

**Do the benefits of youth mentoring persist over time?  
An evaluation of program effects among young adolescents**

Liliokanaio Peaslee, Ph.D.

Amanda C. Teye, Ph.D.



Paper prepared for the Western Political Science Association Annual meeting, April 18-20, 2019; San Diego, CA.

## **Introduction and Overview**

One-to-one youth mentoring programs are a popular strategy to positively influence the life trajectories of at-risk children and youth. Mentoring programs can affect a range of youth outcomes including academic performance, emotional and behavioral wellbeing, family relationships, substance use and other risky behaviors, and juvenile justice involvement (Converse & Lignugaris/Kraft, 2009; Dubois et al., 2011; DuBois & Karcher, 2014; DuBois & Silverthorn, 2005; Herrera et al., 2013; Krebs, Lattimore, Cowell, & Graham, 2010; Rhodes, 2008; Tierney & Grossman, 2000), although program effects have been inconsistent across mentoring programs and outcomes of interest. A significant limitation in youth mentoring research is that few extant evaluations include long enough time horizons to examine whether positive benefits from mentoring are sustained over time (DuBois et al., 2011). This paper presents findings from a five-year quasi-experimental study designed to expand knowledge about the long-term impacts of high quality mentoring programs. Participants included young adults (ages 12 to 16) who received one-to-one youth mentoring as children and a comparison group of youth who were referred to the agency but never received treatment. The project was funded by a major grant award from the Office of Juvenile Justice and Delinquency Prevention (OJJDP) and was conducted in conjunction with an affiliate agency of Big Brothers Big Sisters of America.

We begin with a brief overview of the literature on mentoring and youth outcomes before highlighting our methods, including research design, an overview of instruments and data collection procedures, and a description of full sample participant characteristics. We then turn to procedures for Propensity Score Matching (PSM) used to establish equivalent treatment and comparison groups and present results from two separate PSM models. The first assesses a range of self-reported academic, social, emotional, and behavioral outcomes using a modified Youth Outcome Survey (YOS). The second looks at school-reported outcomes, including grades, absenteeism, and school disciplinary data. Finally, we discuss overall findings and research limitations.

## **Mentoring and Youth Outcomes**

A quality relationship with a caring adult is an important protective factor in mitigating risk factors and fostering resiliency in children and adolescents (Scales, 2003).<sup>1</sup> While natural mentors—non-parental adults who provide youth with support, encouragement, and guidance—are ideal at fulfilling these needs, at-risk youth often have limited access to natural mentor

---

<sup>1</sup> Linqanti (1992) defines resiliency as “that quality in children who, though exposed to significant stress and adversity in their lives, do not succumb to the school failure, substance abuse, mental health and juvenile delinquency problems they are at greater risk of experiencing” (p. 5).

relationships (Klaw, Rhodes, & Fitzgerald, 2003; Hurd & Zimmerman, 2010). The proliferation of assigned mentoring programs provides an alternative for many young people. Mentoring programs for at-risk youth rely on a developmental prevention strategy designed to improve youth assets such as social competency, problem solving skills, autonomy, sense of identity, and sense of purpose and future (Bernard, 2004; Lerner Brittan & Fay, 2007; Lerner, Napolitano, Boyd, Muller & Callina, 2014). Mentoring is also often framed as a way to mitigate individual, familial, social, and community risk-factors that are associated with poor life outcomes like school dropout and juvenile delinquency (Esbensen, 2000; McCord, Widom, & Crowell, 2001; Shader 2003; Moore, 2006). Effective mentoring can affect multiple aspects of child and youth development, including social-emotional, cognitive, and identity development (Rhodes, 2005; Rhodes et al., 2006), though the quality of the mentor-mentee relationship is thought to be a major moderator of positive youth outcomes (Thomson & Zand, 2009; DuBois, Neville, Parra, et al, 2002). The benefits of mentoring are not restricted to a particular age group or developmental stage but can be observed from early childhood through adolescence (DuBois et al., 2011).

However, most outcome evaluations of youth mentoring have addressed proximal or intermediate outcomes and are limited due to the relatively short study periods, which typically include youth outcomes 12 to 18 months after program enrollment. There are few studies examining the possible long-term effects of youth mentoring. Notable exceptions include two recent studies assess participant outcomes in adulthood, largely related to criminal justice outcomes. Blakeslee and Keller (2018) followed participants from two randomized trials of the *My Life* mentoring program for youth in foster care to determine treatment effects on offending in young adulthood. Adolescents were assessed two years after their participation in *My Life* ended (mean age = 20.4 years old). Findings showed statistically significant differences in criminal justice involvement between treatment and control groups, with untreated youth twice as likely to have been justice system involved, and greater effects for males. Positive effects were noted for serving jail time in the previous year and length of sentence.

In a long-term follow-up with participants from the landmark Public/Private Ventures' (P/PV) 1995 randomized study of BBBS community-based programming, DuBois and Herrera (2018) found that being matched for at least one year was a statistically significant predictor of criminal justice involvement. They also observed statistically significant effects on educational outcomes for sample subgroups, including greater likelihood of post-secondary school attendance among racial and ethnic minorities and those with low parental education, and greater degree completion among younger participants. Additionally, the researchers found retrospective reports of high-quality mentoring relationships "approached or reached statistical significance as a predictor" for a wide range of self-reported outcomes, including "lower likelihood of juvenile arrest, less reported

stealing during adulthood, greater reported grit [particularly for females], emotional, psychological, and social well-being, and self-rated health, and less alcohol use during adulthood” (DuBois & Herrera 2018, p.4-5).

Despite these recent advancements, research remains limited on the long-term impact of youth mentoring. Findings are particularly scarce regarding the persistent effects of mentoring relationships with elementary-age children, a group commonly served by mentoring programs and for whom such relationships could provide cost-effective early prevention.

## **Methods**

This study extends the temporal period for observing mentoring program effects in order to better understand the long-term impacts of high-quality one-to-one youth mentoring. Research was conducted with an affiliate of Big Brothers Big Sisters of America that serves two local school divisions: one a small urban city and the other an adjacent rural/suburban county.<sup>2</sup> We employ a quasi-experimental post-test only design to examine a range of participant outcomes among adolescents (ages 12 to 16) who participated in a Big Brothers Big Sisters (BBBS) mentoring program as children and a comparison group of adolescents, who had enrolled in BBBS but who were never matched with a mentor. Research for this study was approved by a university Institutional Review Board; all human subjects in the study have been treated in accordance with the ethical guideline approved by the Board.

## **Instruments, Measures, and Data Collection**

Mentor and mentee characteristics were gathered from youth enrollment and interview forms. Additional data was gathered from the BBBS Agency Information System. Mentee outcome data were collected using the Youth Outcome Survey as well as school records collected from the two local school divisions. These are described in more detail below.

### **Youth Outcome Survey (YOS)**

The original Youth Outcomes Survey (YOS) includes 32 questions across seven key construct areas (defined by BBBSA). These include Social Acceptance, School Competence, Grades, Future Aspirations, Parental Trust, Peers Risky Behavior, and Truancy. Responses are measured using construct-specific scales. Social Acceptance and School Competence were measured on a 4-point Likert scale ranging from “Not at All True” to “Very True” and Future Aspirations was measured on a 4-point scale ranging from “Not at all Sure” to “Very Sure.” Student-reported grades were reported on a 5-point scale, ranging from “Not Good at All (F)” to “Excellent (A),” and Peers Risky Behavior were measured on a 4-point scale with responses ranging from “It’s

---

<sup>2</sup> The 2017 populations of the city and county were approximately 55,000 and 80,000, respectively.

Not Ok” to “It’s Perfectly OK.” Two constructs were measured on 4-point frequency scales: “Hardly Ever” to “Pretty Often” (Parental Trust) and “Never” to “I did it 3 or more times in the last 30 days” (Truancy). An additional question asked youth whether or not they have a special adult in their lives. A comprehensive list of individual questions and scales is provided in Appendix B: Measures and Constructs. Researchers added six questions related to Peers Risky Behavior (e.g. stealing, bullying, gang membership). We also added a 13-item personal Risky Behaviors scale to measure risky behaviors (e.g. been in a fight) and prosocial behaviors (e.g. volunteered in the community) among mentees. Responses were measured on a four-point Likert scale, ranging from “Never” to “5 or more times” in the last 12 months. A 20-item depressive inventory was adapted from the Center for Epidemiological Studies Depressive Scale (CES-D). The Depressive Scale assesses a mentees’ level of depression using a 4-point Likert scale, ranging from “None of the Time” to “Most of the Time.” Finally, we added a Major Life Factors scale comprised of five items assessing a child’s exposure to stressful life events in the previous six months, including breaking up with a boyfriend or girlfriend, losing friends, or knowing someone who died. These were assessed on a simple “Yes”/”No” scale.

*Survey Administration:* A list of eligible participants in both the treatment and comparison groups were given to area school divisions to determine whether they were still enrolled. Once school enrollment was verified, school counselors were then asked to send home parent permission slips allowing parents to opt their child out of the study. Next, counselors worked with BBBS staff to set a date and time where students could be given the YOS instrument in large groups (typically during break periods). As noted above, surveys administered to youth in the treatment group had to be at least 12 months after their match closed. While we anticipated surveys would take 15-20 minutes, the average completion time was 10.68 minutes (7.46) and the maximum completion time was 30 minutes. Most participants were given a coupon after survey completion worth around \$5 for use at a local restaurant.

## **School Records**

Researchers facilitated the development of data sharing agreements between BBBSHR and the two school divisions in its service region in order to collect data on a range of baseline mentee characteristics. These included student grades and absences, grade retention, Special Education and Limited English Proficiency designations, and disciplinary infractions (see Appendix B: Measures and Constructs for a complete list of school data collected for the project). Data was compiled by school division data managers and provided to the research team via a secure server. Variables used as outcome measures are described below.

*Grades:* In earlier research we found that youth are more reliable reporters of overall GPA than of individual grades (Peaslee & Teye, 2015). Therefore, we calculated a core subjects GPA,

using Math, Reading/Language arts (here, we used reading when available, but if not used Language Arts at the recommendation of the school division), English, Science, and Social Studies. Although the difficulty level of these subjects might vary, we did not have a way of capturing this. We report both quarterly core subjects GPA as well as an average across quarterly grades.

*Low or Failing Grades:* We used two variables to capture unsatisfactory grades. First, we identify whether participants received an F in any core subject. Second, we looked at whether they had a failing or unsatisfactory grade (F or D) in any core subjects and also report the frequency.

*Absenteeism:* We use two separate measures of absenteeism. First, we sum the number of total absences per year. Second, we define chronic absenteeism as more than six absences per quarter, or 18 or more missed school days in a year.

*Discipline:* Disciplinary data includes whether or not students received a formal infraction that year, the number of disciplinary incidents, and whether or not they received in-school or out-of-school suspension (yes/no and frequency).

## Participant Selection

Treated youth in the adolescent study were originally identified from a sample of all newly matched youth in two BBBS programs (one school-based and one community-based) from February 2012 to November 2013 (n = 480). Among these, the following eligibility requirements were set for participation in the long-term adolescent study: 1) Participant must have been matched during childhood (ages 6-11 years old); 2) Child must have had a successful match relationship (defined as a minimum of 12 months);<sup>3</sup> 3) The match must have closed at least 12 months prior to participation in the adolescent study; and 4) Participants should be between 12 and 16 years of old. The average age of the initial enrolled population during this period was 8.7 years old (most were under age 10).

Overall, 130 treated youth were surveyed during the five-year study period. Among those in the treatment group, 67 had been enrolled in the community-based program. However, ten of these had matches that had lasted less than 12 months, two had been served in adolescence (past age 12), and six youth were surveyed within 12 months of match closure. An additional six were repeat assessments. These cases were all dropped from further analysis, resulting in a final sample population of 43 community-based subjects. Additionally, 63 treated youth had participated in school-based matches. However, five were in matches that did not last at least 12 months and

---

<sup>3</sup> Four cases were re-matched. The initial case did not make it to 12 months, for this study we are using the longer match.

one was surveyed prior to adolescence. Two matches did not meet follow-up criteria (i.e., the survey given within 12 months of match closure), and six had been previously assessed. Therefore, the final sample population of school-based treatment subjects was 50.

Members of the comparison group were first identified from historical agency waitlists of youth who had never been served and who would be in the target age range during the study period. During the first round of inquiry, we used cases that had been removed from the waitlist (either because the child aged out of the program's service range or because the family had lost contact with the agency) from 2007 through 2011. That process yielded 133 eligible youth. In 2017, we expanded the criteria to include those who had been on the waiting list at least 18 months. We revisited the waitlist to generate an additional 64 eligible participants based on previously stated criteria. Ultimately, 106 unmatched youth reached age 12 during the study period and were, therefore, eligible to participate.<sup>4</sup> BBBS staff worked with the two local school divisions to identify targeted youth who were still enrolled locally and send home parent permissions allowing them to opt-out of the study. Ultimately, 68 surveys were administered to youth who were never matched in the BBBS program. Since our aim was to reach only youth ages 12-16, we dropped nine cases of youth surveyed at age 17 and 18. Additionally, one respondent had been matched earlier under a different last name and one had been previously assessed. The final sample population of comparison subjects was 57.

Table 1 shows the distribution of respondent age at enrollment and at the time of the YOS administration. Although the average age was nine for all three groups, age distribution was statistically different across groups ( $F = 7.37(2)$ ;  $p = .001$ ). The difference, here, was largely driven by the inclusion of youth who were adolescents at the time of enrollment in the comparison group (i.e., those added in 2017). Twenty youth were age twelve or older during the time of referral and enrollment in the program. The oldest was 13 years old. By contrast, the oldest child during referral and enrollment in the treatment group was 11 years old. Age at YOS was also significantly different across groups ( $F=19.82(2)$ ;  $p = .000$ ). The comparison group was significantly older than the community-based and school-based treatment groups. As noted, the comparison group included sixteen year-olds ( $n = 12$ ) while treatment did not. Because we drew from members of the waitlist back to 2007, several eligible comparison youth were already sixteen at the start of the study period.

---

<sup>4</sup> It should be noted, however, that because we expanded eligibility for the comparison group in 2017, a fundamental assumption in comparability of youth was compromised. That is, our initial parameters assumed that youth in the study population were all referred in childhood (ages 6-11) and assessed in adolescence. However, in order to increase sample size, in our second round of participant identification we allowed some who had been referred to the program after age 12 to be part of the comparison group. We will discuss the consequences of this later in the analysis.

