



EVALUATION OF AGGRESSION REPLACEMENT TRAINING (ART)

In Washington State Juvenile Rehabilitation





Washington State Department of
CHILDREN, YOUTH & FAMILIES

CONTENTS

Summary 1

Aggression Replacement Training (ART)..... 1

Current Study 2

Methods..... 2

 Propensity Score Matching 3

Implementation Findings 4

 ART Eligibility 4

 Treatment Quality 5

 JR WSART Starts and Completions 6

 Impact of ART 7

 Consistency of Impact 8

 Facilitator Adherence 9

Conclusion and Discussion 12

 Recommendation 1: Update Eligibility Criteria..... 12

 Recommendation 2: Explore Dosage Variations..... 13

 Recommendation 3: Ensure Data Consistency 13

 Recommendation 4: Track Alternative Outcome Measures..... 13

 Recommendation 5: Establish a Pattern of Program Evaluation..... 14

References 15

Appendix A..... 17

Appendix B 20

Appendix C 23

Andrew M. Fox, PhD | Sarah Veele, PhD | Alexandra Goins, MA

Suggested Citation:

Fox, A. M., Veele, S., and Goins, A. (2021). Evaluation of Aggression Replacement Training (ART) in Washington State’s Juvenile Rehabilitation. Olympia, WA: Department of Children, Youth, and Families, Office of Innovation, Alignment, and Accountability.

Summary

Juvenile Rehabilitation (JR) has been implementing Washington State’s version of Aggression Replacement Training (WSART) in secure institutions since 2008. From 2008 to 2018, 2,631 youth started WSART. This is the first evaluation of the program in JR. Using propensity score matching, this evaluation finds that WSART did not have a significant impact on 18-month conviction recidivism rates overall; however, the program was related to a significant reduction in misdemeanor recidivism. Subgroup analysis further indicates that there are some risk and protective domains that can be used as eligibility criteria for WSART in JR, potentially increasing the effectiveness of the program by better matching the program with those youth who may benefit most from the intervention. Recommendations include scaling back implementation with some updated eligibility criteria, ensuring facilitator adherence data be routinely collected, and exploring the use of alternative, short-term outcomes for use in future evaluations.

Aggression Replacement Training

Youth violence significantly impacts communities across the United States, and state and local jurisdictions are seeking to identify treatment programs that can reduce it. Aggression Replacement Training (ART), one such program, is a 10-week psychoeducational intervention originally designed for incarcerated youth identified as hostile and assaultive. The program teaches a series of prosocial behaviors through a structured learning environment to small groups of six to eight youth. The training is comprised of three components: social skills, anger control, and moral reasoning. Glick and Goldstein (1987) first demonstrated that ART reduced the number and intensity of behavioral incidents of youth while in an institution and showed that youth were able to apply and transfer these skills into the community after being released. This program has shown promise in the decades since first implemented, and ART has been used in school and family-based settings and with youth on the autism spectrum (Calame & Parker, 2003; Moynahan, 2003; Roth & Striepling-Goldstein, 2003).

Brannstrom et al. (2016) conducted a systematic review of the literature in which they identified 16 studies that examined the impact of ART on recidivism. Their review concluded that “the primary studies of ART do not provide a sufficient base for substantiating the claim that the program is effective for reducing antisocial behavior in adolescents and adults” (p. 40). A major concern was that almost half the studies were completed by researchers who had a vested interest in the program. Additionally, the methodological rigor of the current body of research is limited. The National Institute of Justice rates ART as an effective program¹ based on two studies (Barnoski, 2004; Gundersen & Svartdal, 2006), but more recent research is not conclusive about the effectiveness of the program.

Washington State has made a significant investment in ART. First implemented in 1999 for justice-involved youth in a community setting, the state JR agency expanded the program and implemented within state juvenile residential facilities in 2008. Several studies have been conducted on Washington State ART (WSART)² in community settings, with the most recent studies causing concern about the effectiveness of the intervention. In 2004, Barnoski evaluated WSART implemented in 26 courts across 28 counties by comparing 704 youth who started the program in the year 2000 to 525 eligible youth who were wait-listed due to a dearth of resources. The study found that ART resulted in a non-significant reduction in 18-month recidivism;

¹ National Institute of Justice. (2012, June 14). *Program profile: Aggression Replacement Training (ART)*.

<https://www.crimesolutions.gov/ProgramDetails.aspx?ID=254>

² WSART is the Washington State adaptation of ART. The research on WSART is exclusively from implementing it through the Washington State Juvenile Courts in which participants live in the community while in the program. WSART has also been implemented in Juvenile Rehabilitation.

however, for those courts that administered WSART with high fidelity, there was a significant reduction in felony recidivism. This finding highlighted the importance of implementation fidelity.

In 2017, using propensity score matching, Peterson compared 951 youth in a community probation setting from 23 juvenile courts who started WSART to 951 youth who did not receive an evidence-based program (ART or otherwise). While the author noted some missing data and data quality issues, Peterson found that youth who started WSART had higher felony recidivism than a matched control group. More recently, Knoth et al. (2019), using a sample from 2006 to 2016, found that WSART participants were significantly more likely (23% compared to 19%) to recidivate than matched youth who did not participate in the intervention. Additionally, the authors compared WSART (N=6,453) to a treatment as usual group (N=6,453) using propensity score matching, and found that the trainer competence score did not have an effect on the results.

All outcomes on WSART come from implementation through the courts, where the youth were not living in a state facility. Although the training, quality assurance protocols, and assessments in the Juvenile Rehabilitation (JR) institutions are identical to those that have been used in the courts with youth who are in the community, little is known about the effectiveness of the program for juveniles in residential programs. This is important because the implementation of WSART in JR is different from WSART in the community setting because JR youth are learning these skills while incarcerated with other youth with whom they live and from staff who also provide full time care for them. Additionally, JR is usually a last resort in the juvenile justice system and the youth in JR, on average, are at a lot higher risk than those in court programs. Therefore, it is important to identify whether the effectiveness of WSART varies by implementation context.

Current Study

WSART was first implemented in Washington State JR residential facilities in 2008. WSART in JR has never been evaluated. The current study has sought to replicate the recent study conducted by the Washington State Institute for Public Policy (Knoth et al., 2019), which examined WSART for youth in the community. This study looked at the level of WSART implementation in JR residential facilities in Washington State. We examined implementation qualitatively, assessed the rate of program completion, and reviewed whether WSART significantly reduces recidivism compared to a matched control group. Additionally, the study examined the consistency of WSART across risk assessment domain scores and how facilitator competency relates to participant outcomes.

Methods

In order to understand the implementation of WSART in JR, we met with the current treatment administrator who oversees the program and master trainers from three JR institutions. Staff members also provided insight into current program implementation in JR institutions specifically and the WSART model more generally. Additionally, we reviewed all manuals, quality assurance plans, training materials, and client satisfaction surveys. The findings from the implementation assessment are provided in this study.

The authors compiled the data for the outcome analysis in this study from a number of sources. First, program information on who started ART, when they started and finished the program, whether they completed it, and the names of facilitators are all kept in a master Excel spreadsheet. The previous treatment administrators have maintained this spreadsheet, and it includes all WSART program data since the start of the program in JR. This Excel spreadsheet was cleaned, and these data were used to identify who started and completed the program while in JR. Using an intent-to-treat design, the main independent variable in this study was any JR

WSART start. These data were matched to the JR records management system, called the Automated Client Tracking (ACT) system.

A number of control variables (also used for matching) were applied to this study, including age at admission, gender, time served, race or ethnicity, most serious offense, offense class, whether they served parole, sentence type (regular or adult), release year, whether they spent time in a community facility, and all the historical and current risk and protective domain scores on the Integrated Treatment Assessment (ITA). All WSART participation in the community, as part of the juvenile justice system, was also included as a control variable.³ The data for WSART participation in the community was obtained from the Washington State Center for Court Research. A full list of the variables used to create the propensity scores, as well as those controlled for in the outcome models, can be found in Appendix A.

