)	.309**	206.204**	
None			37%
One or two			56%
Three or more			70%
Comply w/facility (n = 2153)	.164**	60.174**	
High compliance			50%
Moderate compliance			66%
No/min. compliance			72%
History escape (n = 2153)	.156**	52.223**	
None			52%
Some			72%
Prior commitments (n = 2153)	.096**[80]	21.786**	
None			54%
One			66%
Two or more			67%

[76] n.s. indicates no relationship
* indicates statistically significant relationship at p < .05
** indicates statistically significant relationship at p < .01
*** indicates statistically significant relationship at p < .001
[77] Bivariate correlation calculated using a recoded variable that reflected the categorical weights.
[78] Bivariate correlation calculated using a recoded variable that reflected the categorical weights.
[79] Bivariate correlation calculated using a recoded variable that reflected the categorical weights.
[80] Bivariate correlation calculated using a recoded variable that reflected the categorical weights.

Table 8. (cont.)
Last RAI, Male juveniles selected – Outcome: Any conviction, 18 month follow-up period

Item	r	chi-square	% recid.
Sex offender (n = 2153)	.287**	177.493**	
No			63%
Yes			28%
Manifest injustice (n = 2153)	.080**	13.780**	
No			54%
Yes			62%
Maximum sentence (n = 2153)	-.053*	5.989*	
< one year			58%
One year +			53%
Admit off. Asslt/robbery (n = 2153)	-.018 (n.s.)	.690 (n.s.)	
No			57%
Yes			55%

Summary statistics for entire scale

Risk category	% recid.
Level 1 (very low 0-19)	36%
Level 2 (low 20-50)	56%
Level 3 (moderate 51-58)	61%
Level 4 (high 59-73)	77%
Level 5 (very high 74-high)	72%

Chi-square = 105.113***

Zero-order correlation: r = .224***
AUC-ROC = .624
Alpha = .617[81]

[81] Cronbach's alpha calculated using recoded categorical variables.

Despite the aforementioned deficiencies the scale as a whole did "work." Recidivism rates generally increased with each increase in level of risk (though some of the differentiation between categories was narrower than what would be considered ideal, particularly between levels 2 and 3, and rates actually decreased from level 4 to 5). The overall zero-order correlation was statistically significant, positive, and above .200, and the AUC-ROC score was acceptable (thought he Alpha score was low, which was likely due to the number of items that presented issues).

Rating: Mixed results

First RAI, Female juveniles, one or more isolations, 5+ hours

The RAI was applied to its intended outcome variable (institutional misconduct) via Tables 9 through 12. In Table 9 female juveniles were selected, and the first RAI was examined. Despite the appropriate outcome being used only one item ("Compliance within facility") revealed a statistically significant relationship without any problems. "Impulsive," while significant and positive had categories that did not differentiate rates of the outcome variable (rates were equal for "Generally no" and "Unknown") and rates actually decreased between "Occasional" and "Frequent" impulsiveness. "Chemical/alcohol use" had a significant chi-square value but rates of the outcome increased from the first to second category then decreased (dramatically) from the second to third category. Likewise "Prior adjudications" showed inconsistency across ascending categories regarding rates of the failure criteria occurring. "Prior commitments," though revealing a statistically significant and positive relationship did not show good differentiation between rates of outcome for the top two categories. The remaining six items on the scale showed no relationship with the outcome. Although overall the zero-order correlation for the entire scale was significant and positive with an AUC-ROC above .600, the rates of the outcome variable varied greatly across ascending levels of risk.

Rating: Poor

First RAI, Male juveniles, one or more isolations, 5+ hours

Much stronger results for the RAI were revealed via Table 10, which contains the results for analyses involving the first RAI for male juveniles when predicting institutional misconduct. Every variable save one ("Manifest injustice") revealed a statistically significant and positive

Table 9.
First RAI, Female juveniles selected – Outcome: room isolation, 5+ hours[82]

Item	r	chi-square	% room isolation
Prior assault behavior (n = 420)	.010[83] (n.s.)	2.302 (n.s.)	
None			14%
Unknown			29%
Prior asslt beh.			18%
Impulsive (n = 420)	.140**[84]	10.937*	
Generally no			15%
Unknown			15%
Occasional impulsive			33%
Frequent impulsive			28%
Chemical/Alcohol use (n = 420)	.018[85] (n.s.)	10.483**	
None/no impairment			11%
Unknown			42%
Impairment			18%
Prior adjudications (n = 420)	.096*	6.239*	
None			15%
One or two			11%
Three or more			23%
Comply w/facility (n = 413[86])	.117[87]*	5.753 (n.s.)	
High compliance			15%
Moderate compliance			21%
No/min. compliance			30%
History escape (n = 413[88])	.091 (n.s.)	3.402 (n.s.)	
None			15%
Some			23%

[82] n.s. indicates no relationship
* indicates statistically significant relationship at p < .05
** indicates statistically significant relationship at p < .01
*** indicates statistically significant relationship at p < .001
[83] Bivariate correlation calculated using a recoded variable that reflected the categorical weights.
[84] Bivariate correlation calculated using a recoded variable that reflected the categorical weights.
[85] Bivariate correlation calculated using a recoded variable that reflected the categorical weights.
[86] The original variable contained 7 cases that were coded as "7." Those cases were not included in the analyses.
[87] Bivariate correlation calculated using a recoded variable that reflected the categorical weights.
[88] The original variable contained 7 cases that were coded as "1." Those cases were not included in the analyses.

Table 9. (cont.)
First RAI, Female juveniles selected – Outcome: room isolation, 5+ hours

Item	r	chi-square	% room isolation
Prior commitments (n = 420)	.123*[89]	6.625*	
None			17%
One			30%
Two or more			33%[90]
Sex offender (n = 420)	.085 (n.s.)	3.060 (n.s.)	
No			19%
Yes			0%[91]
Manifest injustice (n = 420)	-.045 (n.s.)	.849 (n.s.)	
No			20%
Yes			17%
Maximum sentence (n = 420)	.058 (n.s.)	1.414 (n.s.)	
< one year			17%
One year +			22%
Admit off. Asslt/robbery (n = 420)	-.043 (n.s.)	.794 (n.s.)	
No			20%
Yes			17%

Summary statistics for entire scale

Risk category	% room isolation	
Level 1 (very low 0-19)	19%	(6/31)
Level 2 (low 20-50)	15%	(45/302)
Level 3 (moderate 51-58)	33%	(12/36)
Level 4 (high 59-73)	34%	(13/38)
Level 5 (very high 74-high)	15%	(2/13)

Chi-square = 14.124**

Zero-order correlation: r = .149**
AUC-ROC = .625
Alpha = .437[92]

[89] Bivariate correlation calculated using a recoded variable that reflected the categorical weights.
[90] Percentage based on a total of 6 cases.
[91] This percentage is based on 13 cases.
[92] Cronbach's alpha calculated using recoded categorical variables.

Table 10.
First RAI, Male juveniles selected – Outcome: Any isolation 5+ hours[93]

Item	r	chi-square	% room isolation
Prior assault behavior (n = 3549)	.141***[94]	72.276***	
None			21%
Unknown			33%
Prior assaultive			39%
Impulsive (n = 3549)	.209***[95]	167.769***	
Generally no			21%
Unknown			36%
Occasional impulsive			39%
Frequent impulsive			49%
Chemical/Alcohol use (n = 3549)	.114***[96]	49.402***	
None/no impairment			23%
Unknown			36%
Impairment			38%
Prior adjudications (n = 3549)	.163***	94.132***	
None			25%
One or two			33%
Three or more			42%
Comply w/facility (n =3515[97])	.224***[98]	177.793***	
High compliance			27%
Moderate compliance			44%
No/min. compliance			56%
History escape (n = 3515[99])	.111***	43.206***	
None			32%
Some			45%

[93] n.s. indicates no relationship
* indicates statistically significant relationship at p < .05
** indicates statistically significant relationship at p < .01
*** indicates statistically significant relationship at p < .001
[94] Bivariate correlation calculated using a recoded variable that reflected the categorical weights.
[95] Bivariate correlation calculated using a recoded variable that reflected the categorical weights.
[96] Bivariate correlation calculated using a recoded variable that reflected the categorical weights.
[97] The original variable contained 34 cases that were coded as "7." Those cases were not included in the analyses.
[98] Bivariate correlation calculated using a recoded variable that reflected the categorical weights.
[99] The original variable contained 34 cases that were coded as "1." Those cases were not included in the analyses.

Table 10. (cont.)
First RAI, Male juveniles selected – Outcome: Any isolation 5+ hours

Item	r	chi-square	% room isolation
Prior commitments (n = 3549)	.123***[100]	54.438***	
None			32%
One			45%
Two or more			52%
Sex offender (n = 3549)	.142***	71.672***	
No			38%
Yes			21%
Manifest injustice (n = 3549)	.018 (n.s.)	1.121 (n.s.)	
No			34%
Yes			36%
Maximum sentence (n = 3549)	.116***	47.356***	
< one year			30%
One year +			41%
Admit off. Asslt/robbery (n = 3549)	.051**	9.258**	
No			33%
Yes			38%

Summary statistics for entire scale

Risk category	% room isolation
Level 1 (very low 0-19)	14%
Level 2 (low 20-50)	31%
Level 3 (moderate 51-58)	45%
Level 4 (high 59-73)	55%
Level 5 (very high 74-high)	58%

Chi-square = 240.935***

Zero-order correlation: r = .292***
AUC-ROC = .678
Alpha = .636[101]

[100] Bivariate correlation calculated using a recoded variable that reflected the categorical weights.
[101] Cronbach's alpha calculated using recoded categorical variables.

relationship with the outcome variable, and likewise revealed increasing rates for ascending response categories. “Admitting offense” was the only other item that revealed results that were less than ideal (e.g., a significant and positive but weak correlation, and narrow differentiation in rates of outcome) but not to the extent that the item warrants reconsideration regarding its inclusion in the scale.

The scale as a whole performed adequately, with a statistically significant and relatively strong zero-order correlation (.292), an acceptable AUC-ROC (.678), and good differentiation in rates of outcome across ascending levels of risk, though the difference was narrower than ideal between Levels 4 and 5.

Rating: Acceptable

Last RAI, Female juveniles, one or more isolations, 5+ hours

The last RAI for female juveniles performed very poorly when predicting institutional misconduct (see Table 11). None of the individual items were without substantial problems (with 7 of the items revealing no relationship at all). Of the items revealing some statistical significance, the rates of outcome did not ascend across increasing weights of categories (e.g., “Chemical/alcohol use,” “Prior adjudications,” “Prior commitments”) and in the instance of “Admitting offense” the correlation was negative and rates of outcome were opposite what they should be.

In light of the aforementioned problems the scale as a whole unsurprisingly did not perform well. The zero-order correlation was not significant indicating no relationship with the outcome, the AUC-ROC was well below what is considered acceptable, the Alpha score was very poor, and the rates of outcome across ascending categories of risk were inconsistent.

Rating: Poor

Last RAI, Male juveniles, one or more isolations, 5+ hours

The analyses involving the last RAI for male juveniles revealed some effectiveness when predicting institutional misconduct (see Table 12). Most of the items revealed a statistically significant and positive relationship with the outcome, along with ascending rates of outcome across increases in category weights. “Impulsive” was an exception to this regarding rates of outcome (e.g., the percentage of cases with one or more institutional misconduct incidents

Table 11.
Last RAI, Female juveniles selected – Outcome: isolation 5+ hours[102]

Item	r	chi-square	% room isolation
Prior assault behavior (n = 380)	.021[103] (n.s.)	2.057 (n.s.)	
None			14%
Unknown			0%[104]
Prior asslt beh.			15%
Impulsive (n = 380)	.073[105] (n.s.)	3.479 (n.s.)	
Generally no			12%
Unknown			13%
Occasional impulsive			0%[106]
Frequent impulsive			20%
Chemical/Alcohol use (n = 380)	.074[107] (n.s.)	7.541*	
None/no impairment			2.8%
Unknown			33%
Impairment			15%
Prior adjudications (n = 380)	.125*	6.777*	
None			9.4%
One or two			9.7%
Three or more			19%
Comply w/facility (n = 380)	.071 (n.s.)	2.555 (n.s.)	
High compliance			13%
Moderate compliance			15%
No/min. compliance			25%
History escape (n = 380)	.030 (n.s.)	.333 (n.s.)	
None			14%
Some			16%

[102] n.s. indicates no relationship
* indicates statistically significant relationship at p < .05
** indicates statistically significant relationship at p < .01
*** indicates statistically significant relationship at p < .001
[103] Bivariate correlation calculated using a recoded variable that reflected the categorical weights.
[104] Percentage based on 12 cases.
[105] Bivariate correlation calculated using a recoded variable that reflected the categorical weights.
[106] Percentages based on 6 cases.
[107] Bivariate correlation calculated using a recoded variable that reflected the categorical weights.

Table 11. (cont.)
Last RAI, Female juveniles selected – Outcome: isolation 5+ hours

Item	r	chi-square	% room isolation
Prior commitments (n = 380)	.153**[108]	10.859**	
None			12%
One			29%
Two or more			20%[109]
Sex offender (n = 380)	.070 (n.s.)	1.876 (n.s.)	
No			15%
Yes			0%[110]
Manifest injustice (n = 380)	-.016 (n.s.)	.092 (n.s.)	
No			15%
Yes			14%
Maximum sentence (n = 380)	.012 (n.s.)	.052 (n.s.)	
< one year			14%
One year +			15%
Admit off. Asslt/robbery (n = 380)	-.122*	5.678*	
No			18%
Yes			9%

Summary statistics for entire scale

Risk category	% room isolation	
Level 1 (very low 0-19)	16%	(5/32)
Level 2 (low 20-50)	13%	(37/291)
Level 3 (moderate 51-58)	18%	(4/22)
Level 4 (high 59-73)	28%	(8/29)
Level 5 (very high 74-high)	0%	(0/6)

Chi-square = 6.121 (n.s.)

Zero-order correlation: r = .083 (n.s.)
AUC-ROC = .576
Alpha = .407[111]

[108] Bivariate correlation calculated using a recoded variable that reflected the categorical weights.
[109] Percentage based on a total of 5 cases.
[110] Percentage based on 11 cases.
[111] Cronbach's alpha calculated using recoded categorical variables.

Table 12.
Last RAI, Male juveniles selected – Outcome: Any isolation 5+ hours[112]

Item	r	chi-square	% room isolation
Prior assault behavior (n = 3037)	.135***[113]	56.310***	
None			21%
Unknown			25%
Prior assaultive			37%
Impulsive (n = 3037)	.188***[114]	118.127***	
Generally no			23%
Unknown			33%
Occasional impulsive			26%
Frequent impulsive			48%
Chemical/Alcohol use (n = 3037)	.126***[115]	51.627***	
None/no impairment			19%
Unknown			35%
Impairment			36%
Prior adjudications (n = 3037)	.152***	71.015***	
None			25%
One or two			30%
Three or more			40%
Comply w/facility (n = 3037)	.205***	127.757***	
High compliance			27%
Moderate compliance			42%
No/min. compliance			54%
History escape (n = 3037)	.094***	26.996***	
None			31%
Some			42%
Prior commitments (n = 3037)	.119***[116]	46.353***	
None			30%
One			44%
Two or more			46%

[112] n.s. indicates no relationship
* indicates statistically significant relationship at p < .05
** indicates statistically significant relationship at p < .01
*** indicates statistically significant relationship at p < .001
[113] Bivariate correlation calculated using a recoded variable that reflected the categorical weights.
[114] Bivariate correlation calculated using a recoded variable that reflected the categorical weights.
[115] Bivariate correlation calculated using a recoded variable that reflected the categorical weights.
[116] Bivariate correlation calculated using a recoded variable that reflected the categorical weights.

Table 12. (cont.)
Last RAI, Male juveniles selected – Outcome: Any isolation 5+ hours

Item	r	chi-square	% room isolation
Sex offender (n = 3037)	.142***	60.998***	
No			36%
Yes			20%
Manifest injustice (n = 3037)	.015 (n.s.)	.682 (n.s.)	
No			33%
Yes			34%
Maximum sentence (n = 3037)	.113***	38.927***	
< one year			29%
One year +			40%
Admit off. Asslt/robbery (n = 3037)	.059**	10.659**	
No			31%
Yes			37%

Summary statistics for entire scale

Risk category	% room isolation
Level 1 (very low 0-19)	14%
Level 2 (low 20-50)	30%
Level 3 (moderate 51-58)	45%
Level 4 (high 59-73)	57%
Level 5 (very high 74-high)	52%

Chi-square = 196.522***

Zero-order correlation: r = .284***
AUC-ROC = .677
Alpha = .617[117]

[117] Cronbach's alpha calculated using recoded categorical variables.

decreased between the second and third categories but then increased dramatically for the last category). In addition "Chemical/alcohol use" had very little differentiation in rates of outcome between the second and third categories, a characteristic that "Prior commitments" and "Admitting offense" shared to some extent. "Manifest injustice" again revealed no relationship with the outcome variable.

Despite the aforementioned deficiencies the scale as a whole revealed a relatively strong and statistically significant zero-order correlation (.284), an acceptable AUC-ROC (.677) and (relative to prior analyses) a stronger than usual Alpha score. The differentiation in rates of the outcome occurring across ascending levels of risk were good, though it should be noted that rates decreased slightly from level 4 to 5.

Rating: Mixed results

First and last RAI, Female juveniles, number of isolation events

Table 13 presents analyses for the first and last RAIs for female juveniles, using number of isolation events as the outcome variable (with the number of outcome events truncated at 7 or more). For the first RAI only four of the 11 items ("Impulsive," "History of escape," "Prior commitments," and "Maximum sentence") revealed a significant and positive relationship with the outcome variable, though all four were above .100. The entire scale was likewise significantly and positively related to the outcome (r = .185).

The last RAI appeared to perform worse than the first, with only two items ("Prior adjudication" and ""Prior commitments") showing any relationship with the outcome at all (though both correlations were significant, positive, and of an acceptable magnitude). Not surprisingly the scale as a whole revealed a substantially weaker relationship with the outcome variable (r = .103).

Rating first RAI: Poor

Rating second RAI: Poor

First and last RAI, Male juveniles, number of isolation events

Table 14 presents similar analyses as those contained in Table 13, though male juveniles were selected for inclusion. Overall both the first and last RAIs performed better for male juveniles compared to female juveniles. With the exception of "Manifest injustice" which showed

Table 13.
First and Last RAI, Female juveniles selected – Outcome: Number isolation events (7 = 7+)[118]

Item	r-first[119]	r-last[120]
Prior assault behavior	.029 (n.s.)	.044 (n.s.)
Impulsive	.188***	.094 (n.s.)
Chemical/alcohol	-.067 (n.s.)	-.009 (n.s.)
Prior adjudication	.081 (n.s.)	.155**
Comply w/facility	.096 (n.s.)	.019 (n.s.)
History of escape	.108*	.034 (n.s.)
Prior commitments	.101*	.201***
Sex offender	.064 (n.s.)	.052 (n.s.)
Manifest	-.021 (n.s.)	-.006 (n.s.)
Maximum sentence	.107*	.016 (n.s.)
Assault/robb. admitting offense	-.002 (n.s.)	-.070 (n.s.)
Entire scale	.185***	.103*

[118] n.s. indicates no relationship
* indicates statistically significant relationship at p < .05
** indicates statistically significant relationship at p < .01
*** indicates statistically significant relationship at p < .001
[119] n = 420 for each analysis except for compliance within the facility and history of escape where n = 413.
[120] n = 380 for each analysis.

Table 14.
First and Last RAI, Male juveniles selected – Outcome: Number isolation events (7 = 7+)[121]

Item	r-first[122]	r-last[123]
Prior assault behavior	.148***	.138***
Impulsive	.296***	.275***
Chemical/alcohol	.105***	.127***
Prior adjudication	.180***	.167***
Comply w/facility	.310***	.286***
History of escape	.135***	.125***
Prior commitments	.130***	.117***
Sex offender	.128***	.131***
Manifest	.012 (n.s.)	.013 (n.s.)
Maximum sentence	.153***	.147***
Assault/robb. admitting offense	.068***	.078***
Entire scale	.365***	.356***

[121] n.s. indicates no relationship
* indicates statistically significant relationship at p < .05
** indicates statistically significant relationship at p < .01
*** indicates statistically significant relationship at p < .001
[122] n = 3549 for each analysis except for compliance within the facility and history of escape where n = 3515.
[123] n = 3037 for each analysis.

no relationship with the outcome variable (number of isolation events truncated at 7+), all items revealed a statistically significant and positive relationship for both the first and last RAI (though in the case of "Admitting offense" the strength was below .100). Further, the correlation for the entire scale was very strong for both the first and last RAI (r = .365 and .356 respectively).

Rating first RAI: Acceptable

Rating second RAI: Acceptable

First RAR, Female juveniles, felony convictions

In Table 15, the RAR for female juveniles is reexamined, though the outcome variable isolates those who received a felony conviction within the 18 month follow-up period as recidivists. None of the items on their own revealed a statistically significant relationship with the outcome variable, though in some cases the rates of recidivism did increase accordingly with ascending weighted categories (see for example "Drug offender," "Gang member," "Prior admissions," "Mental health needs," and "Chemical/alcohol use").

The RAR as a whole did reveal a statistically significant and positive correlation with felony conviction (r = .148), and the AUC-ROC was slightly above .600 (the Alpha score was extremely poor). Further, the rates of the outcome criteria did ascend somewhat with increasing levels of risk (level 3 to 4 excepted), but moreover appear to reveal two levels overall, as opposed to five. For example, levels 1 and 2 had rates around 10%, while levels 3 through 5 had rates around 25%.

Rating: Poor

First RAR, Male juveniles, felony convictions

The first RAR for male juveniles was used to predict felony conviction in Table 16. Of the 12 items on the instrument, three revealed no relationship with the outcome variable ("Placement," "Age at first adjudication," and "Age at release"). While every other item was significantly (and positively) related to the outcome variable, three of them ("Violent offender," "Prior commitment" and "History of assault") had correlations less than .100. The scale as a whole did reveal a statistically significant and positive correlation of an acceptable magnitude (r = .195) and the AUC-ROC was above .600. It should be noted that levels 2 and 3 did not differentiate well between rates of recidivism, though as in previous analyses the categories were based on natural quartiles.

Table 15.
First RAR, Female – Outcome: Felony conviction, 18 month follow-up period[124]

Item	r	chi-square	% recid.
Violent offender (n = 292)	.106 (n.s.)	3.254 (n.s.)	
No			14%
Yes			23%
Drug offender (n = 292)	.065 (n.s.)	1.251 (n.s.)	
No			19%
Yes			28%
Suicide[125] (n = 292)	-.081 (n.s.)	1.929 (n.s.)	
None/4			21%
1/2/3			10%
Gang member (n = 292)	.072 (n.s.)	1.506 (n.s.)	
No			18%
Yes			25%
Comm Fac. Incid. (n = 292)	.062[126] (n.s.)	1.995 (n.s.)	
None			0%[127]
Never/not scored			20%
One or more			0%[128]
Prior admissions (n = 292)	.108 (n.s.)	3.402 (n.s.)	
None			18%
Not yet scored[129]			---
One or more			29%
Mental health needs (n = 283)	.111 (n.s.)	3.466 (n.s.)	
No			15%
Not yet scored[130]			---
Yes			25%

[124] n.s. indicates no relationship
* indicates statistically significant relationship at p < .05
** indicates statistically significant relationship at p < .01
*** indicates statistically significant relationship at p < .001
[125] 8 cases had values of "2" which were recoded to "20" (factor present) for this analysis.
[126] Bivariate correlation calculated using a categorized variable that corresponded with original weights.
[127] Percentage based on 7 cases.
[128] Percentage based on 1 case.
[129] No cases had "not yet scored" as a value.
[130] Cases assessed as "not yet scored" were not included in the analyses.

Table 15. (cont.)
First RAR, Female – Outcome: Felony conviction, 18 month follow-up period

Item	r	chi-square	% recid.
Chemical/alcohol use (n = 282)	.053 (n.s.)	.798 (n.s.)	
No			13%
Not yet scored[131]			---
Yes			21%
Age at release (n = 292)	.036[132] (n.s.)	1.159 (n.s.)	
Over 16			17%
15 or 16			22%
Under 15			17%
Savy points (n = 292)	.015[133] (n.s.)	.326 (n.s.)	
Vulnerable			18%
Neither			20%
Aggressive[134]			14%

Summary statistics for entire scale

Risk category	% recid.	
Level 1 (20 – 62)	9%	(6/69)
Level 2 (65 – 72)	11%	(2/18)
Level 3 (75 – 100)	24%	(34/144)
Level 4 (105 – 110)	23%	(7/30)
Level 5 (112 – 142)	26%	(8/31)

Chi-square = 8.548 (n.s.)

Zero-order correlation: r = .148*
AUC-ROC = .611
Alpha = .069[135]

[131] Cases assessed as "not yet scored" were not included in the analyses.
[132] Bivariate correlation calculated using a recoded variable with values 0, 1, 2 to correspond with ascending categories.
[133] Bivariate correlation calculated using a recoded variable with values 0, 1, 2 to correspond with ascending categories.
[134] Only 7 cases were assessed as "aggressive."
[135] Cronbach's alpha calculated using recoded categorical variables.

Table 16.
First RAR, Male – Outcome: Felony conviction, 18 month follow-up period[136]

Item	r	chi-square	% recid.
Violent offender (n = 2430)	.082***	16.255***	
Yes			22%
No			29%
Gang member (n = 2430)	.129***	40.171***	
No			21%
Yes			33%
Placement (n = 2430)	.033[137] (n.s.)	5.752 (n.s.)	
Comm 90+			11%
Comm < 90			28%
Inst. 90 in comm			50%[138]
Inst. no comm[139]			25%
Inst. Comm < 90			40%[140]
Prior commitment (n = 2430)	.081***	15.925***	
None			23%
One or more			32%
History of assault (n = 2294[141])	.087***	17.227***[142]	
No			19%
Yes			28%
Age first adjudication (2359[143])	.035 (n.s.)[144]	5.905 (n.s.)	
16 +			21%
14-15			27%
13 -			26%

[136] n.s. indicates no relationship
* indicates statistically significant relationship at p < .05
** indicates statistically significant relationship at p < .01
*** indicates statistically significant relationship at p < .001
[137] Bivariate correlation calculated using a categorical recoded variable.
[138] This percentage was based on 2 cases.
[139] 97% of the sample scored in this category.
[140] This percentage was based on 5 cases.
[141] The original variable had cases coded as "2" which were not part of the original assessment.
[142] Chi-square and percentages were calculated using a 2 category variable even though the original data contained 3 categories (0, 2, 3, as opposed to 0, 3).
[143] The original data contained an additional category not included on the assessment (weight of 7, very few cases).
[144] Bivariate correlation calculated using three-category variable corresponding with original weights.

Table 16. (cont.)
First RAR, Male – Outcome: Felony conviction, 18 month follow-up period

Item	r	chi-square	% recid.
Chemical/Alc. Use (n = 2430)	.243***[145]	50.510***	
None			14%
Unknown/not avail			18%
Impairment			29%
Prior adjudications (n = 2430)	.148***[146]	53.105***	
None			16%
One or two			24%
Three or more			31%
Compliance prior (n = 2430)	.105***[147]	26.675***	
High level			22%
Mod/not scored			28%
Non/minimal			35%
Sex offender (n = 2430)	.160***	62.028***	
Yes			12%
No			29%
Age at release (n = 2428[148])	-.015 (n.s.)[149]	3.726 (n.s.)	
Over 16			25%
15 or 16			27%
Under 15			20%
Compliance w/in JRA (n = 2430)	.114***[150]	31.919***	
High			18%
Mod/not scored			27%
Non/minimal			35%

[145] Bivariate correlation calculated using three-category variable corresponding with original weights.
[146] Bivariate correlation calculated using three-category variable corresponding with original weights.
[147] Bivariate correlation calculated using three-category variable corresponding with original weights.
[148] Original data contained 3 cases scored/weighted as "1" which did not correspond with instrument. These cases were not included in the analyses.
[149] Bivariate correlation calculated using three-category variable corresponding with original weights.
[150] Bivariate correlation calculated using three-category variable corresponding with original weights.

Table 16. (cont.)
First RAR, Male – Outcome: Felony conviction, 18 month follow-up period

Summary statistics for entire scale[151]

Risk category	% recid.
Level 1 (Low-36)	12%
Level 2 (37-56)	27%
Level 3 (57-69)	28%
Level 4 (70-high)	34%

Chi-square = 89.412***

Zero-order correlation: r = .195***
AUC-ROC = .626
Alpha = .618

[151] Risk categories were based on quartiles.

Rating: Acceptable

Last RAR, Female juveniles, felony convictions

Table 17 contains the analyses examining the last RAR for female juveniles using felony conviction as the criteria for recidivism. Only two items revealed a statistically significant relationship with the outcome variable ("Violent offender" and "Prior admissions"). Both correlations were positive with adequate and appropriate differentiation regarding rates of the outcome. The remaining eight items revealed no relationship with the outcome variable.

The last RAR scale as a whole did reveal a statistically significant, positive, though weak correlation (r = .122) with the incidence of a felony conviction. The AUC-ROC was just above .600 (the Alpha score was unacceptably low). Despite the statistically significant correlation, the levels of risk did not differentiate adequately regarding the likelihood of recidivism. Most notably level 4 had lower recidivism rates than level 3, and rates of outcome were similar for levels 1 and 2 (9% and 10% respectively).

Rating: Poor

Last RAR, Male juveniles, felony convictions

Although the last RAR for males appears to have performed better than that for females when using felony conviction as the recidivism criteria, there were some deficiencies worthy of note (see Table 18). Two of the 12 items revealed no relationship with the outcome variable ("Placement" and "Age at release"). In addition, five more items revealed correlations less than .100 (albeit statistically significant and positive – see "Violent offender," "Prior commitment," "History of assault," "Age at first adjudication," and "Compliance within JRA"). Further, "Violent offender," "Placement" and "Age at first adjudication" showed limitations regarding the rates of recidivism across increasingly weighted categories.

Despite the aforementioned limitations, the scale as a whole did reveal a statistically significant zero-order correlation that was of moderate strength and positive (r = .186), and the AUC-ROC value was above .600. The risk levels (based on natural quartiles) displayed appropriately ascending rates of recidivism, though levels 3 and 4 were quite similar (28% and 29% respectively).

Rating: Mixed results

Table 17.
Last RAR, Female – Outcome: Felony conviction, 18 month follow-up period[152]

Item	r	chi-square	% recid.
Violent offender (n = 346)	.124*	5.320*	
Yes			11%
No			21%
Drug offender (n = 346)	.066 (n.s.)	1.506 (n.s.)	
No			17%
Yes			26%
Suicide[153] (n = 346)	-.086 (n.s.)	2.581 (n.s.)	
None/4			19%
1/2/3			11%
Gang member (n = 346)	.075 (n.s.)	1.932 (n.s.)	
No			16%
Yes			23%
Comm Fac. Incid. (n = 346)	.018[154] (n.s.)	.749 (n.s.)	
None			18%
Never/not scored			16%
One or more			22%
Prior admissions (n = 346)	.106*	3.920*	
None			16%
Not yet scored[155]			---
One or more			27%
Mental health needs (n = 342)	.104 (n.s.)	3.714 (n.s.)	
No			13%
Not yet scored[156]			---
Yes			21%
Chemical/alcohol use (n = 342)	.048 (n.s.)	.797 (n.s.)	
No			13%
Not yet scored[157]			---
Yes			18%

[152] n.s. indicates no relationship
* indicates statistically significant relationship at p < .05
** indicates statistically significant relationship at p < .01
*** indicates statistically significant relationship at p < .001
[153] 8 cases had values of "2" which were recoded to "20" (factor present) for this analysis.
[154] Bivariate correlation calculated using a categorized variable that corresponded with original weights.
[155] No cases had "not yet scored" as a value.
[156] Cases assessed as "not yet scored" were not included in the analyses.
[157] Cases assessed as "not yet scored" were not included in the analyses.

Table 17. (cont.)
Last RAR, Female – Outcome: Felony conviction, 18 month follow-up period

Item	r	chi-square	% recid.
Age at release (n = 346)	.033[158] (n.s.)	.693 (n.s.)	
Over 16			16%
15 or 16			19%
Under 15			17%
Savy points (n = 346)	.004 (n.s.)	.026 (n.s.)	
Vulnerable			17%
Neither			17%
Aggressive[159]			20%

Summary statistics for entire scale

Risk category	% recid.	
Level 1 (Low – 69)	9%	(7/75)
Level 2 (70 – 84)	10%	(9/87)
Level 3 (85 – 99)	30%	(24/81)
Level 4 (100 – High)	19%	(20/103)

Chi-square = 15.169**

Zero-order correlation: r = .122*
AUC-ROC = .606
Alpha = .204[160]

[158] Bivariate correlation calculated using a recoded variable with values 0, 1, 2 to correspond with ascending categories.
[159] Only 7 cases were assessed as "aggressive."
[160] Cronbach's alpha calculated using recoded categorical variables.

Table 18.
Last RAR, Male – Outcome: Felony conviction, 18 month follow-up period[161]

Item	r	chi-square	% recid.
Violent offender (n = 2808)	.084***	19.975***	
Yes			21%
No			28%
Gang member (n = 2808)	.119***	39.989***	
No			21%
Yes			31%
Placement (n = 2808)	.007[162] (n.s.)	11.574*	
Comm 90+			22%
Comm < 90			28%
Inst. 90 in comm			34%[163]
Inst. no comm			24%
Inst. Comm < 90			30%
Prior commitment (n = 2808)	.080***	17.800***	
None			23%
One or more			31%
History of assault (n = 2681[164])	.078***	16.415***[165]	
No			18%
Yes			26%
Age first adjudication (2808)	.049*[166]	9.708**	
16 +			19%
14-15			26%
13 -			26%
Chemical/Alc. Use (n = 2808)	.139***[167]	54.828***	
None			14%
Unknown/not avail			18%
Impairment			28%

[161] n.s. indicates no relationship
* indicates statistically significant relationship at p < .05
** indicates statistically significant relationship at p < .01
*** indicates statistically significant relationship at p < .001
[162] Bivariate correlation calculated using a categorical recoded variable.
[163] This percentage was based on 67 cases.
[164] The original variable had cases coded as "2" which were not part of the original assessment.
[165] Chi-square and percentages were calculated using a 2 category variable even though the original data contained 3 categories (0, 2, 3, as opposed to 0, 3).
[166] Bivariate correlation calculated using three-category variable corresponding with original weights.
[167] Bivariate correlation calculated using three-category variable corresponding with original weights.

Table 18. (cont.)
Last RAR, Male – Outcome: Felony conviction, 18 month follow-up period

Item	r	chi-square	% recid.
Prior adjudications (n = 2808)	.146***[168]	60.280***	
None			16%
One or two			24%
Three or more			30%
Compliance prior (n = 2808)	.102***[169]	29.256***	
High level			22%
Mod/not scored			27%
Non/minimal			34%
Sex offender (n = 2808)	.148***	61.337***	
Yes			12%
No			28%
Age at release (n = 2807[170])	-.013 (n.s.)[171]	3.532 (n.s.)	
Over 16			25%
15 or 16			26%
Under 15			20%
Compliance w/in JRA (n = 2808)	.078***[172]	17.295***	
High			20%
Mod/not scored			26%
Non/minimal			32%

Summary statistics for entire scale[173]

Risk category	% recid.
Level 1 (Low-36)	6%
Level 2 (37-56)	17%
Level 3 (57-69)	28%
Level 4 (70-high)	29%

Chi-square = 82.647***

Zero-order correlation: r = .186***
AUC-ROC = .622
Alpha = .568

[168] Bivariate correlation calculated using three-category variable corresponding with original weights.
[169] Bivariate correlation calculated using three-category variable corresponding with original weights.
[170] Original data contained 3 cases scored/weighted as "1" which did not correspond with instrument. These cases were not included in the analyses.
[171] Bivariate correlation calculated using three-category variable corresponding with original weights.
[172] Bivariate correlation calculated using three-category variable corresponding with original weights.
[173] Risk categories were based on quartiles.

First RAR, Female juveniles, violent felony conviction

Tables 19 through 22 examine the relationship between the RAR and violent felony conviction, with Table 19 focusing on female juveniles. None of the 10 items displayed a statistical relationship with the outcome variable (though it should be noted that in addition to the female juvenile sample being smaller to begin with, the rates of occurrence were very low for violence, which undoubtedly contributed to the lack of relationship). The scale as a whole did reveal a significant (though moderate) and positive correlation with outcome (r = .134) and a relatively high AUC-ROC value (.695). These seemingly positive results should be at least somewhat tempered by the rates of recidivism across levels of risk. Levels 2 and 3 fail to differentiate between the likelihood of recidivism, and level 4 has a lower rate than levels 2 and 3 (though again, very low base rates and low sample size likely influenced these results).

Rating: Poor

First RAR, Male juveniles, violent felony conviction

Table 20 contains the analyses examining the relationship between the first RAR and violent felony conviction as well, but does so for male juveniles. Somewhat more promising results were revealed, though again there are some deficiencies worthy of note. Four of the 12 items failed to reveal a statistically significant relationship with the outcome ("Violent offender," "Placement," "Age at first adjudication," and "Age at release"). Four more items ("Prior commitments," "History of assault," "Chemical/alcohol use," and "Prior adjudications") revealed statistically significant and positive relationships with outcome, though the magnitude was below .100 in each case. Of the aforementioned significant but weak correlations three ("Prior commitments," "Chemical/alcohol use," and "Prior adjudications") failed to adequately differentiate regarding rates of recidivism.

Despite the aforementioned limitations, the scale as a whole did reveal a statistically significant and positive (though moderate) correlation with violent felony conviction (r = .136), as well an AUC-ROC above .600. The risk levels (based on natural quartiles) revealed appropriately increasing rates of recidivism, though levels 2 and 3 were very narrow (11% and 12% respectively).

Rating: Mixed results (bordering on Poor)

Table 19.
First RAR, Female – Outcome: Violent conviction, 18 month follow-up period[174]

Item	r	chi-square	% recid.
Violent offender (n = 292)	-.029 (n.s.)	.245 (n.s.)	
No			6%
Yes			4%
Drug offender (n = 292)	.103 (n.s.)	3.110 (n.s.)	
No			4%
Yes			12%
Suicide[175] (n = 292)	.030 (n.s.)	.257 (n.s.)	
None/4			5%
1/2/3			7%
Gang member (n = 292)	.026 (n.s.)	.199 (n.s.)	
No			5%
Yes			6%
Comm Fac. Incid. (n = 292)	.028[176] (n.s.)	.414 (n.s.)	
None			0%[177]
Never/not scored			5%
One or more			0%[178]
Prior admissions (n = 292)	-.013 (n.s.)	.050 (n.s.)	
None			5%
Not yet scored[179]			---
One or more			4%
Mental health needs (n = 283)	.096 (n.s.)	2.599 (n.s.)	
No			3%
Not yet scored[180]			---
Yes			7%

[174] n.s. indicates no relationship
* indicates statistically significant relationship at p < .05
** indicates statistically significant relationship at p < .01
*** indicates statistically significant relationship at p < .001
[175] 8 cases had values of "2" which were recoded to "20" (factor present) for this analysis.
[176] Bivariate correlation calculated using a categorized variable that corresponded with original weights.
[177] Percentage based on 7 cases.
[178] Percentage based on 1 case.
[179] No cases had "not yet scored" as a value.
[180] Cases assessed as "not yet scored" were not included in the analyses.

Table 19. (cont.)
First RAR, Female – Outcome: Violent conviction, 18 month follow-up period

Item	r	chi-square	% recid.
Chemical/alcohol use (n = 282)	.068 (n.s.)	1.308 (n.s.)	
No			0%
Not yet scored[181]			---
Yes			5%
Age at release (n = 292)	.040[182] (n.s.)	.973 (n.s.)	
Over 16			4%
15 or 16			6%
Under 15			4%
Savy points (n = 292)	.088[183] (n.s.)	4.128 (n.s.)	
Vulnerable			0%
Neither			6%
Aggressive[184]			0%

Summary statistics for entire scale

Risk category	% recid.	
Level 1 (20 – 62)	0%	(0/69)
Level 2 (65 – 72)	6%	(1/18)
Level 3 (75 – 100)	6%	(8/144)
Level 4 (105 – 110)	3%	(1/30)
Level 5 (112 – 142)	13%	(4/31)

Chi-square = 8.286 (n.s.)

Zero-order correlation: r = .134*
AUC-ROC = .695
Alpha = .069[185]

[181] Cases assessed as "not yet scored" were not included in the analyses.
[182] Bivariate correlation calculated using a recoded variable with values 0, 1, 2 to correspond with ascending categories.
[183] Bivariate correlation calculated using a recoded variable with values 0, 1, 2 to correspond with ascending categories.
[184] Only 7 cases were assessed as "aggressive."
[185] Cronbach's alpha calculated using recoded categorical variables.

Table 20.
First RAR, Male – Outcome: Violent conviction, 18 month follow-up period[186]

Item	r	chi-square	% recid.
Violent offender (n = 2430)	.013 (n.s.)	.412 (n.s.)	
Yes			11%
No			12%
Gang member (n = 2430)	.148***	52.968***	
No			8%
Yes			18%
Placement (n = 2430)	.022[187] (n.s.)	4.463 (n.s.)	
Comm 90+			3%
Comm < 90			16%
Inst. 90 in comm			0%[188]
Inst. no comm			12%[189]
Inst. Comm < 90			0%[190]
Prior commitment (n = 2430)	.055***	7.306**	
None			11%
One or more			15%
History of assault (n = 2294[191])	.087***	17.469***[192]	
No			7%
Yes			14%
Age first adjudication (n = 2359)	.035 (n.s.)[193]	6.876*	
16 +			8%
14-15			13%
13 -			12%

[186] n.s. indicates no relationship
* indicates statistically significant relationship at p < .05
** indicates statistically significant relationship at p < .01
*** indicates statistically significant relationship at p < .001
[187] Bivariate correlation calculated using a categorical recoded variable.
[188] This percentage was based on 2 cases.
[189] 97% of the cases were assessed as scoring in this category.
[190] This percentage was based on 5 cases.
[191] The original variable had cases coded as "2" which were not part of the original assessment.
[192] Chi-square and percentages were calculated using a 2 category variable even though the original data contained 3 categories (0, 2, 3, as opposed to 0, 3).
[193] Bivariate correlation calculated using three-category variable corresponding with original weights.

