Estimating Impacts on Program-Related Subgroups using Propensity Score Matching: Evidence from the Early College High School Study

Society for Research on Educational Excellence
Proposal for Paper Presentation in Methodology Section

Fatih Unlu, Abt Associates*
Ryoko Yamaguchi, Abt Associates
Larry Bernstein, RTI
Julie Edmunds, SERVE Center

Note: * Fatih Unlu serves as the contact for this paper. Please contact fatih_unlu@abtassoc.com for inquiries.

Background

This paper addresses methodological issues arising from an experimental study of North Carolina's Early College High School Initiative, a four-year longitudinal experimental study funded by Institute for Education Sciