Washington State Institute for Public Policy
110 Fifth Avenue SE, Suite 214 • PO Box 40999 • Olympia, WA 98504 • 360.586.2677 • www.wsipp.wa.gov

December 2015

The King County Education and Employment Training (EET) Program: *Outcome Evaluation and Benefit-Cost Analysis*

Over the past two decades, the Washington State Legislature has focused its operating budget on increasing the use of evidence- and research-based interventions in the juvenile justice system. The Washington State Institute for Public Policy (WSIPP) has evaluated several juvenile justice interventions to determine whether these programs reduce recidivism and whether the benefits outweigh the program costs.

In this report, we evaluate the effect of the Education and Employment Training (EET) program for moderate- to high-risk juvenile offenders. EET was originally established by the King County Juvenile Court in 1982.[1]

The report is organized as follows:

- Section I provides background on research-based interventions in juvenile justice and EET, specifically.
- Section II outlines WSIPP's methodology.
- Section III summarizes the key findings from our evaluation.
- Section IV presents our benefit-cost analysis of EET.

A Technical Appendix is provided for supplemental analysis and technical detail.

Summary

Education and Employment Training (EET) is a program, currently operating exclusively in King County, for juvenile offenders at moderate- to high-risk to re-offend.

For youth in school, the program provides job readiness training and connects youth with jobs—King County pays the wages. Youth who are not in school must re-engage in school, or the program provides assistance to prepare for General Equivalence Diploma (GED).

In 2010, EET was designated a "promising program" by the Community Juvenile Accountability Act oversight committee. At that time, the Washington State Institute for Public Policy agreed to evaluate the program when enough time had passed to measure the program's effect on recidivism. This study compares recidivism rates for youth served by EET to that of similar juvenile offenders served by other court programs in Pierce and Snohomish Counties.

We find that EET reduces overall recidivism by 12 percentage points, from 51% to 39%, compared to youth who participated in typical juvenile court programs.

The overall economic benefits of EET exceed the cost of providing the program to eligible youth. We estimate EET produces $34 in benefits per $1 of costs.

Suggested citation: Miller, M., Fumia, D., & He, L. (2015). *The King County Education and Employment Training (EET) Program: Outcome evaluation and benefit-cost analysis.* (Document Number 15-12-3901). Olympia: Washington State Institute for Public Policy.

[1] The program was initially called the Vocation, Education, Service, Training (VEST) Program and has evolved since it was established. Personal communication, Steven Noble, King County EET/Community Programs Supervisor.

I. Background

In 1997, the Washington State Legislature passed the Community Juvenile Accountability Act (CJAA).[2] This act established a goal of using research-based programs that cost-effectively increase juvenile accountability and reduce criminal recidivism.[3] State funding to county juvenile courts for research-based programs is administered by the Juvenile Rehabilitation Administration through a block grant. The funding formula used to distribute the block grant allocates 25% of these funds to programs defined as "evidence-based."[4]

Promising Programs.

In 2009, WSIPP assisted the CJAA block grant oversight committee with developing a protocol for courts to use state dollars for "promising programs"—programs that are similar to an evidence-based program but without the outcome evaluation evidence.

A court must first identify a given area of concern, such as lack of employment, for the population and then develop a program proposal. The CJAA oversight committee reviews the proposal and either accepts or rejects it. If a program proposal is accepted, the court implements the program. If the proposal is rejected, the court can choose to modify it. When enough time has passed, the program must be evaluated to determine if it qualifies as an evidence-based program.

In 2010, King County's Education Employment Training program (EET) was designated as promising. The legislature provided funding for WSIPP to conduct evaluations of criminal justice programs and WSIPP used these funds to evaluate this program.[5]

[2] RCW 13.40.500-540.

[3] RCW 13.40.500-510. In this context, research-based means a program has sufficient scientific evidence to conclude that it can reduce recidivism if properly implemented.

[4] Department of Social and Health Services, Juvenile Rehabilitation, (2013). "Evidence-based" under the funding formula generally means those programs that demonstrate a statistically significant reduction in recidivism. This definition more closely aligns with the "research-based" definition in RCW 71.36.010. We therefore use research-based throughout this report to reflect the legislative definition.

[5] Drake, E.K. (2010). *Washington State juvenile court funding: Applying research in a public policy setting* (Doc. No. 10-12-1201). Olympia: Washington State Institute for Public Policy.

Education and Employment Training (EET).

King County's EET program is designed for youth who are at least 15 years old and have been assessed as moderate- to high-risk on the Washington State Juvenile Court Assessment. All referrals to the program are made by the assigned juvenile probation counselor (JPC).

EET targets three assessment domains:

- Employment,
- School engagement, and
- Use of free time.

Participants in the EET program are assigned to an Education Employment Specialist (EES) who assesses engagement in school and their readiness for employment. Together, an offender and the EES develop an individual action plan.

For youth engaged in school, the program provides job readiness training and connects youth with jobs. The county pays the minimum wage for up to 20 hours per week—up to a total of 150 hours over a four- to six-month period. Wages also provide a means for juvenile offenders to make financial restitution to victims (if required by the court).

Youth who complete job readiness training receive either a $50 stipend or a gift card.

Youth not in school must re-engage in school or work towards obtaining a General Equivalence Diploma (GED) to continue in the program. The program provides assistance to prepare for the GED.

Each year, EET serves over 200 juvenile offenders.[6]

[6] According to *King County Children, Youth and Young Adult Services*" 230 youth were served in 2012. http://www.kingcounty.gov/~/media/operations/DCHS/2012_KC_Children_Youth_YA_Services_Rev8_31_12.ashx

II. Evaluation Methodology

Estimating the effect of EET on recidivism rates requires comparing EET participants (treated group) to a sufficiently similar group of individuals who are eligible for, but did not receive, EET (comparison group). Ideally, we would estimate this effect using an experimental research design where EET-eligible youth are randomly assigned to either the treated or comparison groups. In a well-implemented experimental design, assignment of eligible youth to the treated and comparison groups occurs only by chance, thus any differences in later outcomes could be confidently attributed to EET.

Without random assignment, however, we must consider that those who participate in EET could be less (or more) likely to recidivate even in the absence of participation due to some other factor unobservable to the researcher. For example, youth who are most motivated to reduce their criminal behavior or increase their chances of employment may be more likely to participate in the program. We would expect these youth to have lower recidivism rates because of their higher motivation regardless of participation.

Because random assignment did not occur in implementing EET, we rely on observational data to evaluate the program. We use an advanced statistical technique—propensity score matching—which can approximate group comparability on observed factors achieved with random assignment. We recognize, however, that propensity score matching may not eliminate all differences in unobservable characteristics.

Study Groups

The "treated group" includes individuals who started in EET between January 1, 2011 and December 31, 2012; these youth come from King County (the only court that has implemented EET). We include all youth who participated in EET regardless of completion.[7] The "comparison group" comes from the population of moderate- to high-risk juvenile offenders in Pierce and Snohomish Counties ("non-EET courts") who met other eligibility requirements and began another juvenile court program between January 1, 2011 and December 31, 2012.

Youth in both the treated and comparison groups were excluded from the analysis if they were younger than 15 at the start date. Because almost all moderate- and high-risk youth receive some sort of other block grant-funded, research-based programming, youth who received no program were eliminated from the data. Treated group participants were also excluded if:

1) Their most recent risk assessment prior to program start indicated low-risk;
2) They had no risk assessment recorded prior to starting the program;
3) Their most recent risk assessment occurred before July 2010; or
4) Youth could not be matched to court or risk assessment data.

[7] We estimate the treatment effect on the treated. We also retain all treated youth regardless of completion to avoid biasing the results toward the treated group because those that complete the program may also be more motivated and less likely to recidivate.

Finally, we excluded ten youth with missing data.[8]

We chose to use a comparison group from non-EET courts primarily because, as noted above, we are unable to control for unobserved factors that could impact a youth's participation in EET. Some youth with access to EET did not participate because they were assigned to another program or because their probation officer felt they were unsuited to the program. By drawing a comparison group from courts that do not offer EET, however, comparison group assignment is based on location rather than self-selection or selection on the part of others, such as parents or juvenile probation counselors. We chose to limit the comparison group to youth in Pierce and Snohomish Counties as they are among the largest counties in western Washington and are the counties most economically similar to King County.

Recidivism Measure

We define recidivism as any offense committed in the 18 months following the "at-risk" date that results in a Washington State conviction.[9] We define the "at-risk" date as the program start date for the EET group and the start date of the first program a comparison youth received between January 1, 2011 and December 31, 2012.

Methods

We use propensity score matching to select the matched comparison group from the pool of EET-eligible youth—i.e., moderate- to high-risk youth at least 15 years of age and meeting other EET criteria—in Pierce and Snohomish County courts. Propensity score matching has three steps. First, we estimate the propensity score—defined as the probability that a youth participates in EET—using a statistical model controlling for demographic, criminal, and behavior characteristics (see Exhibit 1 for the list of variables).

Second, we randomly sort the individuals and match each treated individual to the nearest comparison group individual with a similar propensity score. After matching, our final sample is 266 treated and 266 comparison group youth.[10]

Third, we perform an outcome analysis using the matched sample. We employ logistic regression to estimate the likelihood that a youth will recidivate, conditional on EET participation, as well as other variables included in the propensity score model.

Exhibit 1 reports the means and percentages for all variables used in the analysis for the treated and comparison groups before and after matching.[11] After matching, the two groups were very similar on all observed characteristics.

[8] Prior to propensity score matching there were 272 youth in the treated group and 628 in the comparison group.

[9] The recidivism measurement period includes the 18-month follow-up plus an additional six months to allow for adjudication. We consider juvenile diversion a conviction for the purposes of measuring recidivism. This definition is established in a legislatively directed study, see Barnoski, R. (1997). *Standards for improving research effectiveness in adult and juvenile justice.* (Doc. No. 97-12-1201). Olympia: Washington State Institute for Public Policy.

[10] We use 1:1 nearest neighbor matching without replacement.

[11] More detailed methods for this evaluation are described in the Technical Appendix.

Exhibit 1

Study Group Characteristics

Variable	Before matching[2]			After matching[2]		
	EET	Comparison group		EET	Comparison group	
Age	17.01	16.53	***	17.00	16.83	
Percent male	80%	78%		80%	78%	
Percent white	26%	59%	***	26%	40%	*
Percent black	42%	26%	***	42%	34%	*
Percent Latino	17%	8%	***	16%	13%	
Percent other race	16%	7%	***	16%	12%	
Criminal history score[1]	9.24	7.79	***	9.18	8.71	
Social history score[1]	8.14	8.65	**	8.18	8.15	
Number of previous juvenile court programs	0.68	0.33	***	0.64	0.51	
Whether youth:						
Is high-risk[1]	60%	54%		60%	57%	
Committed first offense before age 13[1,]	8%	13%	*	9%	11%	
Is law abiding (0/1)[1,]	24%	48%	***	24%	27%	
Is currently using alcohol/drugs[1]	72%	55%	***	72%	67%	
Is anti-social (0/1)[1]	72%	46%	***	72%	68%	
Was ever suspended/expelled from school[1]	90%	93%		91%	91%	
Demonstrates verbal aggression (0/1)[1,]	82%	57%	***	81%	76%	*
Demonstrates physical aggression (0/1)[1,]	89%	73%	***	89%	86%	
Demonstrates violent or sexual aggression	50%	23%	***	49%	43%	*
Is employed[1]	11%	12%		11%	11%	
Is in school[1]	89%	84%	*	88%	85%	
Whether youth started program in 2011 (0/1)	47%	56%	*	47%	49%	
Number of youth	272	628		266	266	

Notes:

[1] These measures come from the juvenile risk assessment; see Barnoski, R. (2004). *Washington state juvenile court assessment manual, version 2.1* (Doc. No. 04-03-1203). Olympia: Washington State Institute for Public Policy.

[2] Stars indicate statistical significance; * p<0.1; ** p<0.05; *** p<0.01.

III. Evaluation Findings

We analyzed the effect of EET participation on the following outcomes:

- Total recidivism (any misdemeanor or felony conviction),
- All felony recidivism,
- Violent felony recidivism, and
- Misdemeanor recidivism.

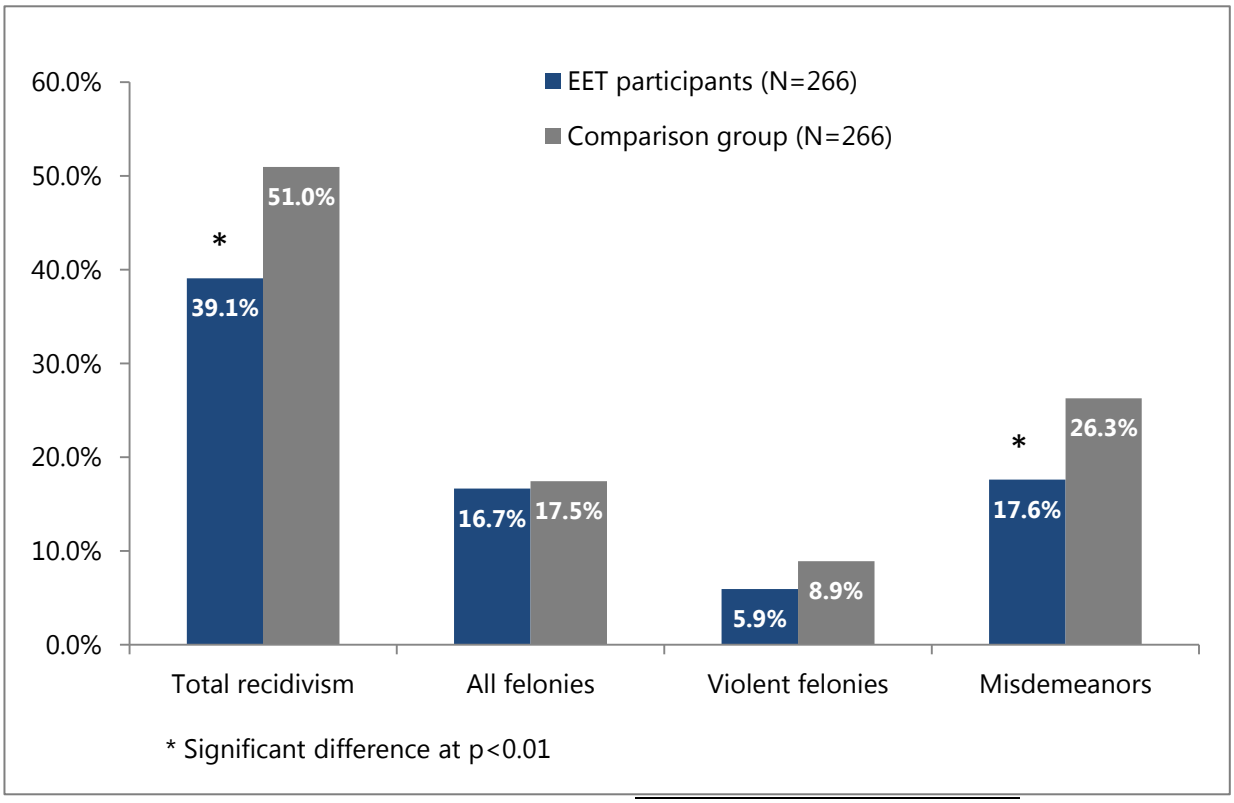
While the program focuses on education and employment, we are unable, at this time, to estimate directly the effect of EET on high school graduation or later employment.

We display our regression-adjusted recidivism rates in Exhibit 2. We find that EET reduces the likelihood of recidivism across all four recidivism measures; results are statistically significant for misdemeanor and total (misdemeanor plus felony) recidivism.[12]

Overall, EET reduces recidivism by 12 percentage points (39% for the treated group compared to 51% for the comparison group). The effect of EET on total recidivism is due mainly to the program's effect on misdemeanor recidivism.

Exhibit 2

18-month Adjusted Re-Conviction Rates Across Treatment Status



[12] We used bootstrapping to arrive at the standard errors for determining statistical significance. Bootstrapped standard errors, as opposed to analytical standard errors, allow us to take into account the fact that the propensity score is estimated in our outcome analysis. Analytic standard errors were slightly smaller than those from bootstrapping; but, as seen in Exhibit A4 in the Technical Appendix, they did not affect our conclusions about statistical significance.

IV. Benefit-Cost Analysis

WSIPP also considers the benefits and costs associated with implementing a program. WSIPP's benefit-cost model provides an internally consistent monetary valuation so program and policy options can be compared on an apples-to-apples basis.[13] Our benefit-cost results are expressed with standard financial statistics—net present values and benefit-cost ratios.