Four outcome variables were used to measure recidivism. We identified all youth who were released from JR residential facilities in calendar years 2010 through June 2017 in administrative records. We then obtained data on all convictions from the Washington State Center for Court Research. Recidivism was defined as an offense that occurred within 18 months of release from a residential facility and that resulted in a conviction or adjudication. The at-risk date was the day youth ended their residential sentence, whether from one of the three state institutions or eight state community facilities. An additional 12 months were allowed for the offense to advance through the courts to determine if a conviction would occur. Recidivism types included any recidivism, misdemeanor, felony, and violent felony. All clients with any type of recidivism were included in any recidivism in this report. We identified the most serious offenses that occurred in the first recidivism event to determine the type of recidivism event. All youth with a violent felony conviction were also included in the felony recidivism count. In this report, clients with only a misdemeanor offense were included in the misdemeanor recidivism and were not counted in the felony and violent felony recidivism. Youth with a felony could also have a misdemeanor but were only included in the felony recidivism category.

Propensity Score Matching

One of the biggest challenges in this type of evaluation is to ensure that the differences in outcome can be attributed to the intervention and not a demographic, risk level, or other difference between the study groups. In research, this is called selection bias. As an example, if we only examined differences in recidivism between those who started WSART and those who did not without considering whether older youth were enrolled in WSART and younger youth were not, we might see that youth who started WSART had lower recidivism rates. In reality, the lower recidivism rates could be the result of those who were older when they started WSART and the comparison group being younger. In an ideal research world, we would do a randomized control study so that the only differences between the treatment and control groups would be random chance. However, in practice, treatment randomization is not feasible, so we need to use other techniques to ensure we can isolate the effect of WSART (the treatment) on recidivism (the outcome).

³ In this study, 26% of youth in the comparison group did have WSART in the community and 27% of youth in the treatment group received WSART in the community before being sent to JR. Based on the WSART eligibility criteria, which were designed for a population with more variation in risk, almost all JR youth are eligible for WSART. We do not exclude those who had WSART in the community; instead, this variable was used in the creation of the propensity score and in the full regression models predicting recidivism.

To limit selection bias in our analysis, we use propensity score matching (PSM). The first step in PSM is to predict the likelihood that a youth would start ART, and the prediction of whether a youth started WSART is their propensity. Youth who did not start WSART often had the same characteristics of youth who did, but they did not start the program for one reason or another. For example, youth who were eligible might not start WSART because they were going to be transferred to another facility and would not have enough time to complete the program. These propensities were then matched to ensure that the treatment and comparison groups were as similar as possible, without using randomization.

For the current study, we use all the variables listed in Appendix A as the matching criteria. Appendix A also shows the balance before and after matching occurred.⁴ It is clear that for many of the variables, there is imbalance (meaning the treatment and comparison groups are significantly different from each other) before matching, but the two groups are more similar after the matching process. Matching in this way reduces the effects of selection bias. The following analysis uses a 3:1 matching process with replacement and caliper.⁵ This means that to find a match (i.e., a youth who was similar in characteristics but did not start ART), three matching youth were identified, some even selected multiple times (replacement), with propensities within an acceptable range (caliper). To ensure that our matching method was robust, we also tested other matching strategies. The other matching techniques and associated outcomes can be found in Appendix C. The results across the matching techniques were similar. We selected the 3:1 matching process with replacement and caliper because the propensity distribution and balance are the best (see Figure A1 in Appendix A). Using the matched sample, we then used logistic regression to predict the recidivism outcomes using WSART starts as the main independent variables, while controlling for other relevant factors. These models were weighted using sampling weights since we allowed replacement in the matching process.

After identifying the main effect of WSART, we conducted subgroup analysis within demographic groups and both high and low risk and protection for all current and historic domains of the ITA. Accordingly, we dichotomized each domain score with the top third of high risk (or low protection) being a one and all other values being zero.⁶ We rematched youth who started WSART to those who did not within each subgroup analysis. After matching, we tested the effect of WSART on recidivism within the matched group. Rematching appears to be the appropriate method for conducting moderation analysis (Green & Stuart, 2014). The results indicated that the effect of WSART might vary by group or be consistent across groups. Before we present the quantitative findings on the effectiveness of WSART, we will review the findings related to the implementation of the program.

Implementation Findings

ART Eligibility

The ITA is used in JR to determine eligibility for ART. The ITA is a risk and needs assessment administered within 14 days of admission into a JR residential facility. The assessment is similar to the Residential Positive

⁴ Table A1 in Appendix A reports the percent bias, the percent difference of the sample means in the treated and untreated samples for both the full and matched samples. The percentages are standardized using the equation from Rosenbaum & Rubin (1985). A value of 25 indicates severe imbalance and a value greater than 10 indicates a moderate imbalance. There are no values greater than 10 in the matched sample.

⁵ The caliper width was calculated as 20% of the standard deviation from the logit of the standard deviation, as recommended by Austin (2011).

⁶ The current cut points for the ITA domains have not been extensively tested. We decided to use the top third of the domain variation for consistency to ensure there was an adequate number of youth in the top of the distribution to test.

Achievement Change Tool (R-PACT), which has been studied extensively (Hay et al., 2019). However, JR has yet to validate the weighting of the ITA for the JR population, which means that some of the weighting might not be optimally related to risk. Nevertheless, WSART in JR uses the eligibility criteria identified by WSART for youth living in the community (which is implemented in the courts). Specifically, to be eligible for WSART, a youth must have a high score in the following domains: 1 (record of referrals), 10B (current attitudes and beliefs), 11 (aggression), and 12 (skills).

It is encouraging that JR is using the ITA to determine eligibility for WSART because this follows the risk-need-responsivity model (RNR). The RNR model is comprised of three principles. First, the risk principle suggests that those with the highest risk for reoffending should be prioritized for treatments and other interventions. Second, the need principle recommends that the individual needs of each youth be determined, specifically those needs that are most likely to be associated with criminal behavior. Third, the responsivity principle requires that the correct type of programming based on an individual's risk and need profile be offered (Brogan, Haney-Caron, NeMoyer, & DeMatteo, 2015; Crites & Taxman, 2013). In JR ART, the use of the ITA indicates that there is some attention paid to the need and responsivity principles of the RNR model.

None of the research reviewed indicates that the current eligibility criteria are predictive of future violent behavior, which would point to WSART as an appropriate treatment. The current program is using an assessment (ITA) for eligibility, and the assessment is being used appropriately; however, JR has not determined that the selected eligibility criteria are appropriate. The research has not determined who would benefit most from participation in the program. JR needs to identify predictors that identify which youth will benefit from WSART and use this information to inform staff of when WSART is the best response, given a profile of risks and needs of a youth.

Treatment Quality

JR appears to be implementing the treatment according to the design. There are strong training and quality assurance protocols in place. In terms of dosage, under the current design all those who start the program receive the same dosage (if they complete it), which is three sessions per week for 10 weeks. It is not clear whether youth are receiving the right dosage. It is likely that some youth require more treatment and some less, but the current design of WSART does not allow for this type of dosage variation. Some reporting indicates that WSART is only allowing 45 minutes for sessions in some places due to school schedules. This would result in a lower dosage than intended. There is variation by location in terms of when WSART is administered, but the standards and quality assurance for the program are consistent across location.

The current implementation design does not include much integration of the assessment (ITA) or with other treatments or reentry planning. Assessments are used to determine eligibility, but they are not used during WSART to determine dosage or focus since it is a group-based program. WSART has a unique set of skills that are different from Dialectical Behavior Therapy (DBT), and there is little to no training for staff on how the two sets of skills overlap or are complementary. Generally, living unit staff members have not been trained in the WSART skills and are not able to reinforce youth who use the WSART skills.

WSART in JR has a well-established quality assurance plan. Facilitators submit a recording of a session annually, and this session is evaluated by a master trainer. The master trainer then determines whether the facilitator is adherent to the model or not. There is a youth survey in place for performance improvement, but this is not being administered on a routine basis. The program also has an established curriculum for training

new trainers, monthly consultation conference calls, site visits and feedback from the WSART administrator, and semiannual quality assurance committee meetings within JR.

JR WSART Starts and Completions

Table 1 shows the number of youth who started and completed WSART at a JR institution. The data were provided for calendar years 2008 through 2018. Each start was counted, so youth who started more than once during this time period were counted each time they started. A total of 2,631 youth started WSART in JR on one or more occasions with a completion rate of 78.4%. Figure 1 shows the starter and completer trends over time, displaying the information presented in Table 1.