Table 20. (cont.)
First RAR, Male – Outcome: Violent conviction, 18 month follow-up period

Item	r	chi-square	% recid.
Chemical/Alc. Use (n = 2430)	.064**[194]	10.363**	
None			8%
Unknown/not avail			9%
Impairment			13%
Prior adjudications (n = 2430)	.080***[195]	16.908***	
None			9%
One or two			9%
Three or more			14%
Compliance prior (n = 2430)	.104***[196]	26.082***	
High level			9%
Mod/not scored			14%
Non/minimal			19%
Sex offender (n = 2430)	.109***	29.085***	
Yes			5%
No			14%
Age at release (n = 2428[197])	-.011 (n.s.)[198]	5.747 (n.s.)	
Over 16			11%
15 or 16			13%
Under 15			8%
Compliance w/in JRA (n = 2430)	.088***[199]	18.622***	
High			8%
Mod/not scored			13%
Non/minimal			18%

[194] Bivariate correlation calculated using three-category variable corresponding with original weights.
[195] Bivariate correlation calculated using three-category variable corresponding with original weights.
[196] Bivariate correlation calculated using three-category variable corresponding with original weights.
[197] Original data contained 3 cases scored/weighted as "1" which did not correspond with instrument. These cases were not included in the analyses.
[198] Bivariate correlation calculated using three-category variable corresponding with original weights.
[199] Bivariate correlation calculated using three-category variable corresponding with original weights.

Table 20. (cont.)
First RAR, Male – Outcome: Violent conviction, 18 month follow-up period

Summary statistics for entire scale[200]

Risk category	% recid.
Level 1 (Low-36)	6%
Level 2 (37-56)	11%
Level 3 (57-69)	12%
Level 4 (70-high)	17%

Chi-square = 41.132***

Zero-order correlation: r = .136***
AUC-ROC = .621
Alpha = .618

[200] Risk categories were based on quartiles.

Last RAR, Female juveniles, violent felony conviction

Table 21 reveals for female juveniles nearly identical results as those contained in Table 19 when examining the relationship between the last RAR and violent felony conviction. None of the 10 items revealed a statistical relationship with the outcome variable, though some of the response categories for some of the items differentiated between rates of outcome appropriately. The scale as a whole did not have a relationship with the outcome variable, and the risk levels likewise did not reveal appropriately increasing rates of recidivism.

Rating: Poor

Last RAR, Male juveniles, violent felony conviction

Although performing somewhat better for male juveniles (relative to female juveniles), the last RAR revealed some deficiencies when testing the relationship with violent felony conviction (see Table 22). Three of the 12 items ("Violent offender," "Placement" and "Age at release") held no relationship with the outcome variable. Of those items that did reveal a statistically significant relationship with new violent conviction, six were of a magnitude less than .100 ("Prior commitment," "History of assault," "Age at first adjudication," "Chemical/alcohol use," "Prior adjudications," and "Compliance within JRA").

Despite the aforementioned deficiencies the scale as whole performed satisfactorily when predicting new violent felony conviction. Specifically a statistically significant and positive zero-order correlation of moderate strength was revealed (r = .141) and the AUC-ROC value was above .600. In addition the rates of recidivism ascended appropriately across levels of risk.

Rating: Mixed results (bordering on Poor).

Comparison of recidivism rates across risk categories, race, and sex

Table 23 is designed to allow for a comparison of recidivism/isolation rates between white and non-white youth for male and female juveniles separately, who fall in similar risk categories for each instrument under consideration (first and last RAR and the first and last RAI). The table is designed primarily for description and should not be regarded as a test of bias regarding the classification of youth using the risk assessments. However there may be value in noting differences in recidivism and/or isolation rates for sub-groups of youth who are classified similarly in terms of risk. It is important to bear in mind that the measure of recidivism

Table 21.
Last RAR, Female – Outcome: Violent conviction, 18 month follow-up period[201]

Item	r	chi-square	% recid.
Violent offender (n = 346)	-.008 (n.s.)	.0235 (n.s.)	
No			4%
Yes			5%
Drug offender (n = 346)	.097 (n.s.)	3.242 (n.s.)	
No			4%
Yes			11%
Suicide[202] (n = 346)	-.031 (n.s.)	.335 (n.s.)	
None/4			5%
1/2/3			3%
Gang member (n = 346)	.018 (n.s.)	.111 (n.s.)	
No			4%
Yes			5%
Comm Fac. Incid. (n = 346)	-.003[203] (n.s.)	.365 (n.s.)	
None			4%
Never/not scored			5%
One or more			2%
Prior admissions (n = 346)	-.010 (n.s.)	.035 (n.s.)	
None			4%
Not yet scored[204]			---
One or more			4%
Mental health needs (n = 342)	.034 (n.s.)	.398 (n.s.)	
No			4%
Not yet scored[205]			---
Yes			5%
Chemical/alcohol use (n = 342)	.078 (n.s.)	2.078 (n.s.)	
No			0%
Not yet scored[206]			---
Yes			5%

[201] n.s. indicates no relationship
* indicates statistically significant relationship at p < .05
** indicates statistically significant relationship at p < .01
*** indicates statistically significant relationship at p < .001
[202] 8 cases had values of "2" which were recoded to "20" (factor present) for this analysis.
[203] Bivariate correlation calculated using a categorized variable that corresponded with original weights.
[204] Cases assessed as "not yet scored" were not included in the analyses.
[205] Cases assessed as "not yet scored" were not included in the analyses.
[206] Cases assessed as "not yet scored" were not included in the analyses.

Table 21. (cont.)
First RAR, Female – Outcome: Violent conviction, 18 month follow-up period

Item	r	chi-square	% recid.
Age at release (n = 346)	.021[207] (n.s.)	.600 (n.s.)	
Over 16			4%
15 or 16			5%
Under 15			3%
Savy points (n = 346)	.054[208] (n.s.)	1.758 (n.s.)	
Vulnerable			2%
Neither			5%
Aggressive[209]			0%

Summary statistics for entire scale

Risk category	% recid.	
Level 1 (Low – 69)	1%	(1/75)
Level 2 (70 – 84)	2%	(2/87)
Level 3 (85 – 99)	9%	(7/81)
Level 4 (100 - high)	5%	(5/103)

Chi-square = 6.189 (n.s.)

Zero-order correlation: r = .067 (n.s.)
AUC-ROC = .626
Alpha = .204[210]

[207] Bivariate correlation calculated using a recoded variable with values 0, 1, 2 to correspond with ascending categories.
[208] Bivariate correlation calculated using a recoded variable with values 0, 1, 2 to correspond with ascending categories.
[209] Only 5 cases were assessed as "aggressive."
[210] Cronbach's alpha calculated using recoded categorical variables.

Table 22.
Last RAR, Male – Outcome: Violent conviction, 18 month follow-up period[211]

Item	r	chi-square	% recid.
Violent offender (n = 2808)	.016 (n.s.)	.752 (n.s.)	
Yes			11%
No			12%
Gang member (n = 2808)	.150***	63.518***	
No			8%
Yes			18%
Placement (n = 2808)	.006[212] (n.s.)	6.516 (n.s.)	
Comm 90+			12%
Comm < 90			9%
Inst. 90 in comm			19%
Inst. no comm			12%
Inst. Comm < 90			11%
Prior commitment (n = 2808)	.069***	13.440***	
None			10%
One or more			16%
History of assault (n = 2681[213])	.092***	22.543***[214]	
No			6%
Yes			13%
Age first adjudication (n = 2808)	.050**[215]	10.419**	
16 +			7%
14-15			12%
13 -			13%
Chemical/Alc. Use (n = 2808)	.073***[216]	15.213***	
None			7%
Unknown/not avail			9%
Impairment			13%

[211] n.s. indicates no relationship
* indicates statistically significant relationship at p < .05
** indicates statistically significant relationship at p < .01
*** indicates statistically significant relationship at p < .001
[212] Bivariate correlation calculated using a categorical recoded variable.
[213] The original variable had cases coded as "2" which were not part of the original assessment.
[214] Chi-square and percentages were calculated using a 2 category variable even though the original data contained 3 categories (0, 2, 3, as opposed to 0, 3).
[215] Bivariate correlation calculated using three-category variable corresponding with original weights.
[216] Bivariate correlation calculated using three-category variable corresponding with original weights.

Table 22. (cont.)
Last RAR, Male – Outcome: Violent conviction, 18 month follow-up period

Item	r	chi-square	% recid.
Prior adjudications (n = 2808)	.090***[217]	23.285***	
None			8%
One or two			10%
Three or more			14%
Compliance prior (n = 2808)	.102***[218]	29.338***	
High level			9%
Mod/not scored			13%
Non/minimal			19%
Sex offender (n = 2808)	.102***	29.254***	
Yes			5%
No			13%
Age at release (n = 2807[219])	-.012 (n.s.)[220]	6.402*	
Over 16			11%
15 or 16			13%
Under 15			8%
Compliance w/in JRA (n = 2808)	.054**[221]	8.861*	
High			10%
Mod/not scored			12%
Non/minimal			16%

Summary statistics for entire scale

Risk category	% recid.
Level 1 (Low-13)	0%
Level 2 (14-47)	8%
Level 3 (48-53)	11%
Level 4 (54-70)	13%
Level 5 (71-high)	18%

Chi-square = 49.893***
Zero-order correlation: r = .141***
AUC-ROC = .625
Alpha = .568

[217] Bivariate correlation calculated using three-category variable corresponding with original weights.
[218] Bivariate correlation calculated using three-category variable corresponding with original weights.
[219] Original data contained cases scored/weighted as "1" which did not correspond with instrument. These cases were not included in the analyses.
[220] Bivariate correlation calculated using three-category variable corresponding with original weights.
[221] Bivariate correlation calculated using three-category variable corresponding with original weights.