Table 1. Age, Contact, and Match Length Distribution							
Match Type	N	Age at Enrollment <i>M(SD)</i>	Age at YOS <i>M(SD)</i>	Age at YOS Range	Match Length	Time since match closure	Time since enrollment
<b>T(1)</b> <sub>community</sub>	43	9.16 (.76)	13.59 (.75)	12.4-15.1	23.05 (7.84)	30.42 (7.86)	53.7 (9.59)
<b>T(2)</b> <sub>School</sub>	50	9.08 (.68)	13.29 (.79)	12.1-15.6	20.76 (6.22)	30.23 (7.65)	51.2 (6.97)
<b>Comparison</b>	57	9.94 (2.14)	14.55 (1.33)	12-16.6	----	----	56.01 (25.47)

Among those in the treatment, the average age was nine years old, one-third were male, nearly half were Hispanic, 57 percent were from households receiving public assistance, and about half were living in the city. Within the comparison group, the average age was nine (although significantly older than the treated group), nearly two-thirds were male, slightly over half were Hispanic, nearly-two-thirds were from households receiving public assistance, and 57.9 percent were living in the city. In summary, the comparison group reflected characteristics the agency had discussed with us as being “difficult to match” in the area, slightly older, Hispanic, males.

Importantly, the average time since match closure was nearly three years for both treatment groups, and close to five years since initial program enrollment. This period is significantly longer than most current outcome studies.

## Part I: Youth Reported Outcomes

We explore program effects on youth reported outcomes based on long-term follow-up assessment with the YOS. In this section we first describe our selection of the matched comparison group and then present program effects on educational expectations, risky behaviors and protective factors, and depressive symptoms. In Part II, we analyze program impacts on school-reported grades, absenteeism and disciplinary records.

### Selection of the Matched Comparison Groups

Data gathered during in initial program application and, in some cases, youth enrollment were used to establish a matched comparison group using Propensity Score Matching. These data included individual-level child demographics, household risk factors, and environmental risk factors. Child demographics included age at enrollment, gender, and ethnicity/race (African American, Hispanic, white, other). Household risk factors included family living situation (this term was initially conceptualized as living in the home with one parent, two parents, two married parents, or another related adult. For this study, we recoded 1 as married), household receipt of public assistance, household income (coded for this study as 1= below \$20,000), free/reduced



price lunch status, maternal education level (recoded as binary variable, 1=less than a high school degree), parental unemployment, and parental incarceration.

Environmental risk factors included child school district and school poverty level. Child school district, in this case, represents a critical environmental dichotomy. The city has substantial ethnic diversity (this is driven largely by the refugee resettlement status and local agricultural community), poverty levels higher than the state and national averages, and increased crime rates (as compared to other rural Virginia localities). However, the city is also home to two major universities and a thriving local economy. The county region, by contrast, has a typically rural demographic composition. It is over 85% white with pockets of very high socio-economic status neighborhoods and several impoverished agricultural communities. We included a percentage of free/reduced lunch receipts PS model (controlling for age at enrollment to account for differences noted), and school district as a proxy for neighborhood risk. School poverty level was conceptualized as the rate of students receiving free/reduced price lunch at the school. This was recoded as binary variable (1 = > 50%).

Variables influencing treatment provision were similar for both the treatment and comparison groups. Youth were referred to the program during childhood (we discussed exceptions above), their parents followed-through with the application process, and the child was placed on the agency waitlist for matching. These factors are critically important in matching unobserved child and family characteristics across groups. For example, youth in the comparison group were not selected directly through schools or other community agencies.

Propensity score analysis and matching were performed using the binary logistic regression command in SPSS. After ensuring all test assumptions were met, following logistic regression model was specified:

$$\text{logit}(\pi_1) = \beta_0 + \beta_{(\text{child characteristics})} + \beta_{(\text{household risk factors})} + \beta_{(\text{environmental risk factors})} + \varepsilon$$

The specific matching method used was nearest neighbor 1:1, no replacement in either group (BBBS or Comparison). A caliper of .2 was set, which was calculated using the pooled standard deviation of logit of the propensity score. A total of 40 of the original 58 comparison group participants and 40 treated participants were identified through this process. Eighteen matches in the control group were discarded due to lack of overlap with propensity scores in the treatment group. The two groups were well matched on all of the measured characteristics as shown in Table 31 ( $p > .2$  was used as a threshold for assessing significant group differences among the

matched pairs). A full matching model, using a 1:2 ratio where possible is also displayed in Table 2. Although the groups were well-matched (using  $p < .2$  threshold), the model increased the number of participants in the treatment group to 62 and 41 in the comparison group, we are not utilizing the full matching model for analysis.

<b>Table 2. Participant Characteristics (Original and Matched Groups)</b>						
<b>Covariates</b>	<b>Original</b>			<b>Nearest Neighbor Match</b>		
	<b>Treatment n=93</b>	<b>Control n=57</b>	<b>% Bias</b>	<b>Treatment n=40</b>	<b>Control n=40</b>	<b>% Bias</b>
<b>Age at Program Enrollment</b>	9.12 (.72)	9.94 (2.15)	9%	9.39 (.69)	9.47 (2.13)	.85%
<b>Gender (Male)</b>	33.3%	59.6%***	26.3%	42.5%	50%	7.5%
<b>Ethnicity (Hispanic)</b>	48.4%	63.2%*	14.4%	62.5%	60%	2.5%
<b>School District (City)</b>	52.7%	57.9%	5.7%	62.5%	67.5%	5%
<b>Free/Reduced Lunch Status</b>	83.9%	68.4%**	15.5%	72.5%	72.5%	0
<b>Receive Public Assistance</b>	57%	61.4%	4.4%	62.5%	57.5%	5%
<b>Family Income (&lt; \$20,000)</b>	37.1%	37.9%	.8%	42.5%	42.5%	0%
<b>Low Maternal Education Level</b>	16.1%	15.8%	.3%	20%	17.1%	2.9%
<b>Family Living Situation (Two Parent)</b>	47.3%	38.6%	8.7%	52.5%	45%	7.5%
<b>Primary Unemployed</b>	11.8%	12.3%	.5%	12.5%	12.5%	0%
<b>Primary Incarcerated</b>	5.4%	7%	1.6%	12.2%	5%	7.2%
<b>School Free-Reduced Lunch Rate (above 50%)</b>	52.7%	49.1%	3.6%	62.5%	52.5%	10%

\*  $p < .1$ ; \*\* $p < .05$ ; \*\*\* $p < .01$ ; \*\*\*\* $p < .001$  for unmatched pairs  
Using \* $p > .2$  as the threshold for group differences within the matched pairs

## Results

### Educational Expectations

Results suggest that adolescent respondents were generally very positive about their educational future, as educational expectation items were highly endorsed by both groups. For example, none of the sample noted they were ‘Not at all sure’ about finishing high school, only 1.3 percent were unsure about going to college, and just 2.5 percent were fully unsure about finishing college. Therefore, here we modeled the treatment effect as being moderately confident or ‘Mostly sure’ or ‘Not really sure’ using ‘Very sure’ as our reference group. To show positive effect sizes we recoded treatment as zero. Results indicate that treatment had no effect on adolescent confidence in finishing high school. Groups were nearly equivalent in reporting. However, adolescents in the treatment were far more likely to be highly confident that they would go to college (2.5 times more likely to respond ‘very sure’ than ‘mostly sure’). Those in the treatment group were also 3.4 times more likely to report they were ‘very sure’ versus ‘mostly sure’ they would finish college than those in the comparison group (see Table 3).