Table 1: JR WSART Starts and Completions from 2008 to 2018

Calendar Year	Started	Completed	
	n	n	%
2008	38	28	73.7%
2009	181	151	83.4%
2010	318	248	78.0%
2011	212	166	78.3%
2012	316	259	82.0%
2013	331	284	85.8%
2014	292	237	80.9%
2015	252	199	79.0%
2016	277	200	72.2%
2017	272	196	72.1%
2018	142	96	67.6%
Total	2,631	2,064	78.4%

Figure 1: JR ART Starters and Completers by Year



Table 2 shows the number of WSART starts from calendar year 2008 through the end of 2018 by location where the program was held. Of the 2,631 WSART starts, 26% occurred at Echo Glen Children’s Center, 34% were at Green Hill School, and about 30% were at Naselle Youth Camp. Close to 10% were at Maple Lane School, which was closed as a JR facility in June 2011. Less than 1% of youth who received WSART started the program while at a community facility.

Table 2: JR WSART Starts by Location From 2008 to 2018

	n	%
Echo Glen Children’s Center	685	26.0%
Green Hill School	894	34.0%
Naselle Youth Camp	780	29.6%
Maple Lane School	256	9.7%
Other	16	0.6%
Total	2,631	100.0%

Note: WSART was offered at the Maple Lane School from 2008 through June 2011 when it was closed as a JR facility.

Impact of ART

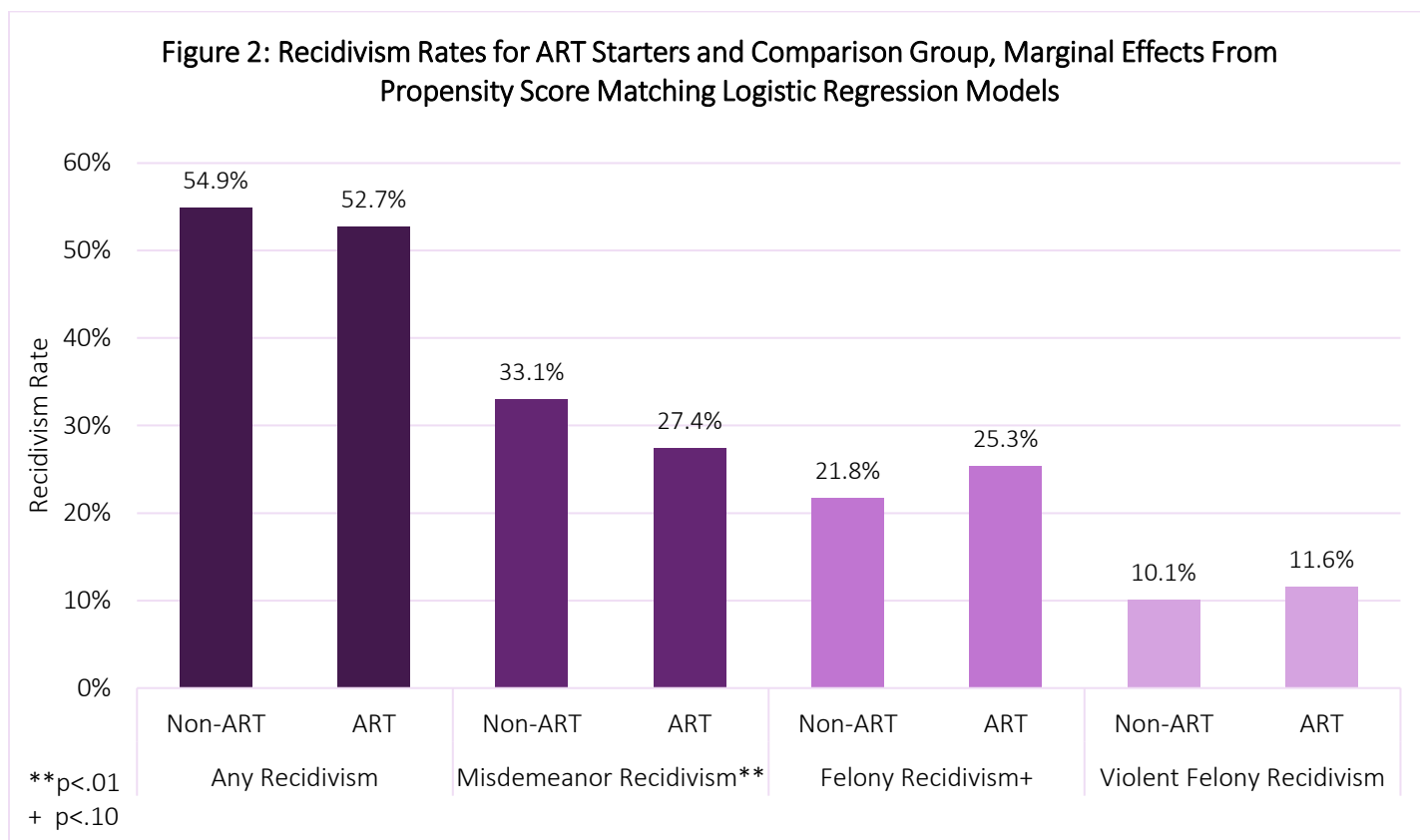
To determine the impact of WSART on recidivism outcomes, we use the propensity score matching outlined in the methods section of this report. We present the models from the 3:1 matching that allowed for replacement, so a youth in the control group could be included more than once, and a caliper, which ensured matched propensity scores were within a reasonable distance. Using this procedure helps to limit the effects of any selection bias that might exist in the data.

Starting with 2,631 WSART starts, 286 were removed because a youth started WSART more than once during a single residential sentence. A youth was counted only once if they started WSART multiple times during a residential sentence. They were counted as a completer if they finished any WSART starts during the same residential sentence. To allow for enough follow-up time for the study’s outcome variable, only youth released from JR before July 2017 were included, resulting in 414 cases being deleted from the analysis. An additional 689 youth were excluded because they did not have a completed ITA; this was particularly true for youth released in 2010, 2011, and 2012 since the ITA was not fully implemented when they were admitted.⁷ Another 115 were excluded since youth with multiple residential obligations during the study time period took WSART during different residential sentences. We randomly selected which residential sentence to include in the study. Finally, two were excluded because their propensity scores were off support, meaning they were outside the range scores from the sample of comparison youth. For the outcome analysis, a total of 1,125 youth, 43% of all WSART starts, were included in the analysis and a sample of 1,752 youth were identified as eligible for the comparison group. The final models included 1,125 in the treatment group and 1,307 in the comparison group, taken from the 1,752 eligible to be included in the comparison group.

Figure 2 shows the marginal effects for each of the four recidivism outcomes for both those who received WSART and those who did not. These estimates of recidivism rates are calculated while keeping all other

⁷ A comparison of demographics between those who had a completed ITA and those who did not indicates that the two groups were identical in age, however, males and Hispanic youth were less likely to have a completed ITA.

variables included in the model constant. Overall, WSART seemed to result in significant reductions in misdemeanor recidivism (33.1% vs. 27.4%), with non-significant but potentially concerning increases in felony recidivism ($p = 0.061$; 21.8% vs. 25.3%). In terms of any recidivism, those who did not have WSART recidivated at a 54.9% rate, compared to 52.7% for those who received ART. The full regression models that produced these estimates can be found in Appendix B.



