What's In a Job? Evaluating the Effect of Private Sector Job Experience on Students' Academic Outcomes

Urbashee Paul[*], Alicia Modestino[†], and Joseph McLaughlin[‡]

October 28, 2020

Abstract

This paper explores the effect of work experience on high school academic outcomes. Specifically, we examine the effects of participation in private sector employment through the Boston summer jobs program on students' academic outcomes. Unlike other cities, Boston places over 3,000 students each summer in private sector jobs through the city's Private Industry Council. Given this non-experimental research setting, we employ propensity score, Mahalanobis distance, and coarsened exact matching techniques to link participants to appropriate comparison groups based on detailed demographic and school characteristics from administrative school records. Preliminary results from each matching method show that participation in the program during the summer of 2015 significantly increased students' school attendance in the following academic year by approximately 2.8 days, decreased truancy by approximately 1.5 days, and increased the probability of graduating on time by 5.5 percentage points. These findings are similar in magnitude to what other researchers have found when using experimental methods to evaluate the impact of community-based job placements through the Boston summer jobs program on school outcomes. We also find significant improvements in standardized test scores, which may lead to better post-secondary outcomes. Our results are driven mainly by Black and Hispanic students, indicating that private sector summer jobs may be an effective approach to reduce inequality among youth.

Keywords: Attendance, Graduation, Summer Jobs, Matching Models

JEL Codes: I24; C13; D04

[*]Department of Economics, Northeastern University
[†]School of Public Policy and Urban Affairs and Department of Economics, Northeastern University
[‡]Boston Private Industry Council

1 Introduction

The past several decades have witnessed little improvement in narrowing the academic achievement gap that exists along socioeconomic and racial lines. A recent study conducted by the National Center for Education Statistics (de Brey et al., 2019) finds that the White-Black achievement gap in reading scores in 12th grade was wider in 2015 (30 points) than in 1992 (24 points). For math scores in 12th grade, neither the White-Black nor the White-Hispanic achievement gaps in 2015 were measurably different from corresponding gaps in 2005. These disparities in achievement persist at the postsecondary level. In 2000, 34 percent of White young adults between the ages of 25–29 had a bachelor's degree, compared to 17.8 percent of Black and 9.7 percent of Hispanic young adults. In 2017, these inequalities changed very little; the disparities between White (42.1 percent) and Black Americans (22.8 percent) increased, and the disparities between White and Hispanic Americans (18.5 percent) declined by just 1 percentage point.

While considerable progress has been made in reducing the Black-White and Hispanic-White gap in high school graduation rates, in 2015–16, the adjusted cohort graduation rates (ACGR)[1] for Black (76 percent) and Hispanic (79 percent) public high school students were below the national average of 84 percent (McFarland et al., 2018).

This paper estimates the effect of participating in a private sector job placement through the Boston Private Industry Council's (BPIC) summer jobs program. Due to the non-experimental setting, we comparatively apply several quasi-experimental matching models to estimate the program's impact. To the best of our knowledge, this paper is the first to examine the effects of private-sector, as opposed to public and subsidized, summer job experiences on high school students' academic achievements. We find evidence to suggest that private summer youth employment programs (SYEP) play an instrumental role in improving school outcomes, especially for Black and Hispanic students—thus helping to mitigate the achievement gap that exists between White and minority students.

[1] The adjusted cohort graduation rate provides information about the percentage of public high school students who graduate on time (i.e., 4 years after starting 9th grade for the first time) with a regular diploma. Those students who were awarded an alternate credential, such as a GED, are not included as graduates in the ACGR calculations.

The remainder of the paper is organized as follows. Section 2 offers an overview of recent relevant research. Section 3 describes the institutional background and placement process of the BPIC SYEP. Section 4 discusses the data and quasi-experimental matching methods we use to identify the effect of participation in summer jobs on students' academic outcomes. Following this, we present our preliminary results in Section 5. We conclude and share our plans for future expansions of this research in Section 6.

2 Literature

Policymakers and researchers have examined how time spent outside of the classroom can affect student outcomes, including reducing chronic absenteeism and improving high school graduation rates. Prior studies have shown, using quasi-experimental methods, that participating in sports boosts graduation rates (Stevenson, 2010) and overall participation in extracurricular activities can reduce dropout rates by up to 18 percentage points (Crispin, 2017).

Although prior literature on SYEPs has found strong positive impacts of work experience for reducing crime (Heller, 2014; Gelber, Isen and Kessler, 2016; Modestino, 2019), the evidence on whether work experience improves academic outcomes is more mixed. In the context of an experimental setting in New York City (NYC), Leos-Urbel (2014), for example, finds significant increases of one to two percentage points in school attendance for the treatment group relative to the control group during the year following participation in the NYC SYEP, with larger improvements for students aged 16 years and older who have low baseline attendance. Building on these results, Schwartz, Leos-Urbel and Wiswall (2015) find that the NYC SYEP increases the number of exams students attempt, the number of exams students pass, and the average score students achieve. Other researchers, however, find that the NYC SYEP did not have a positive effect on longer-term academic outcomes, such as graduating from high school (Valentine et al., 2017) or college enrollment (Gelber, Isen and Kessler, 2016).

Our results, in the setting of Boston's private-sector SYEP, corroborate the findings of earlier research that summer jobs indeed play a role in boosting students' academic performances. Preliminary results from our study show that participation in the program during the summer of 2015 significantly increased students' school attendance in the following academic year by approximately 2.8 days, decreased truancy by approximately 1.5 days, increased the probability of graduating on time by 5.5 percentage points, and improved standardized test scores.

3 Program Context

The main challenge of any impact evaluation study is the construction of a counterfactual outcome—in this case, what the academic outcomes of students who had not participated in the 2015 BPIC SYEP would have been had they not worked in a private sector job over the summer. When possible, researchers conduct experimental evaluations, in which assignment to treatment is random. Such a setting assures that participation in the intervention is the only differentiating factor between the treatment and control groups, thereby allowing the researcher to attribute the average difference in outcomes to the effect of participation in the intervention.

Modestino and Paulsen (2019), for example, conduct such a study on the effect of subsidized summer jobs on school outcomes using data from the Action for Boston Community Development (ABCD), one of the two intermediaries in Boston that makes use of random assignment . The BPIC, another nonprofit organization that places high school students in summer jobs, operates slightly differently from ABCD. This organization, in partnership with the City of Boston and the Boston Public Schools, places high school students (typically those with prior work experience and higher grades) with employer-paid private sector summer internships. Top employers include Massachusetts General Hospital, Brigham and Women's Hospital, Bank of America, Liberty Mutual Group, among others.

In our research setting, a randomized experiment is infeasible. Private employers, on one hand, would not allow for a random selection of high school students to work on their payroll; on the other hand, certain students are more inclined than others to work at private sector jobs during the summer. BPIC, therefore, does not randomly assign students to these private sector job, but rather, the organization works as a facilitator to students in applying for and connecting with participating employers. BPIC career specialists help students complete a résumé, explore available positions, and apply for jobs to employers. Students then interview with at least three employers and typically receive at least one job offer.

Due to our non-experimental setting, we cannot simply measure the impact of BPIC program participation on school outcomes by comparing participants to non-participants, as this method would suffer from selection bias. Students who applied and subsequently participated in the BPIC summer jobs program likely differ on many observable and unobservable characteristics from those who don't. , As we will show in Table 1, these differences include, but are not limited to, age, gender, race, academic records, and students' high school characteristics. This paper aims to disentangle these confounding factors from the effect of participating of Boston high school students in the BPIC summer jobs program on the following academic outcomes: attendance, truancy, grade point average (GPA), probability of graduating on time, and standardized test scores.

4 Data and Methodology

This study makes use of a large administrative database compiled by the Massachusetts Department of Elementary and Secondary Education (DESE). BPIC provides administrative data with only the personally identifiable information of 2015 BPIC participants, not names of those who applied and were rejected; this warrants the use of quasi-experimental techniques to evaluate the effectiveness of private summer jobs on school outcomes. We comparatively implement three matching techniques—Mahalanobis distance matching (MDM), coarsened exact matching (CEM), and propensity score matching (PSM)—to more precisely evaluate the effect of participating in the BPIC summer jobs program on BPS students' academic outcomes. Matching has been shown to reduce model dependence, bias, and variance (Ho et al., 2007).