TABLE 3. Treatment Effect on Adolescent Educational Expectations					
Mentor Characteristic	$\chi^2(df)^*$	% of Treatment	% Comparison	Odds Ratios	
				Estimate	95% CI
Finish High School					
Mostly sure	.00(1)	15%	15%	1.00	(.29 – 3.41)
Not really sure	1.01(1)	2.5%	0%	---	---
Go to College					
Mostly sure	3.16(1)*	35%	17.5%	2.54*	(.89 – 7.2)
Not really sure	.09(1)	15%	17.5%	.83	(.25 – 2.73)
Finish College					
Mostly Sure	5.23(1)**	37.5%	15%	3.4**	(1.15 – 9.99)
Not really Sure	2.05(1)	12.5%	25%	.429	(.13 - 1.39)
†Odd Ratios > 1.5; ††Odds Ratio >3.5 *p<.1; **p<.05; ***p<.01; ****p<.001					

### Risky Behaviors and Protective Factors

For most of the reported risky behaviors, less than 20% of the population selected responses other than ‘Never.’ Therefore, we recoded (Never = 0) and report the effect of treatment on the likelihood to engage in behaviors at any frequency. There were several risk indicators where respondents indicated higher frequency exposure (> 20% selecting a response other than never). For these, we also tested a binary variable for higher frequency engagement (1= 5 or more times). Findings here indicate that treatment had a positive weak effect on likelihood of using tobacco, gang participation, higher frequencies of wanting to fight, and higher frequencies of fighting. Treatment had a positive moderate effect on likelihood to consume alcohol, get a failing

grade more than 5 times, or participate in an organized activity. The only area in which we found negative treatment effects was for high rates of volunteering.

Because the frequency of report for most risky behaviors was very low, we were unable to model the effect of other key covariates such as gender, age, city versus county residency, and socio-economic status. For areas of moderate impact, we specified a logistic regression model incorporating these variables. Results for participation in organized activities indicated that treatment ( $\chi^2 = 7.47(1)$ ;  $p = .006$ ;  $\text{Exp}(B) = 4.65$ ) and child school district ( $\chi^2 = 3.797(1)$ ;  $p = .05$ ;  $\text{Exp}(B) = 6.68$ ) alone were significant predictor of participation in organized activities (pseudo- $r^2 = .21$ ). When testing for reporting high frequency of failing grades, treatment was only marginally significant (which we could have anticipated given the low number reporting) but had a notably high Odds ( $\chi^2 = 2.22(1)$ ;  $p = .13$ ;  $\text{Exp}(B) = 5.47$ ). Other covariates were not significant (pseudo- $r^2 = .14$ ). In both cases, we can confirm that treatment had an important effect on youth reported engagement in risky behaviors.

TABLE 4. Treatment Effect on Adolescent Engagement in Risky Behaviors					
Mentor Characteristic	$\chi^2(df)^*$	% of Treatment	% Comparison	Odds Ratios	
				Estimate	95% CI
<b>Risky Behaviors</b>					
Been Arrested	.213(1)	5%	7.5%	1.5	(.24 – 9.75)
Smoking Cigarettes/Using Tobacco	1.05(1)	2.5%	7.5%	3.162†	(.315 – 31.77)
Drinking Alcohol	3.13(1)*	5%	17.5%	4.03*	(.78 – 20.76)
Using Drugs	.24(1)	7.7%	5%	.632	(.1 – 4)
Participating in Gang Activity	.213(1)	5%	7.5%	1.54†	(.24 – 9.75)
Wanting to Hurt Yourself	.34(1)	17.5%	12.8%	.693	(.2 – 2.4)
Bullied Others	.06(1)	82.5%	84.6%	.86	(.26 – 2.83)
<b>Bullied Others (High Frequency)</b>	1.039(1)	0%	2.6%	----	
Been Bullied	.07(1)	25%	27.5%	1.13	(.42 – 3.08)
<b>Been Bullied (High Frequency)</b>	.346(1)	5%	2.5%	.49	(.04 – 5.59)
Wanting to Fight	.15(1)	37.5%	33.3%	.83	(.33 – 2.09)
<b>Wanting to Fight (High Frequency)</b>	5.47(1)	7.5%	15.4%	2.24†	(.52 – 9.69)
Fighting	.95 (1)	35%	25%	.62	(.23 – 1.63)
<b>Fighting (High Frequency)</b>	.35(1)	2.5%	5%	2.05†	(.18 – 23.59)
Having a Failing Grade	.47(1)	57.5%	65%	1.37	(.56 – 3.86)
<b>Failing Grade (High Frequency)</b>	2.778(1)*	2.6%	12.5%	5.43††	(.6 – 48.78)
†Odd Ratios > 1.5; ††Odds Ratio >3.5					
*p<.1; **p<.05; ***p<.01; ****p<.001					

TABLE 5. Treatment Effect on Adolescent Engagement in Prosocial Activity					
Mentor Characteristic	$\chi^2(df)^*$	% of Treatment	% Comparison	Odds Ratios	
				Estimate	95% CI
<i>Prosocial Behaviors</i>					
Volunteering	.31(1)	51.3%	45%	.77	(.32 – 1.88)
<b>Volunteering (High Frequency)</b>	.32(1)	2.6%	5%	2†	(.17 – 22.9)
Participating in Organized Activities	6.4(1)**	77.5%	50%	.29**	(.11 - .77)
<b>Participating in Organized Activities (High Frequency)</b>	.69(1)	35%	26.3%	.66	(.25 – 1.75)
†Odd Ratios > 1.5; ††Odds Ratio >3.5 *p<.1; **p<.05; ***p<.01; ****p<.001					

### Depressive Symptoms

When variables had abnormal distribution ( $|skewness|>2$  or  $|kurtosis|>2$ ), we removed outliers at 95% confidence. We detected two outliers in the treatment group. In addition, one member of the treatment group did not complete any of the depression items on the survey, but did complete at least 50% of the survey. Results from an independent sample t-test confirm a positive treatment effect on depressive symptoms (see Table 6). The comparison group reported depressive symptoms that were three points higher than the treatment group. Alone, treatment effect had a moderate effect on depressive symptoms.

Table 6. Mean Youth-Reported Depressive Symptoms			
	Mean (SD)	T (p)	Cohen's d
<b>Treatment</b>	10.43 (6.1)	-1.69 (.095)*	.38
<b>Comparison</b>	13.52 (9.6)		
p<.1; **p<.05; ***p<.01; ****p<.001			

We then employed a OLS model to predict depressive symptoms using treatment, gender, ethnicity, age, treatment\*age at time of survey administration, family socio-economic status, city versus county residence, parental incarceration, living situation, maternal education, receipt of public assistance, family income, and school poverty rate (at 50% or higher). Table 7 displays results from the model, removing insignificant covariates ( $d<.15$ ).

Table 7 Results from OLS Analysis Predictive Depressive Symptoms		
	B	Std. error
Intercept	4.26	5.57
Child Gender (Male)	3.24*	1.82
Ethnicity		
Black	-2.18	3.13
Hispanic	-5.01*	2.99
School District (City)	4.811**	2.18
Low Maternal Education	3.04	2.4
Family Income less than \$20,000	-6.89**	2.85
Receiving Family Assistance	4.22**	1.86
Income x Ethnicity	4.86	3.58
Treatment (Control)	-3.47**	1.72
r <sup>2</sup> =.31		
F=2.94(11); p=.004		
n=77		
*p<.1; **p<.05; ***p<.01; ****p<.001		

## Discussion

This study identified 150 adolescents, ages 12 to 16, to understand the impact of mentoring on long-term youth reported outcomes. 93 youth had been in a successful mentoring relationship (at least 12 months long) while 57 had been enrolled by a parent for services but were never given a mentor. We followed up with both groups nearly five years, on average, after initial program enrollment and administered the Youth Outcome Survey, enhanced with several additional measures. As anticipated, our original groups were un-equivalent across several key factors, including demographics (many of which we expect are related to selection into the treatment), youth individual characteristics, and environmental factors (both are known predictors of our outcomes). Thus, prior to exploring group differences across self-reported outcomes, we employed a propensity score matching procedure which lead to 40 well-matched pairs (n = 80). While basic, our analysis yielded important results, demonstrating that the BBBS program has a long-lasting impact on youth served in childhood.