In benefit-cost analyses of juvenile justice programs, reductions in recidivism produce benefits to program participants, crime victims, taxpayers, and other people in society. Reductions in recidivism also produce benefits through avoided costs of crime. Crime produces many costs, including those associated with the criminal justice system as well as those incurred by crime victims. When crime is avoided, these reductions lead to monetary savings or benefits to victims and taxpayers. WSIPP's benefit-cost model estimates the number and types of crimes avoided (or incurred) due to the effect of a policy and the monetary value associated with that reduction.

Although we were unable to measure effects of EET on high school graduation, we know from other research that juvenile offenders who reduce their probability of recidivism experience increases in rates of high school graduation.[14] Such increases benefit youth through increased employment as well as others in society through greater tax revenue and other positive "spillover" effects. Higher rates of high school graduation can also lead to changes in healthcare coverage, as those with high school diplomas are more likely to use private or employer-sponsored health insurance rather than publicly-provided healthcare. We include these benefits in our estimates.

Finally, to account for the inherent uncertainty associated with any statistical or benefit-cost analysis, we perform a "Monte Carlo simulation" in which we vary key factors in our calculations. We can then estimate the degree of risk associated with our estimates. More details on our benefit-cost analysis methods can be found in WSIPP's Technical Documentation.[15]

Exhibit 3 shows the results of our benefit-cost analysis. EET costs $2,857 per youth (in 2014 dollars). We estimate comparison group costs as the weighted average of the alternative programs youth received—$2,006.[16] Participation in EET results in total benefits from increased high school graduation and avoided crime of $29,361 shown in Exhibit 3. Thus, we estimate total net benefits of $28,510. Our risk analysis indicates that EET will yield positive net benefits 100% of the time.

The legislature has identified a three-tiered classification to identify effective programs for children and youth. "Evidence-based" programs—the top tier—are those that have been rigorously evaluated more than once.

[13] Washington State Institute for Public Policy (December 2015). *Benefit-cost technical documentation*. Olympia, WA: Author http://www.wsipp.wa.gov/TechnicalDocumentation/WsippBenefitCostTechnicalDocumentation.pdf

[14] Ibid.

[15] Ibid.

[16] Barnoski, R. (2009). *Providing evidence-based programs with fidelity in Washington State juvenile courts: Cost analysis* (Doc. No. 09-12-1201). Olympia: Washington State Institute for Public Policy.

"Research-based" programs are those that have "some research demonstrating effectiveness, but that does not yet meet the standard of evidence-based practices."[17] "Promising programs" are those that, based upon preliminary information, have potential for becoming a research-based practice.

Based on the findings from this evaluation, we identify EET as a research-based program. That is, the weight of the evidence indicates a significant reduction in recidivism. Additionally, EET produces cost-beneficial outcomes. The program is not evidence-based because there has not been more than one evaluation.

Exhibit 3

Benefits and Costs per Participant for EET vs. Comparison Group (in 2014 Dollars)

Program cost	
EET participants	
Per youth expenditures reported in *King County Children, Youth and Young Adult Services*[1]	$2,857
Comparison group costs	$2,006
Weighted average cost of programs provided to youth in comparison group	
(1) Net EET cost	**-$851**
Recidivism effects	
Decreased taxpayer costs due to decreased recidivism	$6,276
Decreased crime victim costs due to decreased recidivism	$13,834
Decreased deadweight cost of taxation due to decreased taxpayer criminal justice cost	$3,123
Health care-related effects	
Increased healthcare insurance costs to participants due to moving from public to employer or private insurance	-$97
Decreased healthcare insurance costs to taxpayers due to movement from public to employer or private insurance	$353
Decreases deadweight cost of taxation due to decreased taxpayer health care costs	$176
Increased costs to private or employer-sponsored insurance programs	-$388
High school graduation effects	
Increased income to participants due to increased labor market participation	$3,011
Increased tax revenue to taxpayers due to increased labor market participation	$1,284
Positive externalities to society due to greater number of high school graduates	$1,488
Deadweight cost of taxation for program costs	**-$426**
(2) Total benefits	**$29,361**
Bottom line:	
Total net benefits (cost) per participant (3) Net (benefits – costs)	**$28,510**
Benefit-to-cost ratio	**$34.50**
Probability of positive net benefits (risk analysis)	**100%**

[1] Downloaded from http://www.kingcounty.gov/~/media/operations/DCHS/2012_KC_Children_Youth_YA_Services_Rev8_31_12.ashx

[17] RCW 71.36.010.

V. Conclusions and Limitations

Based on this single evaluation, we find that EET participants have lower rates of total recidivism and that EET is cost-beneficial. Therefore, EET meets the criteria of a research-based program.

In this analysis, we compared youth receiving EET to youth receiving other evidence-based programs. It is likely that had we been able to compare outcomes for EET youth with outcomes for similar youth who received no program, the effect on recidivism would have been greater. In that case, however, the cost of the comparison group would be zero so that the net cost of EET would have also been greater.

Our analysis could not entirely eliminate the possibility of selection bias among youth receiving EET. Juvenile probation counselors (JPC) have some discretion in the referral of youth to programs. JPCs may judge that some youth who are eligible for EET would be better served by other programs. Further, engagement in school or working toward a GED is a requirement of EET. Youth who choose not to engage in school, or who are not interested in job-related skills, would likely be referred to another program. Information on JPCs' decisions and youth motivation and interest were not available for this evaluation. Even using propensity score matching, we cannot account for the reasons a King County youth may or may not be assigned to EET.