In the summer of 2015, 1,231 from the Boston public school (BPS) system were placed directly into private-sector summer jobs. In the absence of an experimental design, students participating in the intervention may differ not only in their treatment status, but also in other characteristics that affect both their decision to participate and also in their outcome. The set of characteristics we chose to match on are those that we believe have a theoretical basis for determining selection into the intervention: gender, race, grade, and whether a student is categorized as availing free or reduced price lunches, English language learner (ELL), and special education. Baseline characteristics from the prior year such as the 2014-15 attendance days, truant days, weighted GPA, and course failures are included as matching factors in select models. For example, days attended in the year prior to participating in the program (2014-15) is included as a matching covariate for regressions with specifications that include days attended in the year after participating in the program (2015-16) and probability of graduation (2015-16), since baseline attendance records could affect selection into the program and also predict both outcomes.

As we see from Table 1, there are obvious differences in the characteristics between the average BPS student and the subset of BPS students participating in the BPIC summer jobs program in 2015. Males comprise 51 percent of the student body in the BPS system, and females comprise 49 percent. The majority, 62 percent, of our treatment participants, however, are female—suggesting that gender may be a factor motivating participation. Race may also play a role in selection. While White students make up 26 percent of the student body in the BPS system, they represent only 8 percent of the share of BPIC participants. Black students make up the largest share among participants, at 44 percent, followed by Hispanic students, at 25 percent.

When inspecting the distribution of grades for BPIC participants, we see that the majority of participants are in 11th and 12th grade. This is attributable to the Boston summer jobs ecosystem: younger students are generally triaged into community-based, subsidized employment through either BPIC or other community based organizations. Private sector jobs are more selective in nature and often require prior work experience; thus, older students are more likely to be placed in them.

Almost 70 percent of BPIC participants also participate in the free or reduced lunch program. This may indicate that their families come from more disadvantaged socioeconomic backgrounds.

Table 1: Descriptive Statistics

	BPS			***Participants***		
	Mean	Std.Dev.	Obs	Mean	Std.Dev.	Obs
All Regression Controls						
Male	0.51	0.50	44,494	0.38	0.49	1,231
Female	0.49	0.50	44,494	0.62	0.49	1,231
Race: Asian	0.08	0.28	44,494	0.19	0.39	1,231
Race: Black	0.32	0.46	44,494	0.44	0.50	1,231
Race: Hispanic	0.29	0.46	44,494	0.25	0.43	1,231
Race: White	0.26	0.44	44,494	0.08	0.27	1,231
Race: Other	0.05	0.22	44,494	0.05	0.22	1,231
Grade: 8	0.23	0.42	44,494	0.01	0.08	1,231
Grade: 9	0.23	0.42	44,494	0.06	0.23	1,231
Grade: 10	0.18	0.39	44,494	0.16	0.37	1,231
Grade: 11	0.18	0.38	44,494	0.37	0.48	1,231
Grade: 12	0.18	0.38	44,494	0.41	0.49	1,231
Free or Reduced Lunch	0.49	0.50	44,494	0.68	0.47	1,231
English Language Learner (ELL)	0.13	0.33	44,494	0.11	0.31	1,231
Special Education	0.28	0.45	44,494	0.11	0.32	1,231
Regression Specific Controls						
Attendance Days 2014-15	154.87	42.60	39,077	168.02	14.90	1,230
Truant Days 2014-15	10.00	19.07	39,577	8.05	10.74	1,230
Weighted GPA 2014-15 (increments of .5)	4.21	2.38	28,889	4.67	2.01	734
Course Failures 2014-15	1.67	2.53	30,373	1.38	1.87	735

Description: The above table displays descriptive statistics for (1) Boston public school students, and (2) the 2015 Boston private summer jobs participants. Source: Authors' calculations using Massachusetts Department of Elementary and Secondary Education (DESE) and Boston Private Industry Council (BPIC) data.

Whether this affects selection into treatment is unclear.

Selection bias might not play a large role for English language learners (ELL), given that the percentage of ELL students participating in BPIC is fairly similar to the percentage in the BPS system. We do, however, see a smaller share of special education students participating in BPIC than are present in the BPS system. It may be that special education students are more likely to attend summer school instead of acquiring work experience during the summer so as to maintain academic continuity.

Matching helps to alleviate selection bias issues that would be present if a researcher were to simply evaluate the effect of participation on an outcome using an ordinary least squares (OLS) model. The general purpose of matching, across different techniques, is to find one or more non-treated units that are "similar" across all relevant characteristics to a treated unit. The concept of similarity in characteristics varies from one technique to the next, and also relies on the researcher discretion. Once treated observations are matched appropriately, all other unmatched observations are dropped from the analysis. The researcher must then run a regression on the refined dataset to estimate the average treatment effect on the treated (ATT). Nuances of each matching technique is presented in the following subsections.

4.1 Mahalanobis Distance Matching (MDM)

MDM aims to match observations based on minimizing the distance between the vector of covariates for treated and control units. Specifically, MDM matches are created by minimizing the Mahalanobis distance measure, $M(X_i, X_j)$, between two vectors of characteristics X_i and $X_j \in X$ for individuals i and j, respectively:

$$M(X_i, X_j) = \sqrt{(X_i - X_j)' S^{-1} (X_i - X_j)}, \tag{1}$$

where S^{-1} is the inverse of the sample covariance matrix of X.

When implementing MDM, we allow for one-to-many matching and restrict matches to have a maximum vector distance between covariates of treatment and control distance of 1.5[2]. We know from prior research that characteristics such as gender, grade, race, English proficiency, and socioeconomic status are correlated with applying to the summer jobs program. In addition, the structure of the BPIC program with career coaches being present in some BPS high schools suggest that school would also be highly correlated with applying. These are all characteristics observable in our rich administrative dataset, and thus, we are able to match on these characteristics across all matching models. However, it might still be the case that there are unobservable factors such as motivation or ability that affect selection into which students apply to the program. We use the prior year's school attendance and GPA as proxies for these unobservable characteristics.

Once matching is complete and unmatched units are pruned, we proceed to estimating the average treatment effect on the treated using a one stage estimation estimation procedure, as follows:

$$Y_i = \beta_0 + \beta_1(T_i * w_i) + \gamma X_i + \rho S_i + \epsilon_i, \tag{2}$$

where β_1 is the average treatment effect on the treated, T_i is the treatment indicator, w_i is the weight assigned to individual i, Y_i is the outcome variable, and S_i is the school fixed effect. γ is a vector of coefficients for individual controls, X_i, as denoted in Table 1. Standard errors, ϵ_i, are robust at the student level. When the outcome variable Y_i is dichotomous (whether a student graduates on time, for example), equation 2 is run as a logitic regression.

4.2 Coarsened Exact Matching (CEM)

CEM is a "Monotonic Imbalance Bounding" (MIB) class of matching methods for causal inference, introduced by Iacus, King and Porro (2011). Unlike MDM, CEM guarantees that the imbalance between the matched treated and control units will be bounded ex ante at a user-specified

[2] We experimented with a series of alternative bandwidths (see Appendix A) and found that between a bandwidth of covariate vector distance 1.25 and 1.5, we get improved balance and a larger number of matches, as indicated by the number of covariates with significant differences post-matching (see Table 3). Under these restrictions, we find that each treatment unit is matched with up to 15 comparison units.

level.

CEM is implemented by, first, temporarily coarsening each covariate when two values of a particular variable are substantially indistinguishable. For example, when matching on students' prior attendance records, attendance days may be coarsened into increments of 3 days because a student who attended 97 days of school is likely not substantively different from one who attended 94 days. This coarsening allows for better matches and less trimming of the dataset. Next, strata are formed, where units with the same values for all coarsened characteristics are grouped together. To illustrate, in our dataset, all female, Asian students in grade 10, with perfect attendance records are likely placed in the same stratum. When a stratum does not have at least one treated unit and one control unit, it is pruned from the dataset.

Weights are assigned as follows: treated units receive a weight of one. Control units are weighted as the number of treated units in its stratum divided by the number of control units in the same stratum, normalized so that the sum of the weights equals the total matched sample size[3].