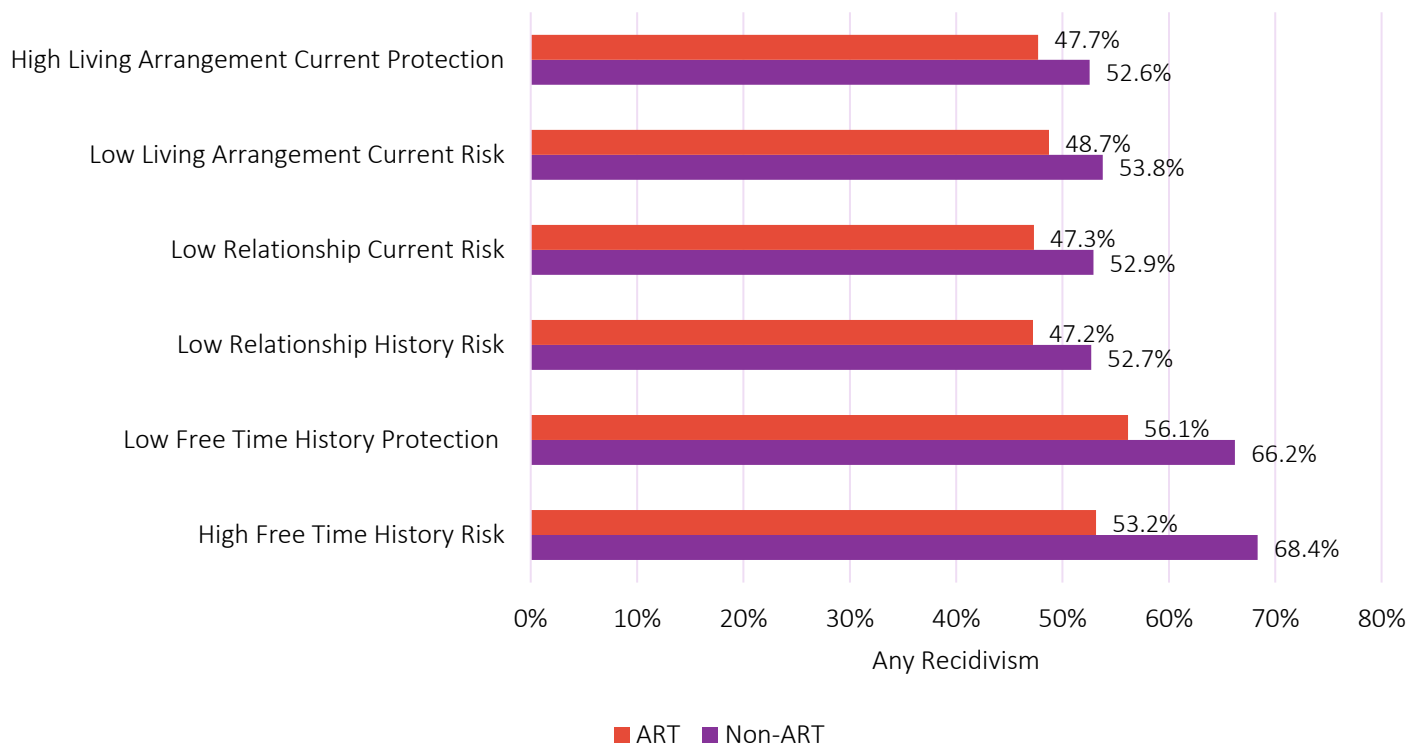
Consistency of Impact

In order to assess the consistency of impact, we conducted subgroup analysis based on age, gender, release year, and level of each risk and protective domain. For each subgroup analysis, we rematched the sample for analysis within each subgroup, before re-estimating the effect of WSART on recidivism. For each risk and protective domain, high risk and low protection was identified based on the top third who were either high in risk or low in protection for each domain score. A total of 84 different subgroup analyses by demographic, risk, and protection level were conducted. For each model, the variable of interest (i.e., the subgrouping variable) was excluded from the controls, but all other variables remained. Based on these analyses, a total of six effects showed that WSART had a significant effect on recidivism within a specific group. In terms of demographics and release year, WSART did not have a significant impact on recidivism within any of those subgroups. Next, we rematched youth who started WSART to those who did not within both low and high groups of each risk and protective domain score (a total of 72 subgroup analyses).

Figure 3 shows the six significant effects and their subgroup that effect was within. In terms of high risk and low protection, there was one domain that showed a significant WSART effect. All youth who had high protection within the current living arrangement domain were identified, and youth who started WSART were matched to those who did not, based on propensity scores. Then, the impact of WSART was tested within that

subgroup of youth. The analysis determined that WSART was related to significant reductions in recidivism for youth who had high protection in the current living arrangement domain, but WSART was not effective for youth with low protection in this domain. Next, three risk domains revealed a significant WSART effect when the risk was low. WSART significantly reduced recidivism for youth with low risk in the current living arrangement, current relationship, and historical relationship domains. WSART was not effective when youth were high risk in these domains. The only domain where those with low protection or high risk experienced a benefit was free time domain. WSART was related to a significant reduction in recidivism for youth with low historical free time protection and those with high risk in historical free time.

Figure 3: Significant Effects of ART on Recidivism Within Subgroups



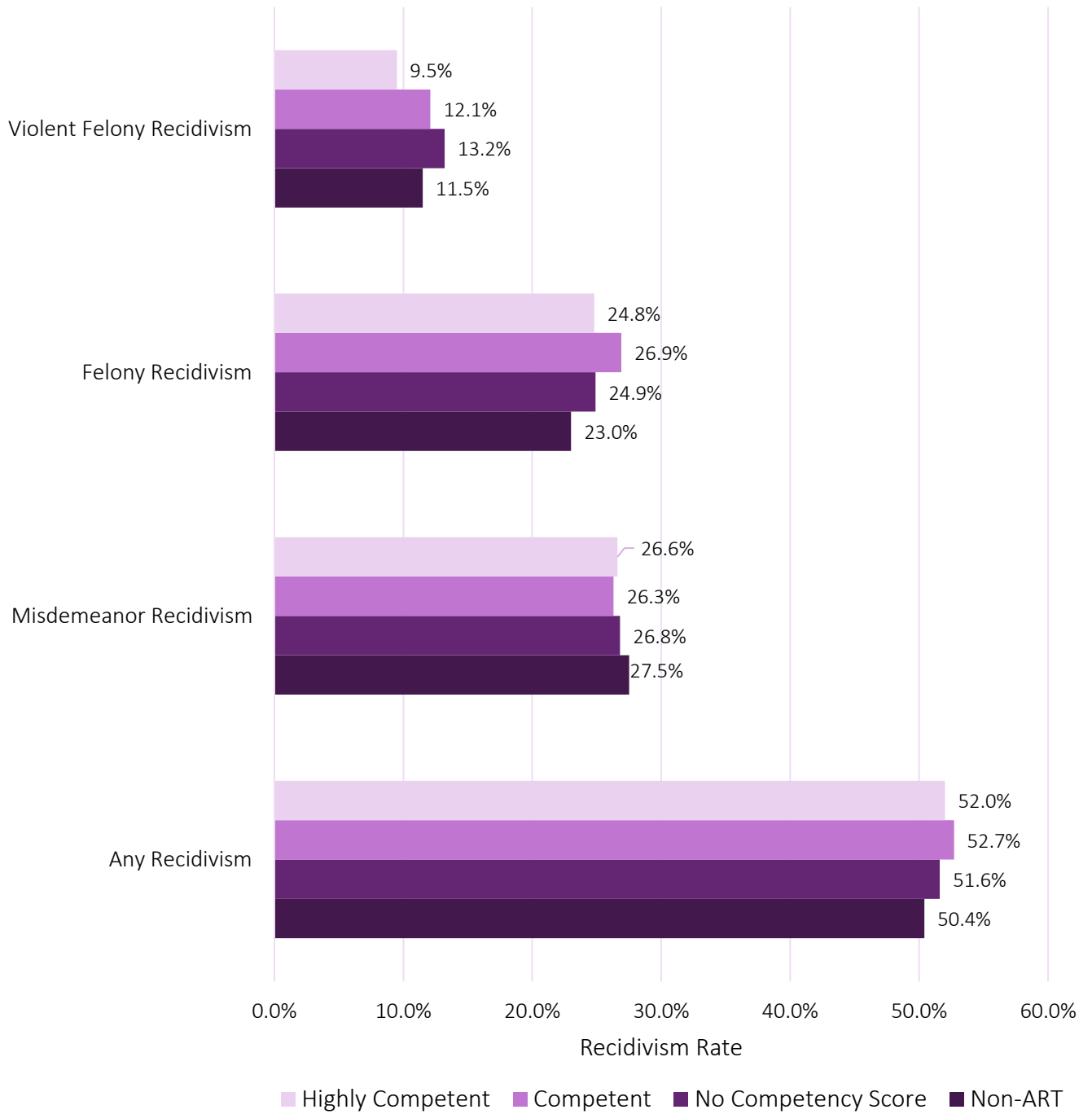
Facilitator Adherence

ART facilitators are supposed to be evaluated on a yearly basis. Additional details about the fidelity rating process for WSART can be found in prior research (Knoth et al., 2019). For the current study, we were able to match the facilitators to the specific courses youth started. Unfortunately, much of the adherence data for JR was kept as a paper file and then discarded when a treatment administrator left the agency. For that reason, many facilitator scores were not available, specifically for those who provided the intervention to youth who were released from JR between 2010 and 2014. Because facilitators can opt to take a refresher training course instead of going through the assessment process, and many facilitators chose this route, adherence data were not available for all facilitators each year. If facilitators did not have a competence score in the year prior to the class, the case was not included in this analysis. In total, we were only able to obtain facilitator competence scores for 465 of the 1,127 (41.3%) WSART starts included in the study. When there was more

than one facilitator listed, we averaged their most recent competence scores (had to have been in the year prior to the start of the course). Scores can potentially range from 0 to 54; in practice, however, scores ranged from 34 to 54. Scores above 42 were counted as highly competent and scores from 34 to 42 were considered competent.