Table 23. Recidivism/failure rates by sex and category

	Any recid.	Fel recid.	Violent recid.
First RAR			
Male -> White -> Lowest	22%	7%	4%
Male -> Non-white -> Lowest	33%	16%	8%
Female -> White -> Lowest	43%	7%	0%
Female -> Non-white -> Lowest	37%	10%	0%
Male -> White -> Highest	71%	32%	12%
Male -> Non-white -> Highest	71%	36%	21%
Female -> White -> Highest	62%	23%	8%
Female -> Non-white -> Highest	67%	28%	17%
Last RAR			
Male -> White -> Lowest	17%	6%	0%
Male -> Non-white -> Lowest	18%	6%	0%
Female -> White -> Lowest	44%	13%	3%
Female -> Non-white -> Lowest	35%	7%	0%
Male -> White -> Highest	76%	34%	14%
Male -> Non-white -> Highest	72%	34%	21%
Female -> White -> Highest	36%	10%	3%
Female -> Non-white -> Highest	63%	25%	6%

	Any isolation
First RAI	
Male -> White -> Lowest	12%
Male -> Non-white -> Lowest	17%
Female -> White -> Lowest	19%
Female -> Non-white -> Lowest	20%
Male -> White -> Highest	54%
Male -> Non-white -> Highest	59%
Female -> White -> Highest	25%
Female -> Non-white -> Highest	11%

Table 23. (cont.) Recidivism/failure rates by sex and category

	Any isolation
Last RAI	
Male -> White -> Lowest	12%
Male -> Non-white -> Lowest	16%
Female -> White -> Lowest	14%
Female -> Non-white -> Lowest	17%
Male -> White -> Highest	55%
Male -> Non-white -> Highest	51%
Female -> White -> Highest	0%
Female -> Non-white -> Highest	0%

(conviction) could be influenced by any number of factors not having to do with individual behavior and/or criminogenic risk. In addition, the number of cases in some of the categories for female youth were very low, making the percentage estimates potentially unstable.

False positive ratios

Table 24 contains an approximation of false positive ratios. False positive ratios were calculated by dividing the number of cases that were deemed false positives, by the sum of false positives and cases that were deemed true negatives. A case was considered a false positive if it had been classified via the instrument as being in the highest risk category, but did not recidivate. A case was considered a true negative if it had been classified via the instrument as being in the lowest risk category, but likewise did not recidivate. The method used is conservative in that only the extreme categorizations were used, and all subgroups (by race and sex) had the same method applied. As such, Table 24 may allow for the comparison of subgroups as it relates to the likelihood of false positive classification.

Overall, male youth appear to have more disparity between white and non-white juveniles regarding false positives classification. Based on the totality of the results, it appears possible that non-white youth are being over-classified in some instances by the instrument (meaning they are placed in a higher risk category than they might actuarially belong in). Additional analysis and data elements may be necessary in order to draw firmer conclusions.

Conclusions and next steps

The vast majority of the analyses presented above appear to indicate that the RAR and the RAI in their current form do not hold an adequate amount of statistical validity for the female youth in JR. The bivariate analyses that focused on female youth largely revealed that individual items very often lacked any relationship with outcome, and the scales as a whole at best possessed weak relationships with the outcome. Multivariate logistic regression (not appearing in tables) further revealed that the first RAR was not at all predictive of the likelihood of recidivism (though race was), while the second RAR was predictive, though in a weak fashion. The logistic regression analyses support the conclusions drawn from the bivariate analyses appearing in the tables.

Table 24. False positive ratios[222]

	Any recid.	Fel recid.	Violent recid.
First RAR			
White male	22%	36%	42%
Non-white male	36%	51%	54%
Last RAR			
White male	66%	82%	84%
Non-white male	88%	94%	95%
First RAR			
White female	24%	28%	30%
Non-white female	19%	26%	27%
Last RAR			
White female	58%	56%	55%
Non-white female	46%	55%	58%
	Any isolation		
First RAI			
White male	11%		
Non-white male	23%		
Last RAI			
White male	8%		
Non-white male	20%		
First RAI			
White female	19%		
Non-white female	40%		
Last RAI			
White female	14%		
Non-white female	21%		

[222] False positive ratios were calculated by first identifying for each group the number of true negatives and false positives. True negatives were the number of cases that were classified as being in the lowest category of risk, who did not recidivate. False positives were the number of cases that were classified as being in the highest category of risk, who did not recidivate. The number of false positives were divided by the number of true negatives plus false positives in order to establish the ratio of false positives to N.

As noted above, it is possible that the analyses for female youth were impacted by relatively small sample sizes, and as such, additional datasets (if available) may be worth pursuing to verify the conclusion that the instruments generally (and severely) lack predictive validity.

The RAR and RAI instruments appear to hold more validity for male youth. Statistically significant findings were commonplace across several different outcome measures. However, as highlighted above in several analyses individual items were found to lack statistical relationships with one or more outcome variables, and of those that did reveal relationships some were not robust. As such further analyses may be warranted in order to fine-tune the instruments in order to determine whether different weighting schemes would improve the instrument, or elimination (or replacement) of certain items might be in order. Interestingly, for male youth, binary logistic regression modeling revealed the instruments to be statistically significant when predicting the likelihood of recidivism/failure in many instances, while controlling for race (which was not statistically significant).

Next steps

In light of the totality of the results presented above, the following recommendations are put forth:

- Obtain additional data elements that will allow for the testing of the RACF.
- If necessary re-run analyses using the risk categories that are in use as opposed to the natural quartiles where noted.
- If a separate risk assessment is deemed necessary, consider re-tooling or replacing the RAR for female youth.
- If a separate risk assessment is deemed necessary, consider re-tooling or replacing the RAR for male youth (though this recommendation is not as urgent as that for the female population).
- Consider making all risk/need (and potentially some responsivity) decisions on the ITA.

Testing the statistical validity of the RACF

An addendum submitted to Washington State Juvenile Rehabilitation

By

Alexander M. Holsinger, Ph.D.

Kristi Holsinger, Ph.D.

August 30, 2020

Testing the statistical validity of the RACF

Introduction

This technical report utilizes data obtained from the State of Washington's Juvenile Rehabilitation (JR) agency that allowed for tests of the statistical validity of the Risk Assessment-Community Facilities (hereafter RACF) instrument. This report is submitted as an addendum[1] to a previous report that tested the statistical validity of the Risk Assessment-Recidivism (RAR) and the Risk Assessment-Institutional (RAI). At the time of the previous analyses and reports, the RACF data were not available, but have since been obtained and constitute the totality of the analyses presented below.

While the RAR and the RAI were designed to assess the likelihood of recidivism and institutional misconduct respectively, the RACF is unique in that its purpose involves assessing the likelihood that a youth will be sent back to a JR institution after spending time in a Community Facility as part of their obligation. As a result the RACF could influence the way a youth is treated in the community facility, and inform any number of decisions regarding their case.

The analyses below are presented in disaggregated form regarding sex, which will allow for a comparison between boys and girls. Each of the individual items that make up the RACF are tested for their relationship with the relevant outcome (for these analyses the outcomes involve whether not a youth is sent back to an institution from a community facility as well as how many times a youth is sent back). In addition the performance of the scale as a whole is tested using appropriate statistical analyses.

Substantially fewer cases were utilized for this addendum (compared to the report analyzing the RAR and the RAI) since not all youth spend time at a community facility as part of their obligation. Only youth who spent time at a community facility, and who likewise had a RACF completed were selected for inclusion in the analyses. This resulted in a sample size of 2,045 youth who had been released during the years 2010 to 2017. Of the total sample (n = 2,045), 1,817 were boys (89%) and 228 were girls (11%). In addition, 903 (44%) were white, while 1,142 (56%) were non-white. Due to the relatively small number of girls in the analyses, many of the

[1] See "Testing the statistical validity of the RAR and the RAI" for detail regarding how to view statistical significance, as well as additional detail regarding specific tests utilized. Also included are descriptions of various types of validity and the role they play in the development and testing of risk assessments of all sorts.

results should be interpreted with caution. Small sample sizes mean that statistical estimates (specific tests as well as percentages and the like) will be unstable and lack generalizability. Nonetheless it can be helpful to see quantitative results disaggregated by sex (as well as other groupings) so observations can be made as to how overall results involving the entire sample can mask ineffective performance for some groups.

In keeping with the previously issued report, the performance of the RACF when predicting the likelihood a youth will be sent back to an institution from a community facility is assessed as being "Acceptable" (not necessarily perfect performance, but at least adequate without glaring problems or deficiencies), "Mixed results" (some indications of effectiveness, but clear room for improvement if not issues that need to be addressed), or "Poor" (tool should be discontinued until it has been fixed; consider adopting something with proven validity and test it prospectively going forward).

Results

Table 1 presents the results for the zero-order correlations that test the relationship between each individual item on the RACF and the binary outcome (youth being sent back to an institution or not). The scale as a whole is tested using zero-order correlations as well. For boys, eight of the 14 items revealed a statistically significant relationship with the outcome. While all the statistically significant relationships were in the 'right' direction (meaning the relationship was positive, which indicates the category that provides more points to the total risk score had a higher failure rate), the majority of them (5) under-performed, revealing a value less than .100 (indicating a relationship that is weaker than it should be). The scale as a whole had a statistically significant and sufficiently strong relationship with the outcome (r = .153). For girls only two (of 14) items maintained a statistically significant relationship with the outcome (prior commitments and history of mental health placement at JRA). In addition, one of the two items (prior commitments) maintained a stronger relationship (r = .225) than the scale as a whole (r = .201) which was statistically significant as well. These results might indicate that a singular item (prior commitments) might perform better or at least more reliably than using the entire scale. The analyses using the scale as a whole (see Total column, Table 1) largely reflect what was revealed

Table 1. zero-order correlations – items and scale x binary outcome; default scoring included[2]

Item	Boys	Girls	Total
Progress w/facility requirements	.069*	.077 n.s.	.071**
Problem solving skills	.035 n.s.	.021 n.s.	.037 n.s.
Hostile response to frustration	.063*	.030 n.s.	.060*
Prior adjudications	.086**	.047 n.s.	.081**
Compliance with facility regulations	.158***	.121 n.s.	.152***
History of escapes	.103***	-.041 n.s.	.087**
Prior commitments	.106***	.225**	.118**
Sex offender	.099***	.027 n.s.	.097***
Mental health needs	.022 n.s.	.098 n.s.	.031 n.s.
History mental health placement at JRA	.027 n.s.	.165*	.058*
Manifest injustice	-.001 n.s.	.037 n.s.	.001 n.s.
Age at admission	.014 n.s.	.022 n.s.	.016 n.s.
Scored as aggressive on SAVY	-.024 n.s.	.018 n.s.	-.023 n.s.
Chemical/alcohol use	.057*	.095 n.s.	.062*
Total score	.153***	.201**	.161***

[2] n.s. = no relationship;
* = significant relationship, p < .05
** = significant relationship, p < .01
*** = significant relationship, p < .001

when isolating the boys. Taken as a whole, Table 1 indicates that the RACF is performing with "Mixed results" bordering on "Poor" for boys, and is clearly "Poor" for girls.

Table 2 likewise presents a test of each individual item and its relationship with the binary outcome, but displays the percentage of cases that were sent back to an institution for each category contained in the item (also contained in Table 2 are additional analyses testing the entire scale, and the rates of outcome for each category of risk). The analyses in Table 2 utilized the items as they were received, which meant "default scoring" categories were included and contributed to the total score. Default scoring results when the answer to an item is unknown, but points are assigned regardless. Ideally for each item the rates of the outcome occurring should increase substantially for each category (since each ascending category assigns an increasing number of points to the total risk score). In short, the categories for each item should substantially differentiate between the likelihood of success and failure (as should the risk categories rendered from the scale as a whole – "Very Low" through "Very High").