For the purpose of our analysis, we restrict further coarsening on all of our covariates except for continuous baseline variables in select models such as pre-period attendance days, truant days, weighted mean GPA, and course failures. In other words, we are forming strata with exact matches on the dichotomous variables for students' gender, race, grade, school, and participation in ELL, special education, and free or reduced lunch. Once strata are formed and unmatched units are pruned, we proceed to estimating the average treatment effect on the treated using equation 2.

4.3 Propensity Score Matching (PSM)

Heckman (1977) and Rosenbaum and Rubin (1983) developed an approach commonly known as PSM for estimating causal effects from non-experimental data. This approach has gained popularity among researchers across disciplines, especially economics and health, in the recent decades. Rosenbaum and Rubin (1983) lay out the theoretical framework of PSM and stipulate two conditions for PSM to correctly estimate the impact of a program: 1) the Condition Independence Assumption

[3]Thus, in a stratum containing one BPIC participant and 5 matches, each matched unit will receive a weight of one-fifth.

(CIA) and 2) the Common Support Condition (CSC). The CIA holds when observations are assigned to treatment based only on observable characteristics. If CIA doesn't hold, or in other words, participation in the treatment is likely driven by unobservable factors, then the matching estimator may be seriously biased. Given our rich administrative dataset that encompasses many observable characteristics that would predispose students' assignment to treatment, we believe CIA is satisfied. Unobservable factors such as motivation or ability that affect selection are proxied for using baseline attendance and GPA in relevant models.

The CSC requires that a substantial overlap exist between propensity scores of treated and untreated individuals. If this condition fails to hold, then we cannot construct a counterfactual comparison group to estimate the impact of the intervention. In our setting, only those enrolled in a Boston public high school were eligible to participate in BPIC and, to our knowledge, each BPS is assigned at least one career coach responsible for disseminating information about the BPIC summer jobs program to students. To ensure that untreated students in our sample would have some propensity to participate in the program, and to meet the CSC requirement, we limit our comparison group to only schools that have at least one BPIC participant.

We implement PSM using the following steps:

1. *Select covariates*

   Just as in OLS regressions, covariates should be selected such that they are related to the participant's self-selection into the intervention. Omitting important covariates related to an individual's likelihood of participating in the treatment will result in bias, and thus, the researcher must strive to include any observable characteristic that is at his/her disposal to mitigate omitted variable bias. We use the same vector of covariates that we use for MDM and CEM, as detailed in Table 1.

2. *Select model for generating propensity scores*

   Given that participation in the treatment is dichotomous, researchers typically use a logit or probit model to create propensity scores (Heinrich et al., 2010). The most frequently used method for creating propensity scores is logistic regression (Austin, 2011; Stuart, 2010). A propensity score $P(T_i = 1|X_i) = P(X_i)$, where $T_i = 1$ indicates that the individual participates in treatment conditional on his observed vector of characteristics X_i and $T_i = 0$ otherwise, is a measure of likelihood of participation in treatment for each individual for individual i. We run the first stage logit regression iteratively[4] with school fixed effects S_i, to control for heterogeneity among Boston schools, as follows:

$$logit(T_i) = \alpha_0 + \alpha X_i + \rho S_i + \eta_i \tag{3}$$

3. *Select a matching algorithm and create matches*

   We implement one-to-many matching with replacement[5], and allow for a caliper width equal to 0.2 of the standard deviation of the logit of the propensity score[6]. After matches are created, we trim any unmatched observations from our dataset before proceeding to estimating treatment effects.

4. *Check balance*

   We assess the quality of our matches by checking the balance of the distribution of our vector of covariates for the treatment and comparison groups. Specifically, we take the difference in means of both groups on each covariate to determine whether the groups differ significantly on any of the relevant matching characteristics. As we see in Table 3, the differences in means after matching using PSM is less significant across most covariates. The balance performance using PSM, however, is lacking when compared to other matching techniques such as CEM and MDM across covariates.

[4]We run one logit regression for each outcome variable with their appropriate matching covariates and school fixed effects as controls. For example, when the outcome of interest in the second stage is attendance days in the 2015-16 academic year, we run a logit regression with a treatment indicator as the outcome on the fixed set of covariates listed in Table 1 as well as baseline 2014-15 attendance days and school fixed effects. When the outcome of interest in the second stage is a test score such as the SAT, we do not include a baseline test score since a large number of students only take the test once.

[5]To maintain consistency with MDM, where we allow for one-to-many matching and find that each observation was matched with, at most, 15 other comparison units, we allow up to 15 matches for our propensity score matching.

[6]Austin (2011) finds that a caliper width of 0.2 times the standard deviation of the logit of the propensity score is optimal because it minimizes the mean square error of the resultant estimated treatment effect when at least some of the covariates are continuous.

5. *Estimate effect of intervention*

   To estimate the average treatment effect of the BPIC summer jobs program on the participants' academic outcomes, we run a second-stage OLS regression for continuous outcome variables and a logit regression for binary outcome variables on the treatment indicator T for student i, weighted by the inverse of the number of matches for each treated student, denoted w_i:

$$Y_i = \beta_0 + \beta_1(T_i * w_i) + \epsilon_{is} \tag{4}$$

   Standard errors are clustered at the school level, as indicated by the subscript s on the error term ϵ_{is}, because we match iteratively at the school level to achieve more accuracy in matching . β_1 represents the average treatment effect on the treated.

4.4 Method Comparison

While PSM appears superior when considering post-matching balance in our sample, Nielsen and King (2019) point out that the method has several weaknesses, and should be used only in conjunction with other matching methods. One weakness is that PSM approximates random matching, rather than a fully blocked experiment. Complete randomization balances the treated and untreated units on average, whereas a fully blocked experiment exactly balances the covariates for the observed treated and untreated units. King et al. (2009) find that standard errors in a fully blocked experiment are, on average, 600% smaller. PSM is efficient relative to complete randomization, but it is inefficient compared to a fully blocked experiment.

PSM also suffers from the "PSM Paradox": as the propensity to be treated or untreated approximates randomization (or a propensity score of 0.5), it gives rise to more pruning at random, which in turn increases imbalance, inefficiency, model dependence, and bias. In addition, PSM is susceptible to the curse of dimensionality: as the number of covariates increases, the logit regression may become worse at predicting the probability of treatment (especially with irrelevant covariates), and the PSM paradox gets significantly worse. These weaknesses are attributable to PSM's two-stage procedure. Much valuable information is lost in the first-stage logit regression, where all covariates playing a strong role in selection into treatment are reduced to a one dimensional propensity score.

MDM and PSM are similar in that for both methods, the researcher may set the caliper width and number of matches per treatment for matching ex ante. Matching and checking imbalance occurs ex post. For CEM, the researcher specifies the desired balance ex ante by restricting matching within strata based on a collection coarsened covariate characteristics. The number of matches are realized ex post. King et al. (2011) conclude that researchers should not necessarily discard PSM as a matching method, but use it in combination with other techniques to compare results.

MDM and CEM both approximate a fully blocked experiment. In the literature, it is not quite clear which method is more superior. After iterating on multiple caliper bandwidths for MDM, we find that post-matching balance is very similar to that of CEM. In addition, for our key academic outcome variables (see Table 1), our MDM regressions capture more matches and, thereby, have more power than our CEM regressions. We find that MDM allows for more flexibility and precision in matching because we can specify which covariates we wish to exactly match on as well as the bandwidth of the Mahalanobis distance. With these considerations in mind, our preferred specification for our analysis is MDM.

4.5 Balance

Our initial comparison group is constructed by keeping only BPS schools that have at least one BPIC participant. This trimming serves the purpose of further alleviating selection bias issues. For example, if a particular school in the BPS system has no participants, it could be that the school differs in terms of summer activities typically availed by students, proximity to jobs, promotion efforts for the BPIC program, and various other unknown factors.

As we see from Table 2, prior to matching, all differences between key covariates are statistically significant. One goal of matching is to achieve less significant differences across all covariates that have a theoretical basis for selection into treatment.