Full logistic regression models are estimated using the same propensity score matching as presented in Figure 1. The marginal effects are presented in Figure 4. Those who started WSART were placed in one of three categories: competence of the facilitator was unknown, competence level was competent, and facilitator was rated as highly competent. Significance testing indicated that when compared to those who received highly competent ART, there were no significant differences in recidivism rates for those who received WSART from facilitators whose competence levels were rated as competent or unknown. From this analysis, we cannot say that facilitator competence level is related to the outcomes of the youth. The data did indicate that the variation at the upper end of the competency range did not seem to differentiate the effect of the program.

Figure 4: Recidivism Rates by Facilitator Competency Scores



Conclusion and Discussion

Washington State has implemented WSART as part of its intervention programming in JR for approximately 12 years. In that time, we had expected to see an overall effect of WSART on youth recidivism rates. The current study did not find this overall effect. WSART in residential facilities does appear to be associated with significant reductions in misdemeanor recidivism and a non-significant increase in felony recidivism. Subgroup analysis indicates that WSART was effective for youth when they were low risk in some domains and high risk in only one domain.

Some notable similarities and differences are identified when we compare the findings of this study to the recent evaluation of WSART in the courts (Knoth et al., 2019). First, we do not find consistent increases in recidivism for youth who participated in WSART whereas WSART was associated with an increase in felony recidivism. The study of WSART in the courts found an increase in misdemeanor recidivism while the current study found a significant decrease in that outcome. Nevertheless, the findings are consistent between the two studies related to the most serious recidivism outcomes.

This finding does not indicate that JR should cease to offer WSART to youth in JR. Rather, this study suggests that implementation should be scaled back and monitored closely moving forward. Our recommendations, supported by the current findings, are outlined below, and some of the recommendations concerning implementation are taken from the recently published report on JR's Integrated Treatment Model (Fox & Veele, 2020).

Recommendation 1: Update Eligibility Criteria

The eligibility criteria for JR WSART should be reconsidered. Initially, ART's criteria were designed for use in the community by juvenile courts, but JR youth are at a much higher risk for recidivism. As a result, most youth in JR are eligible for WSART based on the current eligibility criteria. If the program has the wrong eligibility criteria, it may be determined ineffective by an outcome evaluation when, in fact, the program was incorrectly assigned to a participant. The current analysis found six risk and protective domains that should be targeted; based on the analysis, four of the six would target youth with low risk or high protection in specific domains since these are the youth that are likely to benefit from WSART. The data does not suggest that WSART is harmful for high-risk youth in these domains, only that it was ineffective for them.

The number of youth assigned to WSART should be based on need and not predetermined quotas. This could be a challenge when trying to plan for treatment capacity, but ultimately the need should drive capacity planning. More treatment is not always better; the key is being responsive to need. JR should refine the eligibility criteria so that it is based on evidence accumulated over the past 10 years of implementation, and then focus WSART on those for whom the program can most positively impact. For example, if JR changed the eligibility criteria to the six domains and levels outlined in this report and only provided WSART to youth who met five or six of the criteria, then JR would only need to provide WSART to 30 to 50 youth per year. It is somewhat counterintuitive to the risk-need-responsivity model to target treatment toward youth who are low risk in certain domains. In terms of resources, JR youth are likely better served if treatments that target areas of high need are emphasized (for example, substance use treatment).

Similarly, it is important to highlight the fact that JR uses the ITA as its risk and needs assessment. This tool was developed for juveniles, and validation work has not been done for an adult population. Given the

expanding age of clients served, JR needs to determine whether the risk and need profile of a juvenile is the same as the profile for an adult, especially in terms of whether WSART is the appropriate program.

Recommendation 2: Explore Dosage Variations

JR should consider and test WSART programming dosage variations. Research on crime prevention programming has shown that some interventions, at lower dosages, can fail or even do harm (Linning & Eck, 2018). If we want to see positive results, we must address the difficult but necessary process of identifying and matching the proper dosage to the right youth. JR needs to implement a process that will allow for closely monitored variations of WSART programming. For example, JR could test a shortened version of WSART for youth who have the need and are high risk but do not have a notably long sentence. JR could also start to explore options for a graduated course for youth with longer sentences who need additional reinforcement of the WSART skills. This could include an alumni group that continues to meet so that the skills can be discussed and reinforced. There is also some discussion among staff about whether an eight-week course could be just as effective for youth as the current 10-week structure. These are all important considerations, and JR needs to have a process for continuing to learn about what is most effective for high-risk youth and those who need to reduce their aggression. These types of variations would need robust evaluation protocols to help determine both the strengths and challenges of each. These dosage variations could be monitored and evaluated using short-term outcome metrics so that analysis and delivery modification could happen more quickly.

Recommendation 3: Ensure Data Consistency

JR needs to ensure the consistent collection of facilitator adherence data and program participation data in the agency's ACT records management system. There is currently a quality assurance process in place that requires recording and scoring one session per year for WSART facilitators. The basic structure of this quality assurance process is good; however, we offer a number of recommendations to improve the process. First, some facilitators can participate in an updated WSART training course as an alternative to the facilitator adherence scoring. This should not be permitted. Taking an additional training course does not ensure a facilitator is providing adherent sessions, and the missing competency data prevent us from testing these assumptions. A master trainer should review each facilitator every year, without exception. Second, there is little variation in facilitator adherence scores. More research is needed to determine which aspects of facilitator adherence are most associated with youth success. Understanding how program quality is related to youth outcomes will help continue to improve WSART implementation. Third, all program start and end dates, current and historical, should be verified in the agency's ACT system. Many programs, including ART, maintain a spreadsheet with program data. This practice is fine if ACT is the recognized authority for program and treatment data. Finally, all facilitator adherence data should be maintained in the ACT system. If there are any facilitator adherence paper files remaining elsewhere in the agency, they should be inputted into the ACT system. Many paper files from early in the program's implementation were discarded by an outgoing treatment administrator, and the lack of data from this time severely limited our ability to study how adherence is related to youth outcomes, and we wish to avoid this from happening.

Recommendation 4: Track Alternative Outcome Measures

JR should explore alternative short- and long-term outcome measures. Recidivism is the common outcome measure for testing the effectiveness of programming in juvenile corrections. Recidivism is also problematic

for a number of reasons. First, it takes a long time to obtain recidivism data. In the current study, we had to wait a minimum of 24 months after youth were released from JR. Second, recidivism data rely on the detection of behavior by the criminal justice system. Not all youth who commit a crime are caught and convicted, and the likelihood that youth are arrested and convicted is not always equal across groups. Any bias in the arrest and conviction process will also be present in the recidivism data used for evaluation. For this reason, it is important to establish alternative outcome measures. A short survey could be developed, borrowing from existing surveys, that would measure criminal thinking and coping mechanisms. A pre-post evaluation design could be established to identify how dosage variation affects those who take ART. Another option would be to engage in more regular updates to the ITA. This would provide researchers the opportunity to monitor change in ITA domains as youth move through the different treatment areas within JR.

Recommendation 5: Establish a Pattern of Program Evaluation

JR should establish a pattern of evaluating programs, both their processes and outcomes, every few years. Ideally, corrections or modifications to WSART would have been made every few years. It is unreasonable to wait 12 years for an evaluation and expect the program to have been consistently implemented throughout this time or to have been producing the best outcomes possible. Additionally, we should not discard this, or any other, program based on one study. JR has many well-trained WSART facilitators and well-established quality assurance protocols. The correct approach is to quickly move to make the changes suggested here, and then to reevaluate. Short-term outcomes could be used to understand the effect that these program changes are having. Part of implementing evidence-based practices includes monitoring the impact these programs have on the local population and adjusting appropriately. We take programs that are considered best practices nationally and test to see if, how much, and for whom the programs work for the youth in our care.