For boys, it is again revealed that several items (6 of 14) did not maintain a statistical relationship with the outcome. Moreover, regardless of statistical significance it appears that the "default scoring" categories are causing some incongruous results. For example, the item "Problem solving skills" has a value of "3" in between "Usually appropriate response to problems" and "Rarely appropriate response to problems" but the rates of outcome do not vary much between "Usually appropriate…" (37%) and "3" (39%). The same can be observed for the item "Hostile response to frustration" where the category "Usually does not act out" has an outcome rate of 37% while the default category "2" has an outcome rate of 38%, with the rate of outcome jumping to 54% for "Frequent hostile response." The issue becomes even more marked when examining the item "History of escapes" where the category "None" has an outcome rate of 35% but the default category "3" has a lower rate of 33%. Even in the statistically significant items that do not contain default scoring categories there are problems with outcome rate differentiation (see for example the categories "One or two" vs. "Three or more" for the item "Prior adjudications," the categories "One" vs. "Two or more" for the item "Prior commitments," as well as the item "Mental health needs"). When examining the scale as a whole and the risk categories rendered ("Very Low" through "Very High"), it is a good sign that the rates

Table 2. chi-square and percentage failure – items/score x outcome; default scoring included[3]

Item/categories	Boys	Girls	Total
Progress w/facility core requirements			
High level of compliance	34%	37%	34%
Moderate compliance	40%	46%	40%
No or minima compliance	47%	46%	47%
Chi-square test	6.45*	1.25 n.s.	7.73*
Problem solving skills			
Usually appropriate response to problems	37%	46%	38%
"3"	39%	29%	38%
Rarely appropriate response to problems	45%	50%	47%
Chi-square test	1.76*	3.57 n.s.	3.10 n.s.
Hostile response to frustration			
Usually does not act out	37%	47%	38%
"2"	38%	28%	37%
Frequent hostile response	54%	57%	54%
Chi-square test	5.99*	5.12 n.s.	7.59*
Prior adjudications			
None	30%	42%	32%
One or two	39%	35%	39%
Three or more	41%	46%	41%
Chi-square test	10.85**	1.14 n.s.	10.46**
Compliance with facility regulations			
High level of compliance	33%	39%	33%
"2"	43%[4]	100%[5]	56%
Moderate compliance	47%	55%	48%
No or minimal compliance	54%	50%	54%
Chi-square test	34.86***	5.48 n.s.	37.45***

[3] n.s. = no relationship;
* = significant relationship, p < .05
** = significant relationship, p < .01
*** = significant relationship, p < .001
[4] % based on 7 cases.
[5] % based on 2 cases.

Table 2. (cont.) chi-square and percentage failure – items/score x outcome; default scoring included

Item/categories	Boys	Girls	Total
History of escapes			
None	35%	45%	36%
"3"	33%[6]	40%[7]	35%
Left court-ordered placement/escaped	48%	40%	47%
Chi-square test	15.00**	.29 n.s.	12.13**
Prior commitments			
None	35%	38%	35%
One	48%	58%	49%
Two or more	49%	100%[8]	52%
Chi-square test	17.06***	8.81*	23.01***
Sex offender			
Current or historic	27%	33%[9]	27%
Not a sex offender	40%	44%	41%
Chi-square test	13.23***	.125 n.s.	14.26***
Mental health needs			
No	37%	40%	38%
Yes	40%	50%	41%
Chi-square test	.676 n.s.	1.59 n.s.	1.50 n.s.
History mental health placement at JRA			
No	38%	24%[10]	37%
Yes	43%	47%	46%
Chi-square test	.98 n.s.	4.50*	5.19*
Manifest injustice up or in			
No	38%	43%	39%
Yes	38%	50%	39%
Chi-square test	.001 n.s.	.231 n.s.	.002 n.s.
Age at admission			
Over 16	38%	42%	38%
15 or 16	38%	44%	39%
Under 15	41%	47%[11]	41%
Chi-square test	.452 n.s.	.09 n.s.	.54 n.s.

[6] % based on 18 cases.
[7] % based on 5 cases.
[8] % based on 3 cases.
[9] % based on 3 cases.
[10] % based on 25 cases.
[11] % based on 15 cases.

Table 2. (cont.) chi-square and percentage failure – items/score x outcome; default scoring included

Item/categories	Boys	Girls	Total
Scored as aggressive on SAVY			
No	39%	43%	39%
"1"	27%	100%[12]	28%
Yes	34%	50%[13]	34%
Chi-square test	3.34 n.s.	1.35 n.s.	2.98 n.s.
Chemical/alcohol use			
No	31%	27%[14]	30%
"7"	36%	60%[15]	37%
Yes	39%	45%	40%
Chi-square test	4.47 n.s.	2.35 n.s.	5.89 n.s.
Summary statistics for entire scale			
Risk category			
Very Low	28%	0%[16]	28%
Low	38%	31%	38%
Moderate	45%	40%	44%
High	46%	56%	50%
Very High	67%[17]	53%[18]	60%
Chi-square test	27.26***	9.99*	34.39***
Zero-order correlation	.153***	.201**	.161***
AUC-ROC	.592	.629	.596
Alpha	.323	.334	.312

[12] % based on 1 case.
[13] % based on 2 cases.
[14] % based on 15 cases.
[15] % based on 5 cases.
[16] % based on 5 cases.
[17] % based on 18 cases.
[18] % based on 19 cases.

of outcome do increase from one category to the next, however, it is important to note that there is negligible differentiation between the categories “Moderate” and “High.” In addition, the AUC-ROC (a measure of how well the scale as a whole differentiates between successes and failures) is below .600 (and ideally would be .700 or above), and the Alpha value (a measure of the items’ covariance as a whole – meaning the extent to which they are assessing similar constructs) is well below what is considered acceptable (.600 or higher).

For girls, only two items maintained a statistically significant relationship with the outcome (“Prior commitments” and “History of mental health placement at JRA”). In both instances very good differentiation across categories was revealed, but was likewise based on very few cases in some instances. The performance of the scale overall was not adequate, though two of the risk categories (“Very Low” and “Very High”) had very few cases making the percentages unstable. Nonetheless the rates of outcome decreased with moving from “High” (56%) to “Very High” (53%). Interestingly the zero-order correlation, as noted before in Table 1 was above .200 and statistically significant, and the AUC-ROC was likewise above .600 (though the Alpha score was below the threshold for acceptable). Regardless of these summary statistics, it could be that decisions based on just one item (for example “Prior commitments”) would provide better performance in the long run than the scale as a whole. These results (Table 2) reinforce the conclusions rendered from Table 1 – “Mixed results” bordering on “Poor” performance for boys, and “Poor” performance for girls, in light of the number of items that reveal no statistical relationship with outcome.

The analyses displayed in Table 3 were produced in order to determine what would happen if the default scoring categories were eliminated. Note that the bold-typed items are ones that were changed, eliminating the default scoring (i.e., the cases that had default scoring categories were dropped from the analyses). Regarding the individual items for boys, the performance, based on the zero-order correlations appears to be improved some, but not remarkably so. For example the item “Hostile response to frustration” increased from .063 to .076, “History of escapes” had an increase from .103 to .105, and “Chemical/alcohol use” increased from .057 to .058. Overall the strength of the relationship for the total score increased from .153 (scale rendered using items with default scoring categories) to .170 (scale rendered

Table 3. zero-order correlations – items and scale x binary outcome; no default scoring[19],[20]

Item	Boys	Girls	Total
Progress w/facility requirements	.069*	.077 n.s.	.071**
Problem solving skills	**.040 n.s.**	**.036 n.s.**	**.050 n.s.**
Hostile response to frustration	**.076***	**.066 n.s.**	**.079****
Prior adjudications	.086**	.047 n.s.	.081**
Compliance with facility regulations	**.158*****	**.114 n.s.**	**.152*****
History of escapes	**.105*****	**-.040 n.s.**	**.089****
Prior commitments	.106***	.225**	.118**
Sex offender	.099***	.027 n.s.	.097***
Mental health needs	.022 n.s.	.098 n.s.	.031 n.s.
History mental health placement at JRA	.027 n.s.	.165*	.058*
Manifest injustice	-.001 n.s.	.037 n.s.	.001 n.s.
Age at admission	.014 n.s.	.022 n.s.	.016 n.s.
Scored as aggressive on SAVY	**-.024 n.s.**	**.016 n.s.**	**-.024 n.s.**
Chemical/alcohol use	**.058***	**.105 n.s.**	**.064***
Total score	**.170*****	**.199***	**.184*****

[19] n.s. = no relationship;
* = significant relationship, p < .05
** = significant relationship, p < .01
*** = significant relationship, p < .001
[20] Bold-faced items had default scores that were eliminated for these analyses. Light-faced items are based on the same variables that appeared in Table 1. Total score correlations were re-calculated as the RACF total score was re-calculated using the revised variables that eliminated default scoring.

using items with default scoring eliminated). For girls, the elimination of default scoring categories did nothing to improve the relationships between the items and the outcome. The relationship between the total score and the outcome for girls while maintaining statistical significance, actually decreased in strength slightly (from .201 to .199). In sum, the presence of the default scoring categories do not appear to be helping nor harming the performance of the items or the scale as a whole, with some incongruous results noted above.

Table 4 displays the percentage outcome for each item in the scale as well as summary information for the scale as a whole, however, like Table 3 the default scoring categories were eliminated from the applicable items, and the total score and risk categories were reproduced using the new information. For the statistically significant items for boys (meaning the 'new' items that had default score categories eliminated) differentiation in percentage outcome appears to be good. For example "Compliance with facility regulations" goes from 33% ("High level…") to 47% ("Moderate…") to 54% ("No or minimal…"). For "History of escapes" rates of outcome go from 35% ("None") to 48% ("…escaped"). In addition, the rates of outcome increase for each category of risk (from "Very Low" to "Very High" though it should be noted the lattermost category was based on 12 cases), and the total correlation was acceptable (.170) as was the AUC-ROC (while the Alpha remained abysmally low). Overall however, for boys the performance of the scale appears to render "Mixed results" bordering on "Poor" particularly considering nearly half the items are not revealing a detectable relationship with the outcome.

For girls, the elimination of the default scoring categories did nothing to improve the performance of any of the items, nor the performance of the scale as a whole (though again low frequency of cases for several categories should be noted). Despite what ordinarily would be considered an adequate overall correlation (.199) and an acceptable AUC-ROC value (.625), the scale should be considered to perform at a level of "Poor" in light of the fact that all but two items revealed no relationship with the outcome. Additional cases and future analyses might render different results.