Table 2: Baseline Balance - Prior to Matching

Covariates	Comparison	Treatment	Difference
Male	0.523	0.381	-0.142***
	(0.499)	(0.486)	(0.015)
Female	0.477	0.619	0.142***
	(0.499)	(0.486)	(0.015)
Race: Asian	0.098	0.185	0.087***
	(0.297)	(0.389)	(0.009)
Race: Black	0.333	0.439	0.106***
	(0.471)	(0.496)	(0.014)
Race: Hispanic	0.285	0.248	-0.038***
	(0.452)	(0.432)	(0.013)
Race: White	0.236	0.078	-0.158***
	(0.425)	(0.268)	(0.012)
Race: Other	0.048	0.050	0.002
	(0.214)	(0.219)	(0.006)
Grade: 8	0.057	0.006	-0.051***
	(0.232)	(0.080)	(0.007)
Grade: 9	0.275	0.055	-0.220***
	(0.446)	(0.229)	(0.013)
Grade: 10	0.228	0.158	-0.070***
	(0.420)	(0.365)	(0.012)
Grade: 11	0.219	0.366	0.147***
	(0.414)	(0.482)	(0.012)
Grade: 12	0.221	0.413	0.193***
	(0.415)	(0.493)	(0.012)
Free or Reduced Lunch	0.550	0.675	0.125***
	(0.497)	(0.469)	(0.015)
English Language Learner (ELL)	0.171	0.111	-0.059***
	(0.376)	(0.315)	(0.011)
Special Education	0.179	0.115	-0.064***
	(0.383)	(0.319)	(0.011)
Attendance Days 2014-15	149.967	168.017	18.050***
	(45.451)	(14.904)	(1.300)
Truant Days 2014-15	10.942	8.046	-2.897***
	(19.309)	(10.741)	(0.555)
Weighted GPA 2014-15 (increments of .5)	4.067	4.670	0.603***
	(2.389)	(2.007)	(0.090)
Course Failures 2014-15	1.679	1.376	-0.303***
	(2.218)	(1.868)	(0.083)
Observations	22,538	1,231	23,769

Standard errors in parentheses
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Description: The above table displays baseline means and differences in means for the comparison and treatment groups before matching. Source: Authors' calculations using Massachusetts Department of Elementary and Secondary Education (DESE) and Boston Private Industry Council (BPIC) data.

After matching, we see a reduction in mean differences across all covariates. Table 3 displays differences in means for all covariates in the model where 2015-16 attendance days is the outcome variable. Thus, 2014-15 attendance days appears as the baseline control in the table[7].

Table 3: Balance - Post-Matching - Attendance Days 2015-16

	Difference in Means (Treated - Untreated)			
Covariates	Unmatched	PSM Matched	CEM Matched	MDM Matched
Male	-0.142***	-0.078***	-0.015	-0.026
	(0.015)	(0.020)	(0.023)	(0.020)
Female	0.142***	0.078***	0.015	0.026
	(0.015)	(0.020)	(0.023)	(0.020)
Free or Reduced Lunch	0.125***	0.003	0.019	-0.020
	(0.015)	(0.019)	(0.021)	(0.019)
English Language Learner (ELL)	-0.059***	-0.035**	0.024*	0.020
	(0.011)	(0.015)	(0.014)	(0.013)
Special Education	-0.064***	-0.037***	0.029***	0.065***
	(0.011)	(0.014)	(0.010)	(0.010)
Race: Asian	0.087***	0.032**	0.002	0.026*
	(0.009)	(0.014)	(0.017)	(0.015)
Race: Black	0.106***	0.017	0.040*	0.002
	(0.014)	(0.020)	(0.023)	(0.021)
Race: Hispanic	-0.038***	-0.030	0.015	-0.012
	(0.013)	(0.018)	(0.020)	(0.019)
Race: White	-0.158***	-0.018	-0.070***	-0.056***
	(0.012)	(0.011)	(0.015)	(0.013)
Race: Other	0.002	-0.002	0.013**	0.039***
	(0.006)	(0.009)	(0.005)	(0.006)
Grade: 8	-0.051***	-0.005	-0.004	-0.002
	(0.007)	(0.005)	(0.005)	(0.005)
Grade: 9	-0.220***	-0.052***	-0.024*	-0.037***
	(0.013)	(0.014)	(0.015)	(0.014)
Grade: 10	-0.070***	-0.074***	-0.072***	-0.066***
	(0.012)	(0.019)	(0.022)	(0.019)
Grade: 11	0.147***	0.140***	0.111***	0.117***
	(0.012)	(0.020)	(0.023)	(0.021)
Grade: 12	0.193***	-0.008*	-0.011*	-0.013**
	(0.012)	(0.005)	(0.006)	(0.006)
Attendance Days 2014-15	18.050***	3.612***	-0.064	-0.720
	(1.300)	(0.744)	(0.449)	(0.449)
Observations	23,769	5,348	2,739	3,464

Standard errors in parentheses
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Description: The above table displays baseline means and differences in means for the comparison and treatment groups after matching. Source: Authors' calculations using Massachusetts Department of Elementary and Secondary Education (DESE) and Boston Private Industry Council (BPIC) data.

[7] All other post-matching balance tables by outcome can be shared upon request from the authors.

5 Results

5.1 Secondary School Outcomes

In this section, we present our MDM, CEM, and PSM matching results for all outcome variables. Our first set of models look at school outcome variables: attendance, truancy, weighted mean grade point average (GPA), course failures, and probability of graduating on time.

Our main set of results using our preferred matching technique, MDM, are displayed in Table 4. We find that participation in BPIC during the summer of 2015 led to an increase in attendance of approximately 3 days; this was driven by a reduction of roughly 1.5 truant days. Although not statistically significant, the results show an improvement in GPA[8] and a small reduction in course failures for the treatment group. Finally, students who participated in BPIC saw an increase in probability of graduating on time of around 5.5 percentage points. For comparison, we have also displayed results from using the CEM and PSM models in Table 4. Generally, we find that the alternative matching methods have less power; but across all method results, the coefficients for all key explanatory variables are consistent in signs.

Our quasi-experimental research design yields results almost identical to Modestino and Paulsen (2019)'s experimental study. Their study finds an improvement in attendance days of approximately 3 days, a reduction in truant days of approximately 1.5 days, and an improvement in likelihood of graduating on time of 6.1 percentage points for high school students, resulting from participation in a subsidized Boston summer jobs program with randomized assignment. Although our coefficient for GPA is not statistically significant, it is the same as what Modestino and Paulsen (2019) find as an improvement resulting from participating in subsidized summer jobs—0.08 points higher than non-participants' GPA.

[8]Students' grade point averages were coarsened into increments of 0.5 to allow for better matching and more observation retention.

Table 4: Matching Results - Academic Outcomes

	(1)	(2)	(3)	(4)	(5)
MDM	Attendance	Truancy	GPA	Failures	Pr(Graduation)
Treated	2.794***	-1.589***	0.0854	-0.177	0.0554***
	(0.773)	(0.475)	(0.0989)	(0.104)	(0.0123)
N	3,464	3,552	1,475	1,599	3,526
CEM	Attendance	Truancy	GPA	Failures	Pr(Graduation)
Treated	2.675**	-0.707	0.0618	-0.215	0.0856***
	(0.997)	(0.463)	(0.137)	(0.128)	(0.0173)
N	2,739	2,986	563	1,080	2,761
PSM	Attendance	Truancy	GPA	Failures	Pr(Graduation)
Treated	3.678*	-1.477	0.166	-0.106	0.122***
	(1.367)	(0.893)	(0.158)	(0.113)	(0.0260)
N	5,348	5,300	2,371	2,495	5,948

Standard errors in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Description: This table displays results from running the regression models specified in equation 2 for MDM and CEM, and equation 4 for PSM. Outcome variables are specified under column labels (1) to (5). Source: Authors' calculations using Massachusetts Department of Elementary and Secondary Education (DESE) and Boston Private Industry Council (BPIC) data.

5.2 Test Scores

We explore whether participating in a private-sector summer job improves students' standardized test performance. Because only a minority of students in our sample retake the MCAS AP, and SAT exams, we do not include pre-program test scores as a control for test score regressions. Our MDM model results for standardized test scores are displayed in Table 5. Results from CEM and PSM models can be found in Appendix B.