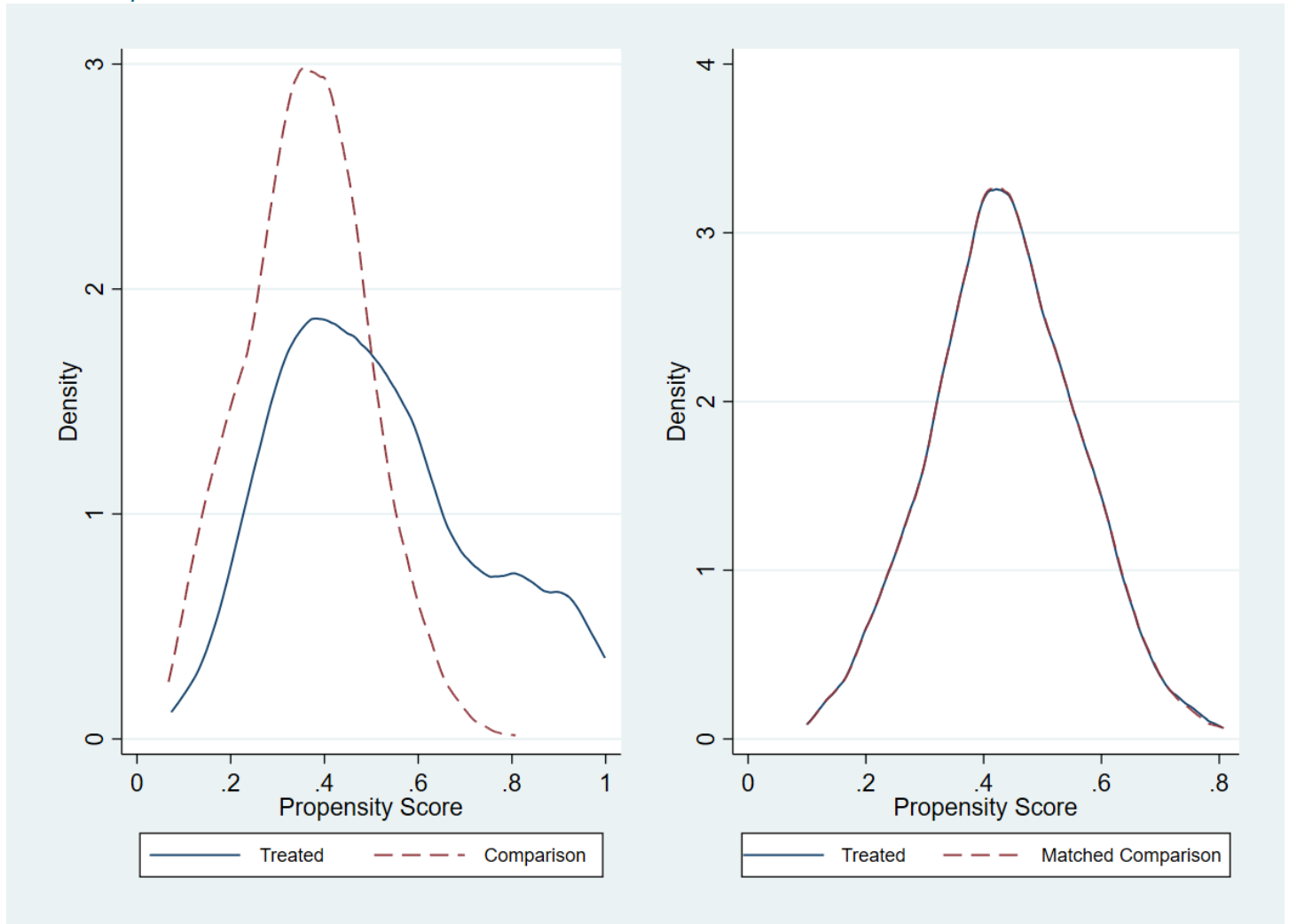
References

- Austin, P. (2011). Optimal caliper widths for propensity-score matching when estimating differences in means and differences in proportions in observational studies. *Pharmaceutical Statistics, 10*, 150–161.
- Barnoski, R. (2004). *Outcome evaluation of Washington State's research-based programs for juvenile offenders*. (Document No. 04-01-1201) Washington State Institute for Public Policy. http://www.wsipp.wa.gov/ReportFile/852/Wsipp_Outcome-Evaluation-of-Washington-States-Research-Based-Programs-for-Juvenile-Offenders_Full-Report.pdf
- Brannstrom, L., Kaunitz, C., Andershed, A., South, S., & Smedslund, G. (2016). Aggression Replacement Training (ART) for reducing antisocial behavior in adolescents and adults: A systematic review. *Aggression and Violent Behavior, 27*, 30–41.
- Brogan, L., Haney-Caron, E., NeMoyer, A., & DeMatteo, D. (2015). Applying the risk-needs-responsivity (RNR) model to juvenile justice. *Criminal Justice Review, 40*(3), 277–302.
- Calame, R., & Parker, K. (2003). Reclaiming youth and families with “Family ART.” *Reclaiming Children and Youth, 12*(3), 154–157.
- Crites E., & Taxman F. (2013). The responsivity principle: Determining the appropriate program and dosage to match risk and needs. In Taxman F., & Pattavina A. (Eds.), *Simulation Strategies to Reduce Recidivism*. Springer, New York, NY.
- Fox, A., & Veele, S. (2020). *Juvenile Rehabilitation Integrated Treatment Model*. Washington State Department of Children, Youth, and Families. <https://www.dcyf.wa.gov/sites/default/files/pdf/reports/jr-itm2020.pdf>
- Glick, B., & Goldstein, A. (1987). Aggression Replacement Training. *Journal of Counseling and Development, 65*(7), 356–362.
- Green, K. M., & Stuart, E. A. (2014). Examining moderation analyses in propensity score methods: Application to depression and substance use. *Journal of Consulting and Clinical Psychology, 82*(5), 773–783.
- Gundersen, K., & Svartdal, F. (2006). Aggression Replacement Training in Norway: Outcome evaluation of 11 Norwegian student projects. *Scandinavian Journal of Educational Research, 50*(1), 63–81.
- Hay, C., Widdowson, A., Bates, M., Baglivio, M., Jackowski, K., & Greenwald, M. (2019). Predicting recidivism among released juvenile offenders in Florida: An evaluation of the Residential Positive Achievement Change Tool. *Youth Violence and Juvenile Justice, 16*(1), 97–116.
- Knoth, L., Wanner, P., & He, L. (2019). *Washington State's Aggression Replacement Training for juvenile court youth: Outcome evaluation*. (Document No. 19-06-1201). Washington State Institute for Public Policy. https://www.wsipp.wa.gov/ReportFile/1707/Wsipp_Washington-States-Aggression-Replacement-Training-for-Juvenile-Court-Youth-Outcome-Evaluation_Report.pdf
- Linning, S., & Eck, J. (2018). Weak intervention backfire and criminal hormesis: Why some otherwise effective crime prevention interventions can fail at low doses. *The British Journal of Criminology, 58*(2), 309–331.

- Moynahan, L. (2003). Enhanced Aggression Replacement Training with children and youth with autism spectrum disorder. *Reclaiming Children and Youth, 12*(3), 174–180.
- Peterson, A. (2017). *Aggression Replacement Training in a probation setting: Outcomes for youths starting treatment January 2010–September 2012*. Washington State Center for Court Research, Administrative Office of the Courts. https://www.courts.wa.gov/subsite/wscrr/docs/ART_Outcomes_2016.pdf
- Roth, B., & Striepling-Goldstein, S. (2003). School-based Aggression Replacement Training. *Reclaiming Children and Youth, 12*(3), 138–141.
- Rosenbaum, P., & Rubin, D. (1985). The bias due to incomplete matching. *Biometrics, 41*(1), 103–116.