Table 4. chi-square and percentage failure – items/score x outcome; no default scoring[21],[22]

Item/categories	Boys	Girls	Total
Progress w/facility core requirements			
High level of compliance	34%	37%	34%
Moderate compliance	40%	46%	40%
No or minima compliance	47%	46%	47%
Chi-square test	6.45*	1.25 n.s.	7.73*
Problem solving skills			
Usually appropriate response to problems	**37%**	**46%**	**38%**
Rarely appropriate response to problems	**45%**	**50%**	**47%**
Chi-square test	**1.70 n.s.**	**.169 n.s.**	**2.95 n.s.**
Hostile response to frustration			
Usually does not act out	**37%**	**47%**	**38%**
Frequent hostile response	**54%**	**57%**	**54%**
Chi-square test	**5.99***	**.562 n.s.[23]**	**7.15****
Prior adjudications			
None	30%	42%	32%
One or two	39%	35%	39%
Three or more	41%	46%	41%
Chi-square test	10.85**	1.14 n.s.	10.46**
Compliance with facility regulations			
High level of compliance	**33%**	**39%**	**33%**
Moderate compliance	**47%**	**55%**	**48%**
No or minimal compliance	**54%**	**50%[24]**	**54%**
Chi-square test	**34.80*****	**2.85 n.s.**	**36.39*****
History of escapes			
None	**35%**	**45%**	**36%**
Left court-ordered placement/escaped	**48%**	**40%**	**47%**
Chi-square test	**14.82*****	**.262 n.s.**	**11.98****

[21] n.s. = no relationship;
* = significant relationship, p < .05
** = significant relationship, p < .01
*** = significant relationship, p < .001
[22] Bold-faced items had default scores that were eliminated for these analyses. Light-faced items are based on the same variables that appeared in Table 1. Total score correlations were re-calculated as the RACF total score was re-calculated using the revised variables that eliminated default scoring.
[23] % based on 14 cases.
[24] % based on 10 cases.

Table 4. (cont.) chi-square and percentage failure – items/score x outcome; no default scoring

Item/categories	Boys	Girls	Total
Prior commitments			
None	35%	38%	35%
One	48%	58%	49%
Two or more	49%	100%[25]	52%
Chi-square test	17.06***	8.81*	23.01***
Sex offender			
Current or historic	27%	33%[26]	27%
Not a sex offender	40%	44%	41%
Chi-square test	13.23***	.125 n.s.	14.26***
Mental health needs			
No	37%	40%	38%
Yes	40%	50%	41%
Chi-square test	.676 n.s.	1.59 n.s.	1.50 n.s.
History mental health placement at JRA			
No	38%	24%[27]	37%
Yes	43%	47%	46%
Chi-square test	.98 n.s.	4.50*	5.19*
Manifest injustice up or in			
No	38%	43%	39%
Yes	38%	50%	39%
Chi-square test	.001 n.s.	.231 n.s.	.002 n.s.
Age at admission			
Over 16	38%	42%	38%
15 or 16	38%	44%	39%
Under 15	41%	47%[28]	41%
Chi-square test	.452 n.s.	.09 n.s.	.54 n.s.
Scored as aggressive on SAVY			
No	**39%**	**43%**	**39%**
Yes	**34%**	**50%[29]**	**34%**
Chi-square test	**.785 n.s.**	**.04 n.s.**	**.84 n.s.**

[25] % based on 3 cases.
[26] % based on 3 cases.
[27] % based on 25 cases.
[28] % based on 15 cases.
[29] % based on 2 cases.

Table 4. (cont.) chi-square and percentage failure – items/score x outcome; no default scoring

Item/categories	Boys	Girls	Total
Chemical/alcohol use			
No	**31%**	**27%**[30]	**30%**
Yes	**39%**	**45%**	**40%**
Chi-square test	**4.16***	**1.77 n.s.**	**5.67***
Summary statistics for entire scale			
Risk category			
Very Low	**28%**	**0%**[31]	**28%**
Low	**36%**	**32%**[32]	**36%**
Moderate	**47%**	**44%**	**46%**
High	**53%**	**62%**	**57%**
Very High	**67%**[33]	**53%**[34]	**59%**[35]
Chi-square test	**22.7*****	**7.91 n.s.**	**32.89*****
Zero-order correlation	**.170*****	**.199***	**.184*****
AUC-ROC	**.600**	**.625**	**.608**
Alpha	**.298**	**.331**	**.295**

[30] % based on 15 cases.
[31] % based on 2 cases.
[32] % based on 25 cases.
[33] % based on 12 cases.
[34] % based on 15 cases.
[35] % based on 27 cases.

Table 5 presents the zero-order correlations testing the relationship between each item and a new outcome[36], total number of times sent back to an institution (also shown is the relationship between the total scale and the new outcome). The results remain largely unchanged compared to what had been revealed previously. For boys six of the 14 items revealed no relationship with the outcome, and of the statistically significant items only two revealed a relationship of adequate strength. The overall correlation between the scale and the outcome (.159) was statistically significant and of adequate strength, however, in light of the number of items that contribute to the scale that have no relationship, the results can be classified as "Mixed results" bordering on "Poor" performance as before. Likewise the analyses for girls remain largely unchanged, however it is interesting to note that one of the two (out of 14 total) statistically significant items revealed a very strong relationship with the outcome (much stronger even than the scale as a whole). "Prior commitments" had a zero-order correlation of .286, while the total scale was at .182. As mentioned above it might very well be that the one item used alone could render more valid decisions than the scale as a whole. The performance of the scale for girls can be classified as "Poor."

Table 6 presents the same analyses as those in Table 5, except the default scoring categories were eliminated and the total score was recalculated using the revised items. For boys there was marginal improvement for some of the items, and likewise marginal improvement for the scale overall (.159 to .172). For girls, the total scale using the revised items loses statistical significance, while the analyses for the individual items remained unchanged.

Table 7 presents hand-calculated false positive ratios, disaggregated by race (white vs. non-white). There does not appear to be any difference in white and non-white boys regarding the number of false positive classifications. There was more difference between white and non-white girls, however as noted in the footnote on Table 7, very low case counts greatly influenced these particular analyses.

Finally, a series of logistic regression models (not shown) were calculated using sex, race, and RACF score (or RACF category) as predictors, and the binary outcome (sent back to institution

[36] Number of times sent back were truncated at 3 – meaning cases that had 4 or more incidents of being sent back to an institutions were recoded to a value of "3."

Table 5. zero-order correlations – items and scale x linear outcome; default scoring included[37]

Item	Boys	Girls	Total
Progress w/facility requirements	.096***	.074 n.s.	.096***
Problem solving skills	.038 n.s.	-.015 n.s.	.036 n.s.
Hostile response to frustration	.069*	.004 n.s.	.061*
Prior adjudications	.088**	.011 n.s.	.079**
Compliance with facility regulations	.169***	.027 n.s.	.151***
History of escapes	.103***	-.035 n.s.	.088**
Prior commitments	.086**	.286***	.108***
Sex offender	.096**	.042 n.s.	.093***
Mental health needs	.013 n.s.	.043 n.s.	.018 n.s.
History mental health placement at JRA	.003 n.s.	.161*	.053*
Manifest injustice	.021 n.s.	.041 n.s.	.021 n.s.
Age at admission	.028 n.s.	.047 n.s.	.032 n.s.
Scored as aggressive on SAVY	-.021 n.s.	-.007 n.s.	-.022 n.s.
Chemical/alcohol use	.068*	.095 n.s.	.072**
Total score	.159***	.182*	.167***

[37] n.s. = no relationship;
* = significant relationship, p < .05
** = significant relationship, p < .01
*** = significant relationship, p < .001

Table 6. zero-order correlations – items and scale x linear outcome; no default scoring[38],[39]

Item	Boys	Girls	Total
Progress w/facility requirements	.096***	.074 n.s.	.096***
Problem solving skills	**.047 n.s.**	**-.003 n.s.**	**.051 n.s.**
Hostile response to frustration	**.085****	**.038 n.s.**	**.082****
Prior adjudications	.088**	.011 n.s.	.079**
Compliance with facility regulations	**.169*****	**.023 n.s.**	**.151*****
History of escapes	**.105*****	**-.031 n.s.**	**.090*****
Prior commitments	.086**	.286***	.108***
Sex offender	.096**	.042 n.s.	.093***
Mental health needs	.013 n.s.	.043 n.s.	.018 n.s.
History mental health placement at JRA	.003 n.s.	.161*	.053*
Manifest injustice	.021 n.s.	.041 n.s.	.021 n.s.
Age at admission	.028 n.s.	.047 n.s.	.032 n.s.
Scored as aggressive on SAVY	**-.021 n.s.**	**-.009 n.s.**	**-.023 n.s.**
Chemical/alcohol use	**.070***	**.098 n.s.**	**.074****
Total score	**.172*****	**.159 n.s.**	**.188*****

[38] n.s. = no relationship;
* = significant relationship, p < .05
** = significant relationship, p < .01
*** = significant relationship, p < .001
[39] Bold-faced items had default scores that were eliminated for these analyses. Light-faced items are based on the same variables that appeared in Table 5. Total score correlations were re-calculated as the RACF total score was re-calculated using the revised variables that eliminated default scoring.

Table 7. False positive ratios[40]

RACF classification[41]	Return to institution
White male	17%
Non-white male	17%
White female[42]	81%
Non-white female	90%

[40] False positive ratios were calculated by first identifying for each group the number of true negatives and false positives. True negatives were the number of cases that were classified as being in the lowest category of risk, who did not recidivate. False positives were the number of cases that were classified as being in the highest two categories of risk, who did not recidivate. The number of false positives were divided by the number of true negatives plus false positives in order to establish the ratio of false positives to N.

[41] Due to low case numbers, the categories of "high" and "very high" were combined.

[42] Exceedingly low case counts make these estimates very unstable.

yes/no) as the dependent variable. In all modeling, sex was not significant (meaning neither boys nor girls were significantly more likely than the other to be sent back to an institution). Race however was statistically significant in all the models, indicating that non-white youth were statistically more likely to be sent to an institution from a community facility compared to white youth[43]. Finally, RACF score (as well as RACF category) was statistically significant when predicting the outcome, while controlling for sex and race.

Conclusions and next steps

If the RACF is kept in place, there is a great deal of room for improvement. Overall several of the items do not reveal a statistical relationship with the outcome. Of those that do reveal a significant relationship, differentiation in the rates of outcome tends to be inadequate in several instances. The effectiveness of the scale when assessing female youth is poor, with low sample size likely affecting the results. The scale could be re-tooled in order to utilize only the items that reveal a statistically valid relationship with the outcome, though in the case of female youth only two items would be left. As noted in "Testing the validity of the RAR and the RAI" given the amount of effort that goes into the ITA, it could be worthwhile to determine whether an instrument that adequately and reliability predicts community facility misconduct could be built.

[43] For the total sample, 32% of white youth were sent to an institution from a community facility, compared to 39% for non-white youth.