In Massachusetts, students sit for the MCAS in 10th grade, and passing the exam is a requirement for graduation. It should be noted that our results reflect the relative performances of students who were in the 9th grade during the summer of 2015 and subsequently took the test after participating in the program. We find that participation in treatment led to a significant increase, of roughly 3 points, in students' MCAS scores for both the English language arts (ELA) section and the math section. This is in contrast to Leos-Urbel (2014) and Modestino and Paulsen (2019), who find no improvement in test scores when studying the effect of participating in subsidized summer job programs in NYC and Boston, respectively. One reason for this difference could be that private sector jobs, often in professional, technical, healthcare, and STEM fields offer experiences that differ from subsidized summer jobs[9]. We do not, however, detect any notable reduction in MCAS failures. This means that although students are perform better on the MCAS after participation in treatment, their improvement in MCAS scores is not the driving force behind the greater likelihood of graduating from high school we see in Table 3.

Very few students in our sample sit for the AP exam, and so we are unable to draw any noteworthy conclusions on students' AP exam performance. A larger subset of our sample has taken the SAT exam, a widely used admission prerequisite for colleges in the US, and the results are promising. Many of the leading SAT test preparation companies guarantee improvements in scores of 100-150 points[10]. Our results show that those who participated in a private-sector summer job scored, on average, 40 points higher on the SAT than our comparison group. Participants' average overall SAT score was 1402 out of 1600, putting them well within the admissions range to apply to most Massachusetts state and private schools[11].

[9]Subsidized summer jobs typically involve working in a daycare or a day camp.

[10]For the current SAT exam with a maximum score of 1600, Ivy Bound guarantees a score improvement of 100 points, Prep Scholar guarantees a score improvement of 100 points, and Princeton Review guarantees a score improvement of 150 points.

[11]Based on 2020 SAT score standards according to College Tuition Compare site.

Table 5: MDM Results - Test Scores

	(1)	(2)	(3)	(4)	(5)	(6)
MCAS	Math	ELA	Math Proficient	ELA Proficient	Math Failing	ELA Failing
Treated	3.439**	3.079***	0.0920*	0.0973***	-0.0273	-0.0720
	(1.079)	(0.640)	(0.0379)	(0.0285)	(0.0314)	(0.0539)
N	1,280	1,289	995	1,137	693	513
SAT	Overall	Math	Reading	Writing		
Treated	41.20***	11.28*	12.50**	17.42***		
	(11.37)	(4.670)	(4.539)	(4.481)		
N	1,771	1,771	1,771	1,771		
AP	Overall	Math/Sci	English/Arts/FL	Social Sci		
Treated	0.0124	-0.0373	-0.0600	0.113		
	(0.0563)	(0.0926)	(0.0899)	(0.236)		
N	2,210	710	593	187		

Standard errors in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Description: This table displays results from running the regression model specified in equation 2 for MDM. Outcome variables are specified under column labels (1) to (6). Test type is specified in bold letters on the leftmost column. Source: Authors' calculations using Massachusetts Department of Elementary and Secondary Education (DESE) and Boston Private Industry Council (BPIC) data.

5.3 Analysis of Heterogeneity Among Subgroups

To analyze whether the impact of participating in the BPIC SYEP was more pronounced for some students than for others, we conduct a subgroup analysis. We inspect four categories of students: i) students with poor attendance records who miss 10 or more percent of school days, ii) students in grades 9 and 10 who have more time to benefit from participating in the program, iii) male students, since historical data show that males exhibit poorer performance than females on a number of academic dimensions including high school completion[12] and dropout rates[13], and iv) Black and Hispanic students, as historical data show that students who identify as Black or Hispanic are less likely than students of other races to graduate from a public high school on time[14].

Tables 6 through 8 exhibit our subgroup analysis results from key MDM estimations for which we saw significant main effects. Estimations for other outcome variables and other methods can be

[12]See National Center for Education Statistics Table 219.65.
[13]See National Center for Education Statistics Table 219.70.
[14]See National Center for Education Statistics Table 219.46.

found in Appendix C.

Table 6 displays the results of our subgroup analysis for the 2015-16 attendance outcome. The column titled "N" shows the number of observations corresponding to the covariate name on the leftmost column. For example, there are 725 treated students that were matched in the MDM regression model with 2015-16 attendance as the outcome variable; 109 students with poor attendance records (below 90%) participated in treatment. The second row of the table displays the number of students that are matched among each subcategory, but not necessarily treated. For example, 777 students are in the 9th and 10th grade in this regression model, 260 of whom are treated.

Black and Hispanic students see an improvement of 4.5 attendance days after participating in the BPIC SYEP. It is worth noting that once we include the interaction term for Black or Hispanic students, the coefficient on the treatment indicator is no longer significant and switches from positive to negative, meaning that Black and Hispanic students are the main driver for our positive and significant main effect of 2.75 days. There is no marginal impact of participating in treatment on attendance days for students in other subgroups of interest.

Table 6: MDM Marginal Effects - Attendance Days 2015-16

	N	Main Effect	Attendance (< 90%)	Grades 9 & 10	Male	Black or Hispanic
			166	777	468	903
Treated	725	2.794***	2.263*	2.019	2.467*	-1.275
		(0.773)	(0.945)	(1.203)	(1.178)	(1.807)
T*(Attendance (< 90%))	109		-2.243			
			(2.940)			
T*(Grades 9 & 10)	260			-0.0832		
				(2.132)		
T*(Male)	266				-1.311	
					(2.124)	
T*(Black or Hispanic)	514					4.502*
						(2.161)
N		3,464	1,249	1,249	1,249	1,249

Standard errors in parentheses
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

In Table 7, we see that there is no marginal effect of participating in treatment on truant days for students in the specified subgroups. Although we lack power to draw statistical conclusions, it appears that Black and Hispanic students are driving much of the reduction in truancy we see in the main effect; in the last column, we see that after including the Black and Hispanic treated interaction term, the main effect coefficient becomes positive, but remains negative for other interaction terms. It is unclear, however, what the mechanism is that is driving these results. It could be that Black and Hispanic students suffer disproportionately high rates of truancy prior to the program due to discriminatory disciplinary practices and that work experience reduces discrimination for this group.

Table 7: MDM Marginal Effects - Truant Days 2015-16

	N	Main Effect	Attendance (< 90%) 183	Grades 9 & 10 766	Male 460	Black or Hispanic 901
Treated	727	-1.589***	-1.107*	-0.792	-1.117	0.635
		(0.475)	(0.513)	(0.711)	(0.735)	(1.116)
T*(Attendance (< 90%))	110		2.262			
			(2.488)			
T*(Grades 9 & 10)	262			-0.0931		
				(1.291)		
T*(Male)	268				0.806	
					(1.166)	
T*(Black or Hispanic)	516					-2.020
						(1.347)
N		3,552	1,237	1,237	1,237	1,237

Standard errors in parentheses
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

In Table 8, we see that Black and Hispanic students also drive the improvement in probability of graduating on time that we see in our main results, with an increase in probability of graduating on time of 6.6 percentage points after participating in the treatment. This may be due to the improvement in attendance we see in Table 6, combined with a reduction in truant days. Our results show, promisingly, that the BPIC SYEP is effective in improving secondary school outcomes across the board, but especially for Black and Hispanic students. This may lead to a reduction in inequality we see in the achievement gap between White and minority students.

Table 8: MDM Marginal Effects - Probability of Graduating on Time

		Main Effect	Attendance (< 90%)	Grades 9 & 10	Male	Black or Hispanic
	N		355	1,617	813	1,467
Treated	1,230	0.0554***	0.0434***	0.0395***	0.0616***	0.00344
		(0.0123)	(0.0104)	(0.0108)	(0.0130)	(0.0120)
T*(Attendance (< 90%))	220		0.0380			
			(0.0237)			
T*(Grades 9 & 10)	263			0.0467		
				(0.0306)		
T*(Male)	468				-0.0320	
					(0.0217)	
T*(Black or Hispanic)	844					0.0657***
						(0.0183)
N		3,526	2,094	2,094	2,094	2,094

Standard errors in parentheses
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

6 Conclusion

Our study shows, using quasi-experimental methods, that participation in private-sector summer jobs leads to improvements in academic outcomes for Boston public high school students. Students who participated in the 2015 BPIC SYEP exhibited an increase in attendance of approximately 3 days, a reduction in truancy of roughly 1.5 days, and an improvement in the probability of graduating on time of 5.5 percentage points. Notably, we find that students who participate in private sector summer jobs also see higher MCAS and SAT scores than comparable non-participants. Earlier papers studying the effects of subsidized summer jobs programs did not see significant improvements in test scores. These outcomes pave the way for promising post-secondary school outcomes for BPIC summer jobs participants. Thus, policymakers should expand partnerships with private employers to allow for more private summer job opportunities for youth.