Appendix A

The Distribution of Propensity Scores Before and After Matching Using 3:1 Matching With Replacement and With a Caliper



Source: Stata/SE 15.1

EVALUATION OF AGGRESSION REPLACEMENT TRAINING (ART)

Table A1: Balance Between ART and Non-ART, Before and After Matching

	Before Matching			After Matching		
	ART	Non-ART	% Bias	ART	Non-ART	% Bias
Age at Admission	16.49	16.45	2.90	16.49	16.44	3.10
Gender	1.94	1.83	34.40	1.94	1.94	0.80
Race/Ethnicity						
African American	0.20	0.19	3.10	0.20	0.19	1.80
Asian	0.02	0.02	0.10	0.02	0.02	-1.30
White	0.44	0.43	2.80	0.44	0.44	-0.02
Hispanic	0.16	0.17	-4.20	0.16	0.15	1.90
Two or More Races	0.13	0.14	-2.10	0.13	0.14	0.09
Native American	0.04	0.04	-0.80	0.04	0.04	-1.10
Other	0.01	0.01	0.01	0.02	-1.80	-0.80
Most Serious Offense						
Person	0.44	0.45	-3.20	0.44	0.43	1.30
Property	0.29	0.29	-0.60	0.29	0.29	-0.03
Sex	0.19	0.13	15.00	0.19	0.20	-2.50
Drug	0.04	0.06	-10.10	0.04	0.03	0.80
Other	0.05	0.07	-6.60	0.05	0.05	1.10
Offense Class						
A	0.20	0.12	22.40	0.20	0.20	-0.20
B	0.55	0.57	-2.80	0.55	0.54	2.40
C	0.16	0.20	-10.30	0.16	0.18	-3.10
Other	0.08	0.11	-9.80	0.08	0.08	0.30
Sentence Type (JR or DOC)	1.97	1.98	-8.40	1.97	1.96	4.10
Domain 1RecordOfReferralRisk	14.95	14.02	17.60	14.93	15.07	-2.50
Domain 3ASchoolHistoryRisk	4.54	4.48	3.50	4.53	4.55	-0.80
Domain 3ASchoolHistoryPro0	0.23	0.23	0.80	0.23	0.23	2.00
Domain 3ASchoolHistoryPro2	0.71	0.72	-2.30	0.70	0.72	-2.60
Domain 3ASchoolHistoryPro4	0.06	0.05	3.00	0.06	0.06	1.40
Domain 3BSchoolCurrentRisk	5.88	6.34	-11.80	5.86	5.87	-0.04
Domain 3BSchoolCurrentPro	8.26	7.90	8.20	8.27	8.35	-1.90
Domain 4AFreeTimeHistoryRisk	0.14	0.11	8.60	0.14	0.13	4.20
Domain 4AFreeTimeHistoryPro	4.35	4.49	-7.90	4.35	4.40	-3.40
Domain 4BFreeTimeCurrentRisk	1.15	1.14	0.80	1.15	1.13	0.90
Domain 4BFreeTimeCurrentPro	4.92	4.87	2.80	4.93	4.99	-3.20
Domain 5AEmploymentHistoryRisk0	0.95	0.95	1.50	0.95	0.94	4.10
Domain 5AEmploymentHistoryRisk1	0.05	0.05	-0.30	0.05	0.06	-4.50
Domain 5AEmploymentHistoryRisk2	0.01	0.01	-3.30	0.01	0.01	0.40
Domain 5AEmploymentHistoryPro	0.86	0.91	-3.30	0.86	0.90	-2.10
Domain 5BEmploymentCurrentRisk	1.09	1.02	8.40	1.08	1.10	-1.10
Domain 5BEmploymentCurrentPro	2.69	2.55	7.00	2.70	2.68	1.00
Domain 6ARelationshipHistoryRisk0	0.15	0.15	-2.30	0.15	0.15	-1.10
Domain 6ARelationshipHistoryRisk1	0.05	0.04	2.90	0.05	0.04	1.30
Domain 6ARelationshipHistoryRisk2	0.50	0.52	-3.60	0.50	0.51	-3.00
Domain 6ARelationshipHistoryRisk3	0.31	0.29	4.50	0.31	0.29	3.50
Domain 6ARelationshipHistoryPro	1.19	1.16	2.00	1.19	1.18	0.80
Domain 6BRelationshipCurrentRisk	3.30	3.46	-7.00	3.28	3.28	-0.10
Domain 6BRelationshipCurrentPro	5.28	5.07	7.10	5.25	5.29	-0.10
Domain 7AFamilyHistoryRisk	3.71	3.84	-4.80	3.70	3.70	0.00
Domain 7AFamilyHistoryPro	3.06	3.10	-3.00	3.06	3.07	-0.60
Domain 7BLivingArrangementCurrentRi	8.20	8.36	-2.90	8.18	8.04	2.60

EVALUATION OF AGGRESSION REPLACEMENT TRAINING (ART)

Domain 7BLivingArrangementCurrentPr	12.43	12.02	8.50	12.44	12.55	-2.40
Domain 8AAcoholHistoryRisk	10.00	10.31	-4.00	9.96	9.99	-0.40
Domain 8AAcoholHistoryPro	3.94	3.72	9.70	3.94	4.02	-3.50
Domain 8BAcoholCurrentRisk	1.84	1.89	-2.40	1.82	1.79	1.40
Domain 8BAcoholCurrentPro	1.84	1.72	9.00	1.84	1.84	-0.30
Domain 9AMentalHealthHistoryRisk	4.96	4.96	0.10	4.96	5.08	-3.20
Domain 9AMentalHealthHistoryPro	8.95	9.31	-10.00	8.96	8.79	4.90
Domain 9BMentalHealthCurrentRisk	0.12	0.11	2.40	0.12	0.14	-5.70
Domain 9BMentalHealthCurrentPro	2.51	2.57	-4.70	2.52	2.53	-1.70
Domain 10AAttitudeHistoryRisk	1.20	0.99	8.40	1.13	1.21	-3.00
Domain 10AAttitudeHistoryPro	0.38	0.31	5.70	0.38	0.42	-3.50
Domain 10BAttitudeCurrentRisk	8.22	8.97	-10.40	8.24	8.24	0.00
Domain 10BAttitudeCurrentPro	14.06	13.68	5.80	14.10	13.88	3.40
Domain 11AggressionRisk	6.83	6.74	2.70	6.82	6.87	-1.80
Domain 11AggressionPro	6.62	6.51	3.20	6.63	6.50	3.80
Domain 12SkillsRisk	6.88	7.27	-7.00	6.86	6.95	-1.60
Domain 12SkillsPro	13.49	13.16	4.90	13.52	13.36	2.30

EVALUATION OF AGGRESSION REPLACEMENT TRAINING (ART)

Appendix B

Table A2: Impact of ART on Recidivism Using Propensity Score Matching

		Any Recidivism			Misdemeanor Recidivism			Felony Recidivism			Violent Felony Recidivism		
		OR	Sig	P Value	OR	Sig	P Value	OR	Sig	P value	OR	Sig	P Value
Non-ART		Ref			Ref			Ref			Ref		
ART		0.894		0.295	0.742	**	0.006	1.261		0.061	1.181		0.331
Age at Admission		0.924		0.148	0.899	*	0.049	1.018		0.780	0.974		0.754
Gender		1.727	**	0.007	1.080		0.716	1.848	*	0.015	2.942	*	0.011
Weeks Served		0.996	*	0.021	0.997		0.126	0.997		0.145	0.995	*	0.039
Race/Ethnicity													
	African American	Ref			Ref			Ref			Ref		
	Asian	0.763		0.461	1.148		0.702	0.651		0.244	1.072		0.871
	White	0.609	**	0.001	1.065		0.687	0.528	**	0.000	0.473	**	0.000
	Hispanic	0.655	*	0.018	1.147		0.453	0.550	**	0.002	0.537	*	0.014
	Two or More Races	0.741		0.111	1.057		0.767	0.665	*	0.039	0.675		0.120
	Native American	0.460	**	0.006	1.052		0.858	0.387	**	0.004	0.369	*	0.050
	Other	0.490		0.101	1.058		0.902	0.421		0.084	0.435		0.193
Most Serious Offense													
	Person	Ref			Ref			Ref			Ref		
	Property	1.571	**	0.000	1.269		0.063	1.235		0.122	0.750		0.115
	Sex	0.516	**	0.001	0.559	*	0.014	0.632		0.081	0.360	*	0.011
	Drug	1.247		0.396	1.017		0.948	1.304		0.366	1.263		0.533
	Other	0.716		0.165	0.794		0.359	0.882		0.655	0.696		0.382
Offense Class													
	A	Ref			Ref			Ref			Ref		
	B	1.342		0.118	1.268		0.285	1.213		0.380	1.138		0.663
	C	1.261		0.319	1.298		0.318	0.996		0.988	0.526		0.086
	Other	1.737	*	0.037	1.702		0.062	1.060		0.850	0.834		0.655
Parole		1.204		0.127	0.983		0.889	1.183		0.220	0.965		0.858
Sentence Type (regular = 1)		2.510	*	0.013	3.266	**	0.009	1.161		0.700	1.934		0.215
Release Year													
	2010	Ref			Ref			Ref			Ref		
	2011	0.772		0.263	0.795		0.303	1.043		0.873	1.127		0.742
	2012	1.117		0.720	0.834		0.541	1.665		0.153	2.850	*	0.022
	2013	0.891		0.637	0.846		0.485	1.202		0.547	1.214		0.663
	2014	1.234		0.383	0.904		0.670	1.811	*	0.050	2.152		0.065
	2015	0.940		0.799	0.609	*	0.035	2.045	*	0.017	3.347	**	0.002
	2016	1.184		0.482	0.737		0.197	2.178	**	0.008	2.543	*	0.018
	2017	1.134		0.629	0.600		0.053	2.397	**	0.007	3.865	**	0.002
Had Community Facility Time		0.957		0.718	0.916		0.473	1.089		0.516	1.189		0.327
Any Court ART Start		0.956		0.713	1.284	*	0.037	0.737	*	0.022	0.919		0.643
Domain 1RecordOfReferralRisk		1.047	**	0.000	1.006		0.671	1.057	**	0.000	1.049	*	0.017
Domain 3ASchoolHistoryRisk		0.987		0.820	0.982		0.759	1.020		0.779	0.895		0.204
Domain 3ASchoolHistoryPro0		Ref			Ref			Ref			Ref		
Domain 3ASchoolHistoryPro2		0.966		0.872	0.819		0.377	1.319		0.244	0.880		0.678
Domain 3ASchoolHistoryPro4		0.580		0.209	0.495	*	0.029	0.840		0.756	0.326		0.147
Domain 3BSchoolCurrentRisk		1.010		0.666	1.022		0.919	0.990		0.685	1.013		0.686
Domain 3BSchoolCurrentPro		0.970		0.168	1.017		0.435	0.936	*	0.012	0.954		0.188
Domain 4AFreeTimeHistoryRisk0		Ref			Ref			Ref			Ref		