Findings indicated that treated youth had greater confidence or ‘surety’ in their educational expectation. While groups did not have differing expectations related to high school graduation, treated youth were statistically more likely to be sure they would go to college and much more likely they would finish college. In addition, we might hope that higher expectations are driving interim outcomes like better academic performance, reduced absenteeism and higher school affect.

While reported engagement in risky behaviors was low for all subjects, results indicated important differences. Youth with successful mentoring relationships were more likely to engage in key pro-social activities, like participating in organized activities. They were also less likely to report smoking, drinking alcohol, and participating in gang activities. Moreover, there were a few outcomes that we found were more likely to occur at higher frequencies in the comparison group. For example, whereas both groups reported fighting, at least one time, at roughly the same rate, the comparison group was much more likely to report fighting in higher frequencies (five or more times). This was also the case with wanting to get in a fight with someone, and having failing grades in school. As we demonstrate in the next section of this report, results related to self-reported frequency of failing were validated in an analysis of actual school records. Finally, results indicated the program had an important effect on youth depressive symptoms. Employing a multi-variate analysis, we found that treatment was a significant predictor of youth depression, members of the treated group indicated lower levels of depressive symptomatology. Results also showed that gender, ethnicity, rural residence, and socio-economic status were also significant predictors.

Importantly, each of the outcomes shown here to be associated with youth mentoring can be directly or indirectly associated with future likelihood for involvement with the juvenile justice system. We anticipate that even modest levels of program effect, as detected here, may have profound influence on continuing youth life choices during their last few years of adolescence and into early adulthood.

## **Part II: School Reported Outcomes**

This section discussed the selection of a second matched comparison group and reports findings related to program impact on school-reported youth outcomes. These include quarterly grades, absenteeism, and school disciplinary records (including in-school and out-of-school suspension records). As discussed previously, historical and ongoing school data were solicited from two local school divisions, per MOUs with school superintendents, for all served youth and youth identified in the comparison group.

### **Selection of Matched Comparison Groups**

As noted, we initially had 150 subjects in the adolescent study (57 in the comparison group and 93 in the treated group). Among these, baseline school data (data on grades and absenteeism at the point of program entry) were available on only 75 (80.7%) of the treated youth and none of the youth in the comparison group. Given, the average length of time from enrollment to assessment in the adolescent study was nearly five years, we did not feel that pursuing missing data for the comparison group would have been a useful process. Grades in early childhood

would only have been useful for the purposes of propensity score matching and may not be predictive of adolescent grades. Rather, for propensity score matching, we focused on other risk factors that would have been more permanent in the child’s life. Thus, the study design explores follow-up grades collected during the year of assessment. In practice, this meant documenting the year and school location that the youth was assessed with the YOS and submitting a unique data requests from each school division. We collected four quarterly observations of grades, four quarterly observations of absenteeism, and annual school discipline records marking the number of formal disciplinary incidences, referral to in-school suspension and referral to out-of-school suspension.

Despite multiple follow-up requests, throughout the grant period we encountered a number of problems with consistency and completeness of school records, particularly from the City school division. For example, where data were incomplete, school officials noted they youth did not take every subject at every quarter and, in some cases, youth did not take core subjects at all. Where data were missing altogether for youth, we were informed that they did not have adequate records or could not confirm that the child was at the school at the time we requested data. Table 8 below shows data completion rates for school grades, absenteeism, and disciplinary infractions in each group. Data were considered missing entirely for absences if we had no record of absences across all four quarters. For grades, we considered data missing in a given quarterly observation if we had less than two observations of core subject grades (Reading, Language Arts, Math, Science, Social Studies). We considered data missing entirely if we had no core subjects grade point average observations across four quarters. Overall, across both groups, ten percent of school reported grade data were missing across both groups. Twenty percent of absentee data and 14.67 percent of disciplinary records were missing. While overall data completion rates were acceptable (80% or higher), we had observably lower data completion in the comparison group.

Table 8. School Data Completion Rates						
	School Reported Grades		School Reported Absenteeism		School Reported Disciplinary Records	
<b>Treatment</b>	87/93	93.5%	78/93	83.8%	90/93	96.7%
<b>Comparison</b>	48/57	84.2%	43/57	73.7%	38/57	66.7%

Propensity score analysis and matching were performed using the binary logistic regression command in SPSS. We employed the model and methodological approach used to assess youth reported outcomes, a nearest neighbor 1:1, with no replacement in either group (BBBS or Comparison). In this case, a caliper of .25 was set. Despite the higher caliper,  $p > .25$  was used as a threshold for assessing significant group differences among the matched pairs. Prior to



executing propensity score matching, we screened data for school record completion. Specifically, we dropped cases missing grades (Treatment n = 87; Comparison n = 48). Table 9, presented here, shows summary statistics and percent bias across groups prior to performing matching and after matched pairs were identified. Further analysis of school reported grades, absenteeism and discipline are conducted with the matched groups from this procedure (missing data are reported for absenteeism and discipline).

<b>Table 9. Participant Characteristics (Original and Matched Groups)</b>						
<b>Covariates</b>	<b>Original Grade Database</b>			<b>Nearest Neighbor Match</b>		
	<b>Treatment n=87</b>	<b>Control n=48</b>	<b>% Bias</b>	<b>Treatment n=35</b>	<b>Control n=35</b>	<b>% Bias</b>
<b>Age at Program Enrollment</b>	9.11 (.74)	10.05*** (2.17)	9.4%	9.35 (.75)	9.58 (2.12)	.85%
<b>Gender (Male)</b>	33.3%	60.4%***	27.1%	45.7%	54.3%	8.7%
<b>Ethnicity (Hispanic)</b>	46%	58.3%	12.3%	57.1%	57.1%	0%
<b>School District (City)</b>	52.9%	56.3%	3.4%	62.9%	62.9%	0%
<b>Free/Reduced Lunch Status</b>	83.9%	64.6%**	19.3%	74.3%	74.3%	0%
<b>Receive Public Assistance</b>	56.3%	56.3%	0%	51.4%	57.1%	5.7%
<b>Family Income (&lt; \$20,000)</b>	35.6%	39.6%	3%	42.9%	45.7%	2.8%
<b>Low Maternal Education Level</b>	16.1%	16.7%	.6%	14.3%	11.4%	2.9%
<b>Family Living Situation (Single Parent)</b>	48.3%	33.3%*	15%	51.4%	54.3%	2.9%
<b>Primary Unemployed</b>	10.3%	12.5%	2.2%	14.3%	14.3%	0%
<b>Primary Incarcerated</b>	4.6%	8.3%	3.7%	8.6%	5.7%	2.9% %
<b>School Free-Reduced Lunch Rate (above 50%)</b>	52.9%	47.9%	5%	54.3%	54.3%	0%

\* p<.1; \*\*p<.05; \*\*\*p<.01; \*\*\*\*p<.001 for unmatched pairs  
Using \*p>.2 as the threshold for group differences within the matched pairs

A total of 35 of the original 48 comparison group participants with grade records and 35 treated participants were identified through this process. Thirteen matches in the control group were discarded due to lack of overlap with propensity scores in the treatment group. The two groups were well matched on all of the measured characteristics as shown in Table 9 ( $p > .25$  was used as a threshold for assessing significant group differences among the matched pairs). Among the 35 treated youth, 13 had been successfully matched in the community-based program and 22 in the school-based program. The average match length among treated youth was 21.46 months (SD = 6.86; Range = 12.23 to 43.52). In the treated group, the average follow-up time from match closure to assessment was 30.57 months (SD = 14.2; Range = 14.23 to 48.9). For both groups, the average time between initial program enrollment and assessment was 54.75 months

(SD = 18.74; Range = 10.26 to 93.33). The most common time lapse from enrollment was 49.3 months (4 years).

Although we matched on baseline characteristics (including age), we noted lack of age at the time of survey administration in the propensity score model as a possible study limitation. While baseline age was included to ensure that all children would have been matched and served during childhood (a key eligibility criteria), age at survey administration was likely to have been an important predictor of academic success, absenteeism, and school behavioral issues. For example, we anticipate that older children are more likely to be engaging in risky behavior including school level risk factors like failing courses, chronic absenteeism and incurring formal behavioral infractions. Within the final matched sample, the average age of treated youth was 13.65 (SD = .8; Range = 12.35 to 15.59). The average age in the comparison group at the time of assessment was 14.28 (SD = 1.2; Range 12.48 to 16.5) ( $t(68)=2.59$ ;  $p=.01$ ). Although a consideration, both groups fell within the stated age range (12 to 16 years old). The difference amounts to half a school year.

## Results

### Grades

Table 10 below presents basic D.O.M tests of youth core subjects GPA by treatment condition for the full matched sample ( $n=70$ ). We observed, moderate effect sizes at all observed time periods. Importantly, we observed a moderate treatment effect on cumulative core subjects GPA. Youth in the treatment group had a notably higher GPA (Mean = 2.55; SD = 1.04) compared to those in the comparison group (Mean = 2.25; SD = .9).<sup>5</sup> In addition, we tested for conditional effects of gender, ethnicity, city residence, and family receipt of public assistance programs. We did not observe any conditional effects for gender, public assistance, or city residence. We did, however, detect an important conditional effect of ethnicity.

Table 10. Program Effect on School Reported Grades				
Core Subjects GPA		Mean (SD)	T (p)	Cohen's d
1 <sup>st</sup> Quarter	Treatment	2.61 (1.14)	1.12(.27)	.27
	Comparison	2.32 (.97)		
2 <sup>nd</sup> Quarter	Treatment	2.3 (1.21)	1.3(.18)	.32
	Comparison	1.91 (1.18)		
3 <sup>rd</sup> Quarter	Treatment	2.54 (1.07)	1.76(.08)*	.41

<sup>5</sup> Findings were similar in the static group comparison ( $n = 135$ ). The treatment group had significantly higher average core subject grades (Mean = 2.76(.89)) than the comparison group (Mean = 2.29 (.94)) ( $t(133) = 2.85$ ;  $p = .05$ ).

	Comparison	2.09 (1.09)		
<b>4<sup>th</sup> Quarter</b>	Treatment	2.52 (1.16)	1.21(.23)	.29
	Comparison	2.2(1.04)		
<b>Cumulative</b>	Treatment	2.55 (1.04)	1.29 (.2)	.31
	Comparison	2.25 (.909)		
p<.1; **p<.05; ***p<.01; ****p<.001				

While Hispanic students in the treatment group performed only modestly better, Black and white students in the treatment group had observably higher grades than in the comparison group. For example, the average cumulative GPA for white students in the comparison group (n = 9) was 1.95 (SD = .77) while it was 2.58 (SD = .99) in the treatment group (n = 10) (Cohen’s d = .72). Similarly, average grades for Black students in the comparison group (n=6) were 1.66 (SD = .95) versus an average of 2.47 (SD = .98) in the treatment group (n = 6) (Cohen’s d = .84). These results suggest that a successful mentoring relationship, especially for white and Black students, is a major factor in adolescent academic performance (see Figure 1). Large effect sizes presented here, however, do not account for the low number of observations for each sub-group, which draws into question the broader generalizability of these findings.

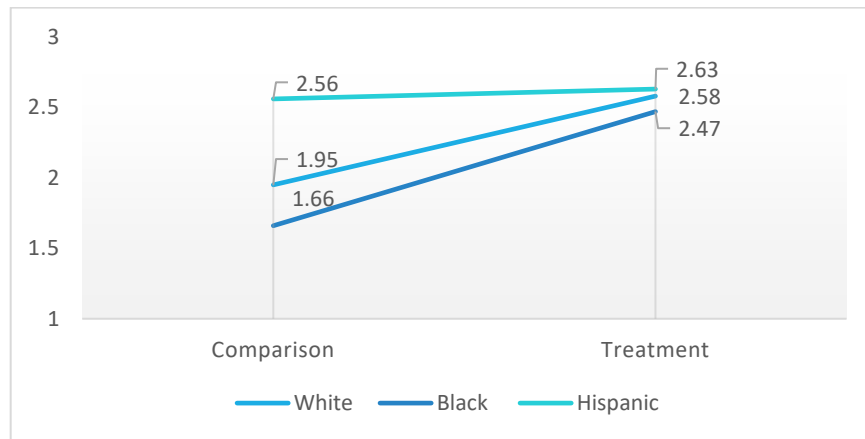


Figure 1: Program Effects on Cumulative GPA by Ethnicity

Second, we tested treatment effect on the likelihood to get a failing or unsatisfactory grade (D or F). Table 40 below presents major findings for the full matched sample. Members of the comparison group were more likely to have a failing grade than those in the treatment (noting small effect sizes). Importantly, members of the comparison group were much more likely to have a failing or unsatisfactory grade in the core subject area than those in the treatment (Odds = 2.53).

In addition, frequency of receiving a failing or unsatisfactory grade was higher for comparison subjects than for treated. On average, youth with successful BBBS matches received an unsatisfactory grade 2.89 (SD = 3.51; Range = 0 to 13) times while youth with no BBBS mentor had, on average, 3.89 (SD = 3.7; Range = 0 to 14) times. While not a statistically significant difference ( $t(68) = -1.16$ ;  $p = .25$ ), we did observe a small treatment effect (Cohens  $d = .277$ ). This is an important confirmation of previous findings suggesting youth in the comparison group were more likely to both fail and receive unsatisfactory grades than treated youth. Here, we can further conclude that they may also be marginally more likely to have a higher frequency of failing/unsatisfactory grades.

Table 11. Likelihood of Receiving a Failing or Unsatisfactory Grade by Treatment Group				
	$\chi^2(df)^*$	%	Odds Ratios	
			Estimate	95% CI
<b>Failing Grade</b>				
Treatment	1.45(1)	37.1%		
Control		51.4%	1.79†	(.69 – 4.65)
<b>Failing or Unsatisfactory Grade</b>				
Treatment	.045(1)*	57.1%	2.53*	(.89 – 7.12)
Control		77.1%		

\* $p < .1$ ; \*\* $p < .05$ ; \*\*\* $p < .01$ ; \*\*\*\* $p < .001$

### Absenteeism

We also explored program impact on student absenteeism focusing on two key measures, cumulative absences and high absenteeism (see description of measures). For this analysis we used the matched pairs previously identified through propensity score matching ( $n=70$ ). However, absence data were not available on all subjects (Treatment  $n = 30$ ; Comparison  $n = 28$ ). Again, we are aware that the limited sample size presents a limitation in generalizability of findings. Results from a t-test indicated the treatment group had significantly lower levels of absenteeism ( $t(55) = 2.39$ ;  $p = .02$ ;  $d = .63$ ). Youth in successful match relationships had an average of 10.43 (SD = 9.08; Range = 0 to 39) annually while the comparison group had an average of 16.89 (SD = 11.23; Range = 2 to 98)

**Table 12. School Disciplinary Infraction**

Classroom/Campus disruption
Defiance/ Refuses Request
Fighting No or Minor Injury
Altercation/Confrontation No Injury
Assault Battery No weapon
Disruptive demonstrations
Sexual harassment
Theft School Property
Bringing Tobacco Paraphernalia to School
Bullying

absences annually.<sup>6</sup> One student in the comparison group with 98 absences, however, was dropped from the sample for analysis of average scores and group differences as an outlier ( $Z = 5.44$ ). Again, we tested for conditional effects of gender, ethnicity, city residence, and socio-economic status. Findings indicated that the effect of treatment may have been conditioned upon gender. Whereas boys in the treatment group had only a modest decline in absences, rates of absenteeism for girls were vastly lowered in the treatment group ( $F(1) = 3.88$ ;  $p = .054$ ) (See Figure 2).

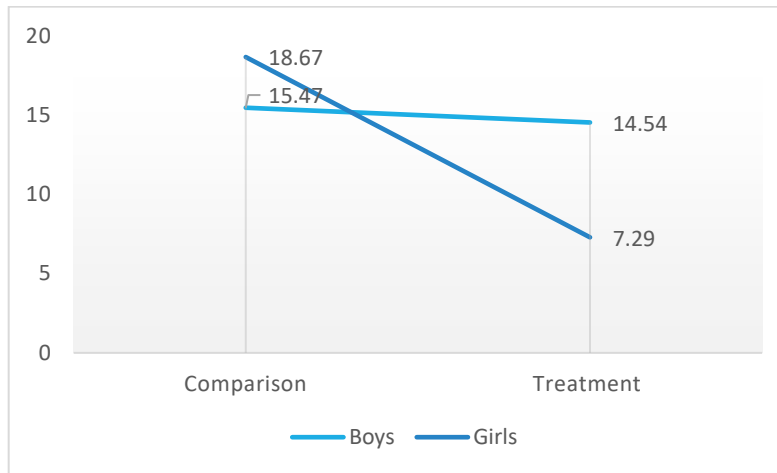


Figure 2: Program Effect on Number Annual Absences by Gender

Second, we explored the program effect on the likelihood of having chronic absenteeism (greater than 18 absences a year), a known risk factor. Although not statistically significant, the comparison group was more likely to have chronic absenteeism (42.9%) than the treatment group (31.4%). Fifteen students in the comparison group had more than 18 annual absences, compared to 11 in the treatment. Mantel-Haenszel Odds suggest a small program effect (Odds = 1.636). Again, although the number of observations was severely limited, we observed a much more pronounced program effect on girls than boys. 43.8% of girls in the comparison group had chronic absenteeism versus 21.1% in the treatment group ( $\chi^2(1) = 2.07$ ;  $p = .14$ ). An Odds Ratio of 2.92 suggests a moderate program effect for girls.

### School Discipline

As noted, we were missing school disciplinary records on 20% of the adolescent study population. Thus, our sample for school discipline analysis was 119 cases (35 in the Comparison

<sup>6</sup> Findings were similar in the static group comparison ( $n = 135$ ). The treatment group had significantly lower absences (Mean = 13.47(11.3)) than the comparison group (Mean = 18.95 (18.27)) ( $t(117) = 2.02$ ;  $p = .046$ ).

and 84 in the Treatment).<sup>7</sup> Among these, 33 had a formal disciplinary infraction during the school year. Fifteen students were given In-School-Suspension (ISS). Among these, eight were referred on more than one occasion (Range = 2 to 12). Thirteen students were referred to Out-of-School Suspension (OSS). Seven had multiple referrals (Range = 2 to 5). Noted offenses were provided (see Table 12). Most common among those noted was classroom disruption.

Once again, we filtered the data using the equivalent pairs identified in the propensity score matching process. Among the 70 cases, 33.3% in the treatment group (11) had a formal infraction, compared to 44% in the comparison group (11) ( $\chi^2(1) = .69$ ;  $p = .29$ ). We found only a small program effect (Odds = 1.57). We found a stronger effect for boys than girls (Odds = 2.13). Among boys, 38.5% treatment group versus 57.1% in the comparison had a formal infraction. 17.1% of those in the treatment and 11.4% in the comparison were referred to ISS, findings indicated no difference and no effect. Similarly, we detected no statistical difference in referral rates to OSS, which were 44.4% in the Treatment and 55.6% in the comparison group.

## **Discussion**

We explored the impact of the BBBS mentoring model on adolescent school reported grades, absenteeism and formal disciplinary records in a long-term follow-up study. We utilized propensity score matching to create equivalent groups, comparing those with at least one year of mentoring to unserved youth. The nearest neighbor propensity score model yielded a well-matched sample of 70 youth. Among these, the average follow-up time post initial program enrollment in both groups was about four and a half years. Treated youth all had successful matches in the BBBSHR program. All matches exceeded one year, the average match was nearly-two years. Follow-up time post-match closure for treated youth was nearly-two and a half years after match closure (at least one year for all subjects).

Results for the entire matched sample suggest the program had a small to moderate effect on overall Core Subjects GPA and likelihood of receiving a failing grade. We observed a 14% difference in rates of failing between the two groups. The program had a stronger impact on likelihood of receiving an unsatisfactory grade (D or F). Here, we observed a 20% difference between groups. Effects on GPA were most present among White and Black students, where we observed a nearly one-point GPA difference. Most notably, findings suggested that youth in mentoring had remarkably lower rates of absenteeism. Youth in the comparison group were absent, on average, six days more annually than those in the treatment. Here, effects were strongest among girls, who were absent over ten days less, on average. In this study, we attempted to identify a critical cut-point in absenteeism, based on the literature, that might

---

<sup>7</sup> Findings among the static groups suggest the treatment group was statistically less likely to have a formal disciplinary infraction (21.4%) than the comparison (42.9%) ( $\chi^2(1) = 5.66$ ;  $p = .02$ ).

indicate chronic loss of learning, directly related with a host of poor academic and life outcomes. We defined this as 18 or more days. Results showed an 11.5% difference in the rate of chronic absenteeism across groups, suggesting a small, but critically important, program impact. Here, again, we detected much more pronounced effect for girls (~22% difference between groups). Finally, we explored impact on formal school disciplinary records. While we did not detect any differences in youth likelihood of referral to ISS or OSS, we did find a small program effect on likelihood of receiving a formal infraction.

Generally, findings from this section of the study provide the clearest evidence of program impact for several reasons. First, data integrity issues with first person risk reporting are not present. School data provides us with clear, mostly objective, and verified information regarding student performance. Second, the performance gap is clear, several years after program delivery in most cases. Whereas there have been a few well controlled studies that demonstrate impact on school related outcomes during and in the short-run after program delivery (especially for those in successful match relationships), this study provides some compelling evidence that school impacts may be long lasting. Even modest differences, from this perspective, are extremely relevant.

## **Overall Limitations and Conclusions**

We were careful throughout this study to note the limited sample size, especially problematic once we created equivalence groups. Not only does this naturally limit the generalizability of the study (in principle), but it, in fact, limited our ability to detect significance. One hopeful note has been the growth of work, in the past few years, with similar findings (e.g., Blakeslee & Keller, 2018; DuBois & Herrera, 2018). Despite our low number of observations, this work, at least, helps to reinforce a trend in long-term impact findings.

While there are possible pre-test effects for treated youth, average length of time sense match closure was nearly three years. Therefore, it is doubtful that children were impacted by their exposure to the pretest. Also, social desirability bias and agency loyalty could have played a role in reporting more positive outcomes. However, a new staff member administered the surveys who did not normally support matches. Very few if any of the students would have recognized the staff member and felt an obligation to answer positively. Observations on pre-test outcomes were not included as covariates in the propensity score matching model.

Additional analysis is needed to follow-up on group differences within program type. Several studies note fundamentally different outcomes and rates of change between youth served in the BBBS school-based mentoring program and the BBBS community-based program. Because of the nature of this study, focusing on comparisons between those never matched with a BBBS

mentor, as well as the limited number of observations, such analysis was beyond the scope of this project. That was also true as it relates to understanding differential program impacts due to length of the match relationship. This is an area for future research. We would also like to explore fundamental differences in outcomes among those with 12 to 24 month matches and those with long-term or chronic match (greater than 24 months).

Despite these limitations, findings from this research project suggest a number of positive, long-term effects of youth mentoring among adolescents, which holds promise for broadening program support and investment. While this study has a relatively low sample size and was implemented with just one agency, our results duplicate important findings in the mentoring field, confirming the relevance of our results (e.g, Blakeslee & Keller, 2018; DuBois & Herrera, 2018). Second, the study does so in rural and small urban regions, whereas most of the larger long-term studies have been conducted in populous, urban areas. This, alone, presents important practical knowledge for mentoring practitioners, the majority of whom serve rural or suburban areas. Additionally, this study makes new contributions to understanding persistent program effect among those served as children (prior to age 12), pointing to mentoring as an important early intervention to prevent more serious risk later in life. Together these findings serve to reinforce a small but growing body of literature in the research community that points to youth mentoring as an effective and lasting intervention to improve the lives of at-risk youth.



## References

- Bernard, B. (2004). *Turning the corner: from risk to resiliency*. Minneapolis, MN: National Resiliency Resource Center.
- Blakeslee, J. E., & Keller, T. E. (2018). *Extending a Randomized Trial of My Life Mentoring Model for Youth in Foster Care to Evaluate Long-term Effects on Offending in Young Adulthood: Final Technical Report* (NCJRS# 251418).
- Converse, N., & Lignugaris/Kraft, B. (2009). Evaluation of a school-based mentoring program for at risk middle school youth. *Remedial and Special Education, 30*(1) 33-46.
- DuBois, D. L. & Silverthorn, N. (2005). Natural mentoring relationships and adolescent health: Evidence from a national study. *American Journal of Public Health, 95*(3), 518-524.
- DuBois, D. L., & Karcher, M. J. (Eds.). (2014). *Handbook of youth mentoring*. Sage Publications.
- DuBois, D. L., Portillo, N., Rhodes, J. E., Silverthorn, N., & Valentine, J.C. (2011). How effective are mentoring programs for youth? A systematic assessment of the evidence. *Psychological Science in the Public Interest, 12*(2), 57-91.
- DuBois, D. L. & Herrera, C. (2018). *Investigation of Long-Term Effects of the Big Brothers Big Sisters Community-Based Mentoring Program: Final Technical Report for OJJDP* (NCJRS# 251521).
- Esbensen, F. (2000). Preventing adolescent gang involvement. *Juvenile Justice Bulletin*. Washington, DC: Office of Juvenile Justice and Delinquency Prevention.
- Herrera, C., DuBois, D. L., & Grossman, J. B. (2013). The Role of Risk: Mentoring Experiences and Outcomes for Youth with Varying Risk Profiles. *MDRC*.
- Hurd, N., & Zimmerman, M. (2010). Natural mentors, mental health, and risk behaviors: A longitudinal analysis of African American adolescents transitioning into adulthood. *American Journal of Community Psychology, 46*(1-2), 36-48.
- Klaw, E. L., Rhodes, J. E., & Fitzgerald, L. F. (2003). Natural mentors in the lives of African American adolescent mothers: Tracking relationships over time. *Journal of Youth and Adolescence, 32*(3), 223-232.
- Krebs, C., Lattimore, P.K. Cowell A. J., & Graham P. (2010). Evaluating the Juvenile Breaking the Cycle Program's impact on recidivism. *Journal of Criminal Justice, 38*(2), 109- 117.
- Lerner, R. M., Napolitano, C. M., Boyd, M. J., Mueller, M. K., & Callina, K.S. (2014) Mentoring and positive youth development. In D. L. DuBois & M. J. Karcher (Eds.). *Handbook of youth mentoring* (pp. 17-27). Thousand Oaks, CA.

- Lerner, R., Brittan, A., & Fay, K. (2007). Mentoring: A key resource for promoting positive youth development. *Research in Action, 1*, 3-9.
- Linquanti, R. (1992). *Using community-wide collaboration to foster resiliency in kids: A conceptual framework*. Portland, OR: Western Regional Center for Drug-Free Schools and Communities.
- McCord, J., Widom, C. S., and Crowell, N. A., Eds. (2001). *Juvenile crime, juvenile justice. Panel on Juvenile Crime: Prevention, Treatment, and Control*. Washington, DC: National Academy Press.
- Moore, K. A. (2006). Defining the term “at-risk.” *Research to Results Brief*. Washington, DC: Child Trends.
- Morrow, K. V. and Styles, M. B. (1995). *Building relationships with youth in program settings: A study of Big Brothers/Big Sisters*. Philadelphia: Public/Private Ventures.
- Parra, G. R., DuBois, D. L., Neville, H. A., Pugh-Lilly, A. O., & Povinelli, N. (2002). Mentoring relationships for youth: Investigation of a process-oriented model. *Journal of Community Psychology, 30*(4), 367-388.
- Peaslee, L., & Teye, A. C. (2015). Testing the Impact of Mentor Training and Peer Support on the Quality of Mentor-Mentee Relationships and Outcomes for At-Risk Youth. (NCJRS# 248719).
- Rhodes, J. E. (2005). A model of youth mentoring. In D. L. DuBois & M. J. Karcher (Eds.), *Handbook of youth mentoring* (pp. 30-43). Thousand Oaks, CA: Sage.
- Rhodes, J. E. (2008). Improving youth mentoring interventions through research-based practice, *American Journal of Community Psychology, 41*(1-2), 35–42.
- Rhodes, J. E., Spencer, R., Keller, T. E., Liang, B., & Noam, G. (2006). A model for the influence of mentoring relationships on youth development. *Journal of Community Psychology, 34*(6), 691–707.
- Scales, P. (2003). *Other People's Kids: Social Expectations and American Adults' Involvement With Children and Adolescents*. Springer Science & Business Media.
- Shader, M. (2003). *Risk factors for delinquency: An overview*. Washington DC: U.S. Department of Justice, Office of Justice Programs, Office of Juvenile Justice and Delinquency Prevention.
- Thomson, N. R., & Zand, D. H. (2010). Mentees’ perceptions of their interpersonal relationships: The role of the mentor–youth bond. *Youth & Society, 41*(3), 434-445.

*DRAFT: DO NOT QUOTE OR DISTRIBUTE WITHOUT PERMISSION*

Tierney, J. P., Grossman, J. & Resch, N. (2000). *Making a difference: An impact study of Big Brothers/Big Sisters*. Philadelphia: Public/Private Ventures.