Our subgroup analysis of the effect of private-sector summer job participation on academic outcomes yielded an important finding: the BPIC SYEP plays a role in reducing the disparity we see in achievements between White and minority, particularly Black and Hispanic, students. We find that Black and Hispanic students see the largest improvements in attendance days and probability of graduating on time as a result of availing private-sector summer jobs.

In the next phase of this project, we will be expanding our results to include post-secondary outcomes: college enrollment, completion, and choice of major using National Student Clearinghouse (NCS) data. We also plan on eventually incorporating post-graduation employment data to analyze whether there is a tangible impact of participating in private-sector summer jobs on students' choice of employment industry and starting income.

References

Austin, Peter C. 2011. "Optimal caliper widths for propensity-score matching when estimating differences in means and differences in proportions in observational studies." *Pharmaceutical statistics* 10(2):150–161.

Crispin, Laura M. 2017. "Extracurricular participation,"at-risk" status, and the high school dropout decision." *Education Finance and Policy* 12(2):166–196.

de Brey, Cristobal, Lauren Musu, Joel McFarland, Sidney Wilkinson-Flicker, Melissa Diliberti, Anlan Zhang, Claire Branstetter and Xiaolei Wang. 2019. "Status and Trends in the Education of Racial and Ethnic Groups 2018. NCES 2019-038." *National Center for Education Statistics* .

Gelber, Alexander, Adam Isen and Judd B Kessler. 2016. "The effects of youth employment: Evidence from new york city lotteries." *The Quarterly Journal of Economics* 131(1):423–460.

Heckman, James J. 1977. Dummy endogenous variables in a simultaneous equation system. Technical report National Bureau of Economic Research.

Heinrich, Carolyn, Alessandro Maffioli, Gonzalo Vazquez et al. 2010. "A primer for applying propensity-score matching." *Inter-American Development Bank* .

Heller, Sara B. 2014. "Summer jobs reduce violence among disadvantaged youth." *Science* 346(6214):1219–1223.
URL: *https://science.sciencemag.org/content/346/6214/1219*

Ho, Daniel E, Kosuke Imai, Gary King and Elizabeth A Stuart. 2007. "Matching as nonparametric preprocessing for reducing model dependence in parametric causal inference." *Political analysis* 15(3):199–236.

Iacus, Stefano M, Gary King and Giuseppe Porro. 2011. "Multivariate matching methods that are monotonic imbalance bounding." *Journal of the American Statistical Association* 106(493):345–361.

King, Gary, Emmanuela Gakidou, Kosuke Imai, Jason Lakin, Ryan T Moore, Clayton Nall, Nirmala Ravishankar, Manett Vargas, Martha María Téllez-Rojo, Juan Eugenio Hernández Ávila et al. 2009. "Public policy for the poor? A randomised assessment of the Mexican universal health insurance programme." *The lancet* 373(9673):1447–1454.

King, Gary, Richard Nielsen, Carter Coberley, James E Pope and Aaron Wells. 2011. "Comparative effectiveness of matching methods for causal inference." *Unpublished manuscript* 15(1):41–67.

Leos-Urbel, Jacob. 2014. "What Is a Summer Job Worth? The Impact of Summer Youth Employment on Academic Outcomes." *Journal of Policy Analysis and Management* 33(4):891–911.
URL: *https://onlinelibrary.wiley.com/doi/abs/10.1002/pam.21780*

McFarland, Joel, Jiashan Cui, Amy Rathbun and Juliet Holmes. 2018. "Trends in High School Dropout and Completion Rates in the United States: 2018. Compendium Report. NCES 2019-117." *National Center for Education Statistics* .

Merolla, David M and Omari Jackson. 2019. "Structural racism as the fundamental cause of the academic achievement gap." *Sociology Compass* 13(6):e12696.

Modestino, Alicia Sasser. 2019. "How Do Summer Youth Employment Programs Improve Criminal Justice Outcomes, and for Whom?" *Journal of Policy Analysis and Management* 38(3):600–628.

Modestino, Alicia Sasser and Richard J Paulsen. 2019. "Reducing inequality summer by summer: Lessons from an evaluation of the Boston Summer Youth Employment Program." *Evaluation and Program Planning* 72:40–53.

Nielsen, Richard and Gary King. 2019. "Replication Data for: Why Propensity Scores Should Not Be Used for Matching.".

Rosenbaum, Paul R and Donald B Rubin. 1983. "The central role of the propensity score in observational studies for causal effects." *Biometrika* 70(1):41–55.

Ross, Martha and Richard Kazis. 2016. "Youth summer jobs programs: Aligning ends and means." *Washington, DC: Brookings Institution* .

Schwartz, Amy Ellen, Jacob Leos-Urbel and Matthew Wiswall. 2015. Making summer matter: The impact of youth employment on academic performance. Technical report National Bureau of Economic Research.

Stevenson, Betsey. 2010. "Beyond the Classroom: Using Title IX to Measure the Return to High School Sports." *The Review of Economics and Statistics* 92(2):284–301.
URL: *https://doi.org/10.1162/rest.2010.11623*

Stuart, Elizabeth 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.

Valentine, Erin Jacobs, Chloe Anderson, Farhana Hossain and Rebecca Unterman. 2017. "An introduction to the world of work: A study of the implementation and impacts of New York City's summer youth employment program.".

Appendices

A MDM Balance Table Iterations

A.1 MDM distance less than or equal to 1.25

Table 9: Balance - Post-Matching - Attendance Days 2015-16 (MDM dist 1.25)

	Difference in Means (Treated - Untreated)			
Covariates	Unmatched	PSM Matched	CEM Matched	MDM Matched
Male	-0.142***	-0.048*	-0.015	-0.025
	(0.015)	(0.025)	(0.023)	(0.020)
Female	0.142***	0.048*	0.015	0.025
	(0.015)	(0.025)	(0.023)	(0.020)
Free or Reduced Lunch	0.125***	0.021	0.019	-0.016
	(0.015)	(0.024)	(0.021)	(0.019)
English Language Learner (ELL)	-0.059***	0.025	0.024*	0.021
	(0.011)	(0.016)	(0.014)	(0.013)
Special Education	-0.064***	0.046***	0.029***	0.067***
	(0.011)	(0.015)	(0.010)	(0.010)
Race: Asian	0.087***	-0.062***	0.002	0.025*
	(0.009)	(0.020)	(0.017)	(0.015)
Race: Black	0.106***	0.045*	0.040*	0.006
	(0.014)	(0.025)	(0.023)	(0.021)
Race: Hispanic	-0.038***	0.043*	0.015	-0.011
	(0.013)	(0.022)	(0.020)	(0.019)
Race: White	-0.158***	-0.038**	-0.070***	-0.058***
	(0.012)	(0.015)	(0.015)	(0.013)
Race: Other	0.002	0.012	0.013**	0.039***
	(0.006)	(0.010)	(0.005)	(0.006)
Grade: 8	-0.051***	0.001	-0.004	-0.002
	(0.007)	(0.005)	(0.005)	(0.005)
Grade: 9	-0.220***	-0.003	-0.024*	-0.036***
	(0.013)	(0.015)	(0.015)	(0.014)
Grade: 10	-0.070***	-0.033	-0.072***	-0.064***
	(0.012)	(0.023)	(0.022)	(0.019)
Grade: 11	0.147***	0.031	0.111***	0.114***
	(0.012)	(0.025)	(0.023)	(0.021)
Grade: 12	0.193***	0.004	-0.011*	-0.012**
	(0.012)	(0.004)	(0.006)	(0.005)
Attendance Days 2014-15	18.050***	-1.047	-0.064	-1.032**
	(1.300)	(0.680)	(0.449)	(0.441)
Observations	23,769	1,574	2,739	3,401