EVALUATION OF AGGRESSION REPLACEMENT TRAINING (ART)

Domain 4AFreeTimeHistoryRisk1	0.770		0.178	0.845		0.096	0.916		0.685	1.520		0.149
Domain 4AFreeTimeHistoryPro	0.937		0.101	0.916	*	0.020	1.040		0.393	1.073	*	0.266
Domain 4BFreeTimeCurrentRisk	1.074		0.245	1.006		0.655	1.078		0.276	1.115		0.259
Domain 4BFreeTimeCurrentPro	1.005		0.906	0.967		0.932	1.034		0.493	1.120		0.069
Domain 5AEmploymentHistoryRisk0	Ref			Ref			Ref			Ref		
Domain 5AEmploymentHistoryRisk1	0.987		0.958	1.296		0.292	0.765		0.304	0.931		0.841
Domain 5AEmploymentHistoryRisk2	2.274		0.189	1.975		0.314	1.076		0.912	1.221		0.821
Domain 5AEmploymentHistoryPro	1.093	*	0.029	1.061		0.153	1.048		0.302	1.016		0.795
Domain 5BEmploymentCurrentRisk	1.014		0.844	0.986		0.839	1.029		0.719	1.039		0.719
Domain 5BEmploymentCurrentPro	0.944		0.124	0.993		0.842	0.943		0.177	1.031		0.600
Domain 6ARelationshipHistoryRisk0	Ref			Ref			Ref			Ref		
Domain 6ARelationshipHistoryRisk1	0.847		0.570	0.700		0.275	1.282		0.496	2.550	*	0.043
Domain 6ARelationshipHistoryRisk2	1.184		0.316	1.240		0.239	1.084		0.690	1.022		0.941
Domain 6ARelationshipHistoryRisk3	1.038		0.863	0.885		0.591	1.373		0.202	1.156		0.679
Domain 6ARelationshipHistoryPro	1.010		0.863	0.961		0.486	1.069		0.298	1.112		0.229
Domain 6BRelationshipCurrentRisk	1.050		0.250	1.000		0.993	1.051		0.258	1.152	*	0.014
Domain 6BRelationshipCurrentPro	0.997		0.912	1.016		0.594	0.964		0.235	0.960		0.318
Domain 7AFamilyHistoryRisk	1.015		0.697	0.993		0.859	1.027		0.527	0.923		0.163
Domain 7AFamilyHistoryPro	0.831	*	0.021	0.931		0.368	0.867		0.115	0.747	*	0.014
Domain 7BLivingArrangementCurrentRisk	0.995		0.687	1.006		0.621	0.990		0.444	0.990		0.573
Domain 7BLivingArrangementCurrentPro	1.016		0.249	1.011		0.468	1.009		0.587	1.040		0.093
Domain 8AAcoholHistoryRisk	1.020	*	0.034	1.021	*	0.023	1.001		0.933	0.992		0.554
Domain 8AAcoholHistoryPro	0.947		0.062	0.965		0.257	0.968		0.337	0.987		0.753
Domain 8BAcoholCurrentRisk	0.971		0.389	0.998		0.952	0.968		0.438	1.054		0.270
Domain 8BAcoholCurrentPro	0.964		0.470	1.049		0.364	0.904		0.082	0.965		0.645
Domain 9AMentalHealthHistoryRisk	0.991		0.772	0.976		0.416	1.034		0.329	1.047		0.302
Domain 9AMentalHealthHistoryPro	0.985		0.595	0.971		0.300	1.029		0.371	1.039		0.373
Domain 9BMentalHealthCurrentRisk	1.045		0.802	0.998		0.988	1.026		0.888	0.941		0.804
Domain 9BMentalHealthCurrentPro	0.942		0.287	1.021		0.713	0.886		0.060	0.853		0.078
Domain 10AAttitudeHistoryRisk	0.989		0.757	0.949		0.205	1.046		0.308	0.974		0.621
Domain 10AAttitudeHistoryPro	1.003		0.958	1.046		0.391	0.905		0.123	0.908		0.259
Domain 7BLivingArrangementCurrentPro	1.007		0.771	0.976		0.284	1.024		0.335	1.005		0.891
Domain 10BAttitudeCurrentPro	0.995		0.849	0.977		0.366	1.009		0.755	0.996		0.926
Domain 11AggressionRisk	0.998		0.963	1.069		0.085	0.903	*	0.033	0.966		0.590
Domain 11AggressionPro	0.973		0.487	1.019		0.639	0.936		0.136	0.916		0.139
Domain 12SkillsRisk	0.971		0.149	0.978		0.262	0.994		0.788	0.984		0.569

EVALUATION OF AGGRESSION REPLACEMENT TRAINING (ART)

Domain 12SkillsPro	0.979		0.194	0.967		0.055	1.011		0.559	0.997		0.891
N	2,432			2,432			2,432			2,434		
Pseudo R-squared	0.171			0.087			0.134			0.143		
* $p < .05$; ** $p < .01$												

Appendix C

Table A3: Sensitivity of Outcome Findings Using Alternative Matching Strategies

	Any Recidivism				Misdemeanor Recidivism			
	Non-ART	ART	% Point Diff.	Sig	Non-ART	ART	% Point Diff.	Sig
No matching	53.2%	53.0%	-0.1%	0.954	32.1%	28.5%	-3.6%	0.055
3:1 matching with replacement, with caliper	54.9%	52.7%	-2.2%	0.295	33.1%	27.4%	-5.7%	0.006
1:1 matching, no replacement, no caliper	54.1%	53.1%	-0.9%	0.654	32.0%	27.4%	-4.7%	0.018
1:1 matching, no replacement, with caliper	53.4%	53.2%	-0.2%	0.938	31.4%	27.4%	-4.0%	0.050
1:1 matching, with replacement, with caliper	55.6%	52.5%	-3.1%	0.187	34.2%	27.0%	-7.2%	0.003
	Felony Recidivism				Violent Felony Recidivism			
	Non-ART	ART	% point Diff.	Sig	Non-ART	ART	% Point Diff.	Sig
No matching	21.1%	24.4%	3.3%	0.049	9.7%	10.7%	1.0%	0.418
3:1 matching with replacement, with caliper	21.8%	25.3%	3.6%	0.061	10.1%	11.6%	1.4%	0.331
1:1 matching, no replacement, no caliper	22.2%	25.6%	3.5%	0.060	10.4%	11.5%	1.2%	0.398
1:1 matching, no replacement, with caliper	22.1%	25.7%	3.6%	0.061	10.2%	11.8%	1.6%	0.272
1:1 matching, with replacement, with caliper	21.5%	25.4%	3.9%	0.077	9.9%	11.4%	1.5%	0.348