In the future, it may be possible to conduct a random assignment study to further evaluate the effect of the program without the possible selection bias, which we were not able to rule out for this report. In such a study, youth eligible for, and willing to participate in, EET would be assigned at random to EET or another research-based program. This would remove the selection bias due to JPC judgement and youth preference.

Technical Appendix

The King County Education and Employment Training (EET) Program: Outcome Evaluation and Benefit-Cost Analysis

Appendix

A. I. Study Group Selection & Matching Procedures

In an ideal research design, offenders eligible for EET would be randomly assigned to EET or a untreated comparison group. With a successfully implemented random assignment, any observed differences in recidivism could be attributed to the effect of EET. Unfortunately, as is the case in many real world settings, random assignment was not possible for this evaluation. Further, almost all moderate- to high-risk juvenile offenders receive some block-grant funded research-based programming, so that such an experiment might be considered unethical.

Instead, we use observational data and rely on a quasi-experimental research design. Unlike random assignment, this type of design cannot eliminate the risk that selection bias or unobserved factors may threaten the validity of the findings. For example, juvenile probation counselors (JPCs), parents, or the youth themselves can base participation on the youth's likelihood of success or motivation. If youth who participate in EET are more motivated, this unobserved factor would bias the results in favor of the treated group. However, if youth in EET are referred to the program because they are perceived as worse off than youth referred to other programs, this selection would bias the results toward the comparison group.

To infer causality from this quasi-experimental study, selection bias must be minimized. To do so, we implement a variety research design methods and statistical techniques that provide the ability to test the sensitivity of our findings. In this section of the Technical Appendix, we describe the study groups and statistical methods we use to arrive at estimates of the effects of EET.

Study groups

We draw our EET participant pool from youth who started EET between January 1, 2011 and December 31, 2012, as reported by the King County Juvenile Court. Comparison group youth are drawn from the population of moderate- to high-risk youth meeting all EET eligibility requirements who began another program between January 1, 2011 and December 31, 2012 in Pierce and Snohomish Counties, where EET is not offered.

We draw the comparison group from a pool of offenders in Pierce and Snohomish courts to ensure that program availability is the primary driver of comparison group assignment. That is, youth from non-EET courts do not participate in the program simply because it is not offered. If we draw a comparison group from King County, however, we cannot easily determine the reasons that a non-EET youth did not participate. Youth may not participate in EET for many reasons. If the researcher cannot observe the reasons for nonparticipation, however, then selection bias would be a serious concern. However, youth from non-EET courts do not have the

option to participate and so we can infer that their nonparticipation has to do with their location rather than unobserved characteristics.

Drawing comparison youth from non-EET courts does not completely eliminate selection bias. Those that participate in EET may still be more motivated than the average comparison group youth, or their JPCs may have felt they were best-suited to EET, for example. Further, the program also requires that youth be engaged in school or working toward obtaining a GED. From the assessment data available to us, we are unable to identify motivation to reduce recidivism, obtain employment or graduate from high school.

Even drawing a comparison pool from Pierce and Snohomish Counties rather than King County, we are unable to identify similarly motivated youth from those counties. We still cannot prevent selection into the treatment group on the part of JPCs, youth, parents, lawyers, etc. Thus, it is important we are able to balance the treated group with comparison youth from non-EET courts on all observables.

Propensity Score Matching

While using a comparison group from other locales and selection bias pose possible threats to the validity of a study, we attempt to minimize these influences using propensity score matching. Propensity score matching allows us to match treated individuals with similar comparison group individuals to obtain balance on observed covariates. This method has many benefits over standard regression analysis, which is often used to control for differences between treated and comparison groups.

First, the outcome plays no part in matching the treated and comparison groups. This emulates an experimental design by separating the research design stage—where we test various matching procedures to obtain a sufficiently matched sample—from the analysis stage—where we estimate the effect of the treatment using our matched sample. Second, matching can limit the importance of functional form in regression analysis.[18] Third, by imposing common support restrictions, we ensure that the comparison group does not differ substantially in their likelihood to participate in EET, i.e. we are not comparing treated youth to youth who we would never expect to participate in EET. Finally, by conducting a logistic regression on the matched sample using the covariates from the matching model, we further reduce any residual bias that may remain after matching and account for any correlation between matched pairs.

Exhibit A1 below reports the results from the coefficients from the first stage model estimating the likelihood of EET participation. We control for demographic characteristics, criminal history and social history scores, and behavior variables from the assessment data. Ideally, to address the fact that treated and comparison group youth come from different locales, we might control for various county and court characteristics. We were not able to do so in this case, however, because EET is offered only in King County and, thus, EET status is perfectly correlated with county.

[18] Ho, D.E., Imai, K., King, G., & Stuart, E.A. (2007). Matching as nonparametric preprocessing for reducing model dependence in parametric causal inference. *Political Analysis*, *15*(3), 199-236.

Exhibit A1

Logit Model Estimating the Likelihood of EET Participation

Covariate	Coefficient		SE
Age	0.396	***	0.082
Male (0/1)	1.629		0.283
Black (0/1)[#]	-0.081	***	0.220
Latino (0/1)[#]	1.253	***	0.211
Other race (0/1)[#]	1.844	***	0.295
Criminal history score	0.040		0.028
Social history score	-0.221	***	0.042
Number of previous juvenile court programs	0.481	***	0.130
Whether youth:			
Is high-risk	0.516	**	0.254
Committed first offense before age 13	-0.583	*	0.299
Is law abiding (0/1)	-0.206		0.255
Is currently using alcohol/drugs	0.617	***	0.207
Is anti-social (0/1)	0.391		0.248
Was ever suspended/expelled from school	-0.579	*	0.337
Demonstrates verbal aggression (0/1)	0.952	***	0.258
Demonstrates physical aggression (0/1)	-0.029		0.300
Demonstrates violent or sexual aggression (0/1)	0.857	***	0.195
Is employed	-0.172		0.286
Is in school	0.762	***	0.260
Whether youth started program in 2011 (0/1)	-0.331	*	0.179
Constant	-8.651	***	1.538
Number of youth	899		
Pseudo R2	0.401		
AUC	0.84		

Notes:

Stars indicate statistical significance; * p< 0.1; ** p<0.05; *** p<0.01

[#]Reference group is white youth.