Standard errors in parentheses
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

A.2 MDM distance less than or equal to 1.50 (optimal)

Table 10: Balance - Post-Matching - Attendance Days 2015-16 (MDM dist 1.50)

	Difference in Means (Treated - Untreated)			
Covariates	Unmatched	PSM Matched	CEM Matched	MDM Matched
Male	-0.142***	-0.048*	-0.015	-0.026
	(0.015)	(0.025)	(0.023)	(0.020)
Female	0.142***	0.048*	0.015	0.026
	(0.015)	(0.025)	(0.023)	(0.020)
Free or Reduced Lunch	0.125***	0.021	0.019	-0.020
	(0.015)	(0.024)	(0.021)	(0.019)
English Language Learner (ELL)	-0.059***	0.025	0.024*	0.020
	(0.011)	(0.016)	(0.014)	(0.013)
Special Education	-0.064***	0.046***	0.029***	0.065***
	(0.011)	(0.015)	(0.010)	(0.010)
Race: Asian	0.087***	-0.062***	0.002	0.026*
	(0.009)	(0.020)	(0.017)	(0.015)
Race: Black	0.106***	0.045*	0.040*	0.002
	(0.014)	(0.025)	(0.023)	(0.021)
Race: Hispanic	-0.038***	0.043*	0.015	-0.012
	(0.013)	(0.022)	(0.020)	(0.019)
Race: White	-0.158***	-0.038**	-0.070***	-0.056***
	(0.012)	(0.015)	(0.015)	(0.013)
Race: Other	0.002	0.012	0.013**	0.039***
	(0.006)	(0.010)	(0.005)	(0.006)
Grade: 8	-0.051***	0.001	-0.004	-0.002
	(0.007)	(0.005)	(0.005)	(0.005)
Grade: 9	-0.220***	-0.003	-0.024*	-0.037***
	(0.013)	(0.015)	(0.015)	(0.014)
Grade: 10	-0.070***	-0.033	-0.072***	-0.066***
	(0.012)	(0.023)	(0.022)	(0.019)
Grade: 11	0.147***	0.031	0.111***	0.117***
	(0.012)	(0.025)	(0.023)	(0.021)
Grade: 12	0.193***	0.004	-0.011*	-0.013**
	(0.012)	(0.004)	(0.006)	(0.006)
Attendance Days 2014-15	18.050***	-1.047	-0.064	-0.720
	(1.300)	(0.680)	(0.449)	(0.449)
Observations	23,769	1,574	2,739	3,464

Standard errors in parentheses
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

A.3 MDM distance less than or equal to 1.75

Table 11: Balance - Post-Matching - Attendance Days 2015-16 (MDM dist 1.75)

	Difference in Means (Treated - Untreated)			
Covariates	Unmatched	PSM Matched	CEM Matched	MDM Matched
Male	-0.142***	-0.048*	-0.015	-0.026
	(0.015)	(0.025)	(0.023)	(0.020)
Female	0.142***	0.048*	0.015	0.026
	(0.015)	(0.025)	(0.023)	(0.020)
Free or Reduced Lunch	0.125***	0.021	0.019	-0.022
	(0.015)	(0.024)	(0.021)	(0.019)
English Language Learner (ELL)	-0.059***	0.025	0.024*	0.020
	(0.011)	(0.016)	(0.014)	(0.013)
Special Education	-0.064***	0.046***	0.029***	0.065***
	(0.011)	(0.015)	(0.010)	(0.010)
Race: Asian	0.087***	-0.062***	0.002	0.028*
	(0.009)	(0.020)	(0.017)	(0.015)
Race: Black	0.106***	0.045*	0.040*	0.001
	(0.014)	(0.025)	(0.023)	(0.021)
Race: Hispanic	-0.038***	0.043*	0.015	-0.012
	(0.013)	(0.022)	(0.020)	(0.019)
Race: White	-0.158***	-0.038**	-0.070***	-0.056***
	(0.012)	(0.015)	(0.015)	(0.013)
Race: Other	0.002	0.012	0.013**	0.039***
	(0.006)	(0.010)	(0.005)	(0.006)
Grade: 8	-0.051***	0.001	-0.004	-0.002
	(0.007)	(0.005)	(0.005)	(0.005)
Grade: 9	-0.220***	-0.003	-0.024*	-0.037***
	(0.013)	(0.015)	(0.015)	(0.014)
Grade: 10	-0.070***	-0.033	-0.072***	-0.065***
	(0.012)	(0.023)	(0.022)	(0.019)
Grade: 11	0.147***	0.031	0.111***	0.118***
	(0.012)	(0.025)	(0.023)	(0.021)
Grade: 12	0.193***	0.004	-0.011*	-0.014**
	(0.012)	(0.004)	(0.006)	(0.006)
Attendance Days 2014-15	18.050***	-1.047	-0.064	-0.472
	(1.300)	(0.680)	(0.449)	(0.456)
Observations	23,769	1,574	2,739	3,503

Standard errors in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

B Test Score Results using CEM and PSM

B.1 CEM

Table 12: CEM - MCAS Scores

	(1) Math	(2) ELA	(3) Math Proficient	(4) ELA Proficient	(5) Math Failing	(6) ELA Failing
Treated	3.958***	3.259***	0.117**	0.137***	-0.0283	-0.0656
	(1.108)	(0.665)	(0.0402)	(0.0369)	(0.0355)	(0.0444)
N	1,263	1,272	981	1,084	683	501

Standard errors in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 13: CEM - AP Scores

	(1) Overall	(2) Math/Sci	(3) English/Arts/FL	(4) Social Sci
Treated	0.0612	0.0263	-0.0280	0.196
	(0.0608)	(0.0973)	(0.0941)	(0.223)
N	2,182	702	583	183

Standard errors in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 14: CEM Regressions - SAT Scores

	(1) Overall	(2) Math	(3) Reading	(4) Writing
Treated	44.04***	11.17*	14.61**	18.26***
	(11.98)	(4.973)	(4.838)	(4.695)
N	1,741	1,741	1,741	1,741

Standard errors in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

B.2 PSM

Table 15: PSM - MCAS Scores

	(1) Math	(2) ELA	(3) Math Proficient	(4) ELA Proficient	(5) Math Failing	(6) ELA Failing
Treated	4.401***	3.476***	0.0974***	0.0931***	-0.0318*	-0.0282**
	(1.041)	(0.638)	(0.0283)	(0.0269)	(0.0124)	(0.00958)
N	1,701	1,701	1,756	1,763	1,701	1,701

Standard errors in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 16: PSM - SAT Scores

	(1) Overall	(2) Math	(3) Reading	(4) Writing
Treated	43.26**	12.99*	12.56**	17.72***
	(13.06)	(5.533)	(4.420)	(4.477)
N	2,135	2,135	2,135	2,135

Standard errors in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 17: PSM - AP Scores

	(1) Overall	(2) Math/Sci	(3) English/Arts/FL	(4) Social Sci
Treated	0.0461	-0.0279	-0.114	0.198
	(0.0599)	(0.123)	(0.0965)	(0.0836)
N	2,507	705	562	169

Standard errors in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

C Subgroup Analysis Tables

C.1 2015-16 Attendance

Table 18: CEM Marginal Effects - Attendance Days 2015-16

	N	Main Effect	Attendance (< 90%)	Grades 9 & 10	Male	Black or Hispanic
			366	1,508	1,065	1,902
Treated	601	2.675**	2.138*	3.336*	3.074*	-0.176
		(0.997)	(0.975)	(1.348)	(1.341)	(1.828)
T*(Attendance (< 90%))	75		4.401			
			(4.743)			
T*(Grades 9 & 10)	378			-1.763		
				(2.044)		
T*(Male)	227				-0.999	
					(2.058)	
T*(Black or Hispanic)	443					3.905
						(2.207)
N		2,739	2,739	2,739	2,739	2,739

Standard errors in parentheses
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 19: PSM Marginal Effects - Attendance Days 2015-16

	N	Main Effect	Attendance (< 90%)	Grades 9 & 10	Male	Black or Hispanic
			1,204	2,751	2,350	3,854
Treated	713	3.678*	7.935***	4.062**	4.501**	3.469
		(1.367)	(1.533)	(1.460)	(1.486)	(2.138)
T*(Attendance (< 90%))	108		-28.11***			
			(3.841)			
T*(Grades 9 & 10)	258			-1.060		
				(2.024)		
T*(Male)	265				-2.215	
					(1.772)	
T*(Black or Hispanic)	506					0.295
						(2.649)
N		5,348	5,348	5,348	5,348	5,348

Standard errors in parentheses
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

C.2 2015-16 Truancy

Table 20: CEM Marginal Effects - Truant Days 2015-16

		Main Effect	Attendance ($< 90\%$)	Grades 9 & 10	Male	Black or Hispanic
	N		293	1,605	1,126	2,087
Treated	609	-0.707	-0.670	-0.547	-0.497	0.804
		(0.463)	(0.378)	(0.525)	(0.551)	(0.658)
T*(Attendance ($< 90\%$))	63		-0.287			
			(2.799)			
T*(Grades 9 & 10)	377			-0.414		
				(1.023)		
T*(Male)	223				-0.558	
					(0.997)	
T*(Black or Hispanic)	445					-2.061*
						(0.883)
N		2,986	2,986	2,986	2,986	2,986

Standard errors in parentheses
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 21: PSM Marginal Effects - Truant Days 2015-16

		Main Effect	Attendance ($< 90\%$)	Grades 9 & 10	Male	Black or Hispanic
	N		1,241	2,667	2,279	3,862
Treated	713	-1.477	-4.500***	-1.789	-1.981*	-2.055
		(0.893)	(0.902)	(1.066)	(0.870)	(1.910)
T*(Attendance ($< 90\%$))	110		19.60***			
			(3.127)			
T*(Grades 9 & 10)	259			0.859		
				(1.613)		
T*(Male)	267				1.346	
					(1.607)	
T*(Black or Hispanic)	506					0.814
						(2.318)
N		5,300	5,300	5,300	5,300	5,300

Standard errors in parentheses
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

C.3 2015-16 GPA

Table 22: MDM Marginal Effects - Weighted Mean Overall GPA 2015-16 (increments of .5)

	N	Main Effect	Attendance (< 90%)	Grades 9 & 10	Male	Black or Hispanic
N			76	40	173	383
Treated	281	0.0854	0.146	0.0417	0.153	-0.292
		(0.0989)	(0.140)	(0.384)	(0.173)	(0.303)
T*(Attendance (< 90%))	36		-0.776*			
			(0.326)			
T*(Grades 9 & 10)	252			-0.000680		
				(0.407)		
T*(Male)	96				-0.307	
					(0.269)	
T*(Black or Hispanic)	225					0.404
						(0.333)
N		1,475	471	471	471	471

Standard errors in parentheses
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 23: CEM Marginal Effects - Weighted Mean Overall GPA 2015-16 (increments of .5)

	N	Main Effect	Attendance (< 90%)	Grades 9 & 10	Male	Black or Hispanic
N			97	53	206	467
Treated	156	0.0618	0.168	-0.371	0.130	0.155
		(0.137)	(0.136)	(0.511)	(0.170)	(0.256)
T*(Attendance (< 90%))	22		-0.770			
			(0.429)			
T*(Grades 9 & 10)	10			0.468		
				(0.528)		
T*(Male)	55				-0.187	
					(0.290)	
T*(Black or Hispanic)	133					-0.108
						(0.299)
N		563	563	563	563	563

Standard errors in parentheses
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 24: PSM Marginal Effects - Weighted Mean Overall GPA 2015-16 (increments of .5)

	N	Main Effect	Attendance (< 90%)	Grades 9 & 10	Male	Black or Hispanic
N			509	205	996	1,766
Treated	251	0.166	0.504**	-1.513*	0.510**	0.0589
		(0.158)	(0.143)	(0.580)	(0.152)	(0.333)
T*(Attendance (< 90%))	35		-2.424***			
			(0.435)			
T*(Grades 9 & 10)	223			1.890***		
				(0.508)		
T*(Male)	95				-0.908***	
					(0.213)	
T*(Black or Hispanic)	196					0.137
						(0.407)
N		2,371	2,371	2,371	2,371	2,371

Standard errors in parentheses
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

C.4 2015-16 Course Failures

Table 25: MDM Marginal Effects - Course Failures 2015-16

	N	Main Effect	Attendance (< 90%)	Grades 9 & 10	Male	Black or Hispanic
N			78	48	169	375
Treated	283	-0.177	-0.516**	0.0407	-0.385	-0.503
		(0.104)	(0.169)	(0.713)	(0.222)	(0.467)
T*(Attendance (< 90%))	37		1.504***			
			(0.430)			
T*(Grades 9 & 10)	252			-0.430		
				(0.742)		
T*(Male)	97				0.127	
					(0.371)	
T*(Black or Hispanic)	226					0.198
						(0.512)
N		1,599	465	465	465	465

Standard errors in parentheses
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 26: CEM Marginal Effects - Course Failures 2015-16

	N	Main Effect	Attendance (< 90%)	Grades 9 & 10	Male	Black or Hispanic
			155	61	363	832
Treated	218	-0.215	-0.363**	-0.365	-0.151	-0.326
		(0.128)	(0.120)	(0.620)	(0.156)	(0.285)
T*(Attendance (< 90%))	25		1.314**			
			(0.507)			
T*(Grades 9 & 10)	11			0.161		
				(0.635)		
T*(Male)	79				-0.177	
					(0.285)	
T*(Black or Hispanic)	186					0.131
						(0.325)
N		1,080	1,080	1,080	1,080	1,080

Standard errors in parentheses
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 27: PSM Marginal Effects - Course Failures 2015-16

	N	Main Effect	Attendance (< 90%)	Grades 9 & 10	Male	Black or Hispanic
			535	171	1,032	1,935
Treated	278	-0.106	-0.348***	1.316*	-0.281*	0.142
		(0.113)	(0.0912)	(0.589)	(0.129)	(0.256)
T*(Attendance (< 90%))	37		1.820**			
			(0.603)			
T*(Grades 9 & 10)	248			-1.594**		
				(0.584)		
T*(Male)	97				0.502**	
					(0.183)	
T*(Black or Hispanic)	221					-0.312
						(0.305)
N		2,495	2,495	2,495	2,495	2,495

Standard errors in parentheses
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

C.5 2015-16 Graduation

Table 28: CEM Marginal Effects - Probability of Graduating on Time

	N	Main Effect	Attendance (< 90%)	Grades 9 & 10	Male	Black or Hispanic
			662	2,997	1,674	2,885
Treated	1,015	0.0856***	0.0466***	0.0397***	0.0694***	-0.000974
		(0.0173)	(0.00889)	(0.00906)	(0.0107)	(0.0109)
T*(Attendance (< 90%))	162		0.0565*			
			(0.0279)			
T*(Grades 9 & 10)	791			0.0745**		
				(0.0267)		
T*(Male)	384				-0.0362	
					(0.0191)	
T*(Black or Hispanic)	718					0.0801***
						(0.0161)
N		2,761	4,246	4,246	4,246	4,246

Standard errors in parentheses
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 29: PSM Marginal Effects - Probability of Graduating on Time

	N	Main Effect	Attendance (< 90%)	Grades 9 & 10	Male	Black or Hispanic
			2,118	5,854	3,846	6,059
Treated	1,221	0.122***	0.170***	0.128***	0.192***	0.153***
		(0.0260)	(0.0312)	(0.0256)	(0.0433)	(0.0391)
T*(Attendance (< 90%))	220		-0.181***			
			(0.0314)			
T*(Grades 9 & 10)	260			-0.0148		
				(0.0344)		
T*(Male)	468				-0.140***	
					(0.0416)	
T*(Black or Hispanic)	840					-0.0421
						(0.0508)
N		5,948	5,948	5,948	5,948	5,948

Standard errors in parentheses
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$