Our preferred matching procedure for the main analysis is 1:1 nearest neighbor matching without replacement. Using 1:1 matching can reduce the bias between the treated and comparison groups by only matching treated individuals with the most similar comparison group individual.

In some cases, where the number in the comparison group pool greatly exceeds the number treated, it may be possible to match two or more comparison youth to each person in the treatment group. In this case, however, the comparison pool was only about twice the size of the treatment group, so 1:1 matching was necessary.

We used various diagnostics to determine the extent to which the propensity score matching improved balance between the treated and comparison groups. A common measure of balance is the standardized difference (or bias) calculated as the difference in the mean/proportion for the treated and comparison groups divided by the pooled standard deviation for each covariate prior to matching. This measure is preferred to traditional t-tests as the standardized difference is not influenced by the study's sample size. Additionally, t-tests are used for making inferences about a population based on a sample; balance, on the other hand, is an in-sample property. Standardized bias values greater than 0.10 usually indicate moderate imbalance while greater than 0.25 indicates severe imbalance.[19] Exhibit A2 displays the standardized bias for each covariate in the propensity score model before and after matching as well as the p-value as a reference. After matching, most differences were greatly reduced although some moderate bias remained. We control for the bias in the logistic regression. This last step is used to "'clean up' residual covariate imbalance between groups."[20]

[19] Austin, P.C. (2009). Balance diagnostics for comparing the distribution of baseline covariates between treatment groups in propensity-score matched samples. *Statistics in Medicine, 28*(25), 3083-3107 and Stuart, E.A. (2010). Matching methods for causal inference: A review and a look forward. *Statistical Science: A Review Journal of the Institute of Mathematical Statistics*, *25*(1), 1–21.
[20] Stuart, E.A. (2010). Matching methods for causal inference: A review and a look forward. *Statistical Science, 25*(2),1-21.

Exhibit A2

Matched Study Groups Characteristics

Variable	Means and proportions after matching			Absolute standardized difference (d)	
	EET	Comparison group	p-value	Before matching[1]	After matching
Age	17.00	16.87	0.206	**0.43**	*0.11*
Percent male	80%	79%	0.80	0.04	0.01
Percent black	41%	34%	0.09	*0.24*	0.10
Percent Latino	16%	14%	0.39	*0.17*	0.05
Percent other race	16%	13%	0.27	*0.20*	0.07
Criminal history score[2]	9.19	8.68	0.16	**0.38**	*0.12*
Social history score[2]	8.19	8.17	0.93	*0.18*	0.01
Number of previous juvenile court programs	0.65	0.50	0.02	**0.42**	*0.24*
Whether youth:					
Is high-risk (0/1)[2]	60%	57%	0.54	0.10	0.04
Committed first offense before age 13	9%	12%	0.25	0.10	0.07
Is law abiding (0/1)[2]	24%	29%	0.20	**0.37**	0.08
Currently uses alcohol/drugs (0/1)[2]	72%	68%	0.34	*0.24*	0.06
Is anti-social (0/1)[2]	72%	66%	0.13	**0.40**	0.09
Was ever expelled/suspended from school (0/1)[2]	91%	91%	0.88	0.07	0.01
Demonstrates verbal aggression (0/1)[2]	81%	75%	0.09	**0.39**	0.09
Demonstrates physical aggression (0/1)[2]	89%	85%	0.20	**0.29**	0.06
Demonstrates violent or sexual aggression (0/1)[2]	49%	42%	0.10	**0.42**	0.06
Is employed (0/1)[2]	11%	10%	0.78	0.03	0.02
In school (0/1)[2]	88%	85%	0.25	0.09	0.07
Started program in 2011 (0/1)[2]	47%	49%	0.66	*0.13*	0.03
Number of youth	267	267			

Notes:

[1] **Bold** text indicates severe imbalance, |d| > 0.25; *italics* text indicates moderate imbalance, |d| >0.1.

[2] These measures come from the juvenile risk assessment developed by the Washington Association of Juvenile Court Administrator and the Washington State Institute for Public Policy.

[3] From the Summary Reporting System obtained from the Office of Financial Management Washington State Criminal Justice Data Book (http://wa-state-ofm.us/CrimeStatsOnline/index.cfm).

[4] Obtained from the Office of Juvenile Justice and Delinquency Prevention "Easy Access to Juvenile Populations" (http://www.ojjdp.gov/ojstatbb/ezapop/).

[5] Calculated from juvenile justice system administrative data obtained from the Administrative Office of the Courts.

A. II. Methods to Estimate the Effects of EET

Timing of Recidivism

Recidivism is defined as any offense committed in the 18 months after program start for EET and comparison youth that resulted in a Washington State conviction.[21] In addition to the follow-up period, time is needed to allow an offense to be processed in the criminal justice system. The criminal justice process also includes the adjudication period—the time period between the date recorded for the commission of a subsequent offense and the resulting conviction for that offense. This analysis allows for a 12-month adjudication period as suggested by Barnoski.[22]

For this analysis, we consider the "at-risk" period to begin on the date youth begin the program (EET or, in the case of the comparison group, the date they began another juvenile court program).

Logistic Regression Analysis on Full (Unmatched) Sample

We begin our outcome analysis using traditional multivariate logistic regression analysis on the full (i.e. unmatched) sample. Regression analysis allows us to control for observed covariates in estimating the treatment effect. However, regression analysis has several limitations. First, regression analysis can only control for observed factors. Second, if treated and comparison group covariate distributions do not overlap, then any causal inferences for regions with few treated or control group members must be based on extrapolation, leading to less precise estimates. Third, to approximate an experimental design, the research design stage of an evaluation should be separate from the outcome analysis stage. With standard regression analysis, the outcome of interest is necessarily part of the regression model and determining model fit requires repeatedly estimating the treatment effect.[23] This can lead to model selection based on the observed treatment effect and also suffers from the multiple comparisons problem, where the likelihood of finding a statistically significant result increases with the number of statistical tests performed. Finally, regression analysis requires making assumptions about functional form, which can increase bias if the wrong functional form is used.

While regression analysis has several limitations, it can outperform matching methods if important unobserved covariates are omitted from the analysis. In this case, regression analysis will produce a less biased estimate than propensity score matching. For this reason, we first estimate the relationship between EET participation and recidivism using standard logistic regression. Row 3) of Exhibit A5 reports the regression-adjusted recidivism rates for the unmatched sample. The effects using standard logistic regression indicate that EET participation reduces recidivism by about 14.6 percentage points, a slightly larger reduction than in our matched sample of 11.4 percentage points (Row 4)).

Outcome Analysis: Logistic Regression on Matched Sample

Our preferred analysis uses logistic regression on the matched sample to estimate the effect of EET on total, felony, misdemeanor recidivism and violent felony recidivism. Our outcome model uses most of the same covariates included in the matching model. Some variables were omitted from the analysis because they were highly correlated with other covariates. Results of the analyses are reported in Exhibit A4.

[21] Barnoski, (1997), pg. 2.
[22] Barnoski, (1997), pg. 4.
[23] Rubin, D.B. (2007). The design versus the analysis of observational studies for causal effects: parallels with the design of randomized trials. *Statistics in medicine*, *26*(1), 20-36.

Exhibit A4

Logistic Regression Estimating Effect of EET on Recidivism
(EET participant N = 267, Comparison group N = 267)

Covariate	Total recidivism		Felony recidivism		Misdemeanor recidivism		Violent felony recidivism	
	Odds ratio	p-value[#]	Odds ratio	p-value[#]	Odds ratio	p-value[#]	Odds ratio	p-value#
EET participation	0.617	0.015	0.946	0.946	0.599	0.020	0.645	0.145
Age	0.908	0.275	0.872	0.872	1.003	0.976	0.940	0.635
Male (0/1)	2.518	0.000	3.415	3.415	1.251	0.425	4.518	0.006
Black (0/1) [##]	1.746	0.017	4.030	4.030	0.663	0.101	5.596	0.000
Latino (0/1) [##]	0.769	0.404	2.024	2.024	0.455	0.031	3.905	0.009
Other race (0/1) [##]	1.340	0.349	3.829	3.829	0.539	0.097	4.450	0.008
Criminal history score	1.070	0.012	1.051	1.051	1.022	0.449	1.028	0.423
Social history score	1.096	0.017	1.143	1.143	0.989	0.802	1.073	0.203
Number of previous juvenile court programs	1.153	0.294	1.345	1.345	0.929	0.625	1.251	0.252
Whether youth:								
Committed first offense before age 13	1.087	0.803	0.926	0.926	1.262	0.503	0.700	0.482
Is currently using alcohol/drugs	1.030	0.898	1.025	1.025	1.040	0.877	1.174	0.655
Is anti-social (0/1)	1.610	0.047	1.115	1.115	1.831	0.035	1.013	0.972
Was ever suspended/expelled from school	1.095	0.809	0.638	0.638	1.791	0.229	0.472	0.192
Demonstrates verbal aggression (0/1)	2.232	0.008	1.946	1.946	1.786	0.100	8.424	0.003
Demonstrates physical aggression (0/1)	0.828	0.595	1.568	1.568	0.549	0.135	0.768	0.723
Is employed	0.857	0.624	0.361	0.361	1.652	0.130	0.344	0.099
Is in school	0.701	0.225	0.789	0.789	0.791	0.451	0.748	0.449
Started program in 2011 (0/1)	1.135	0.511	1.275	1.275	1.037	0.865	1.231	0.481
Constant	0.261	0.418	0.044	0.123	0.172	0.341	0.006	0.046
Pseudo-R2	0.208		0.244		0.093		0.218	
AUC	0.734		0.776		0.666		0.806	

Notes:

[#]P-values based on analytical standard errors rather than bootstrapped standard errors. The analytical standard errors are slightly smaller than those from bootstrapping but had a minimal effect on significance; however, our main findings, reported in Exhibit A5, use bootstrapped standard errors.

[##]Reference group is white youth.

Exhibit A5

Effects of EET With and Without Matching[1]

Matching method	Total recidivism				Felony recidivism				Misdemeanor recidivism				Violent felony			
	EET	Comp[2]	% point difference[3]	SE[4]	EET	Comp[2]	% point difference[3]	SE[4]	EET	Comp[2]	% point difference[3]	SE[4]	EET	Comp[2]	% point difference[3]	SE[4]
Unadjusted recidivism rates																
(1) Unmatched	42.3%	45.2%	-2.9	3.6	23.5%	17.8%	5.7**	3	18.8%	27.4%	-8.6***	3	9.6%	11.8%	-2.2	2.3
(2) Matched	42.5%	49.2%	-6.8	4.3	23.7%	21.1%	2.6	3.6	18.8%	28.2%	-9.4***	3.6	12.0%	13.5%	-1.5	2.9
Regression adjusted recidivism rates																
(3) Unmatched	35.3%	47.1	-11.8***	4.2	15.2%	14.9%	0.2	3.1	17.0%	26.8%	-9.7***	3.6	5.0%	7.4%	-0.2	2.1
(4) Matched	39.1%	51.0%	-11.9***	4.1	16.6%	17.5%	-0.8	3.5	17.6%	26.3%	-8.7***	3.8	5.9%	8.9%	3.0	2.9

Notes:

[1]Unweighted sample sizes are as follows:

Unmatched raw (Treated N = 271, Comparison N = 628); Unmatched regression adjusted (Treated N = 271, Comparison N = 628); 1:1 Nearest neighbor without replacement (both raw and regression adjusted Treated N = 266 Comparison N = 266).

[2] Comparison group

[3]Stars indicate statistical significance; * p< 0.1; ** p<0.05; *** p<0.01.

[4]Standard errors are expressed as percent. For raw (unadjusted) recidivism rates and the regression-adjusted unmatched recidivism rates, standard errors are calculated using the formula:

$$SE = \sqrt{\frac{p_1(1-p_1)}{N_1} + \frac{p_2(1-p_2)}{N_2}}$$

For regression adjusted matched rates, bootstrapped standard errors are presented.

[5]Raw recidivism rates are differences in mean recidivism rates for treated and comparison groups without regression adjustment. Matching on covariates was still used to obtain a matched raw recidivism rate.

Three issues warrant further discussion: 1) treatment group selection, 2) comparison group selection, and 3) standard error estimation.

Treatment Group

In discussions with court personnel, it was suggested that one program unit might not have adequate quality assurance and, thus, should be omitted from the analysis. We conducted analyses three ways to determine whether inclusion of this unit would change our results. These analytical approaches were:

1) Omitting the unit,
2) The unit by itself, and
3) The entire sample.

We found that the conclusions were the same with any approach. We therefore chose to report on the entire sample because the increased sample size increases the statistical power of the findings.

Comparison Group

While untreated youth from within EET courts face the same court and county factors as treated youth, we chose our final comparison group from courts where EET was not available because we could not fully explain why youth in King County did not receive EET when the program was available to them. In other words, we were more concerned about the potential for selection bias in the comparison group when drawing that comparison group from the King County courts.

By using a comparison group from Pierce and Snohomish courts, which did not offer EET, we cannot determine the extent to which the effect we observe is partially due to characteristics that may differ between the King County court and Pierce and Snohomish courts.

Ideally, we would conduct a fixed effects regression which would control for unobserved characteristics that differ across courts, but because court is perfectly predictive of the treatment, we cannot include it in the propensity score model.

Further, the Washington State Juvenile Court assessment identifies programs for which youth are eligible. Sometimes, moderate- to high-risk youth are eligible for more than one program. It is unclear how the final decision of program assignment is made.

Standard Errors

In propensity score matching, the problem of obtaining correct standard errors often arises. Analytical formulas for the standard error such as those from logistic regressions ignore the error associated with estimating the propensity score and the correlation of the matched sample.[24] Thus, the analytical standard errors based on matched data may be inaccurate. To address this issue, we use bootstrapping methods to estimate the standard error of the regression-adjusted effect of EET. Bootstrapping means repeatedly drawing *N* random samples from the matched sample with replacement and computing an effect of EET for each sample using the methods described in the outcome analysis section above. Then, the variance of

[24] Hill, J. (2008). Discussion of research using propensity-score matching: Comments on 'A critical appraisal of propensity-score matching in the medical literature between 1996 and 2003' by Peter Austin, Statistics in Medicine. *Statistics in Medicine*, *27*(12), 2055-2061.

the effect of EET is measured by estimating the variance in the estimated effects of EET across the *N* samples.

Research suggests that bootstrapping standard errors for matched data may only be appropriate in some situations. First, bootstrapping can only be used for population inference rather than in-sample estimates.[25] Second, when performing regression analyses on matched data, it may be unnecessary to employ bootstrapping when the regression analysis includes the covariates in the matching model. The correlation caused by the matched sample design will already be accounted for by regressing the outcome on the treatment and the covariates used in the matching model.[26] Finally, bootstrapping may be inappropriate for nearest neighbor matching with replacement,[27] although these concerns do not apply to matching without replacement,[28] which is our chosen method for this analysis. Given the tradeoffs between underestimating standard errors and using inappropriate methods for correction, we also examined the sensitivity of our conclusions using analytical standard errors.

In this sample, analytical standard errors were only slightly smaller than those obtained through bootstrapping and did not affect our conclusions about statistical significance of effects of EET on recidivism.

[25] Austin, P.C., & Small, D.S. (2014). The use of bootstrapping when using propensity score matching without replacement: a simulation study. *Statistics in medicine*, *33*(24), 4306-4319.
[26] Ho et al., (2007) and Gelman, A., & Hill, J. (2006). *Data analysis using regression and multilevel/hierarchical models*. Cambridge University Press.
[27] Abadie, A., & Imbens, G.W. (2008). On the failure of the bootstrap for matching estimators. *Econometrica*, *76*(6), 1537-1557.
[28] Austin & Small, (2014).

Acknowledgements:

The authors would like to thank staff at the Administrative Office of the Courts, the Juvenile Rehabilitation Administration/DSHS, members of the CJAA Advisory Board, and Steven Noble at the King County Juvenile Court.

For further information, contact:
Marna Miller 360.586.2745, marna.miller@wsipp.wa.gov

Document No. 15-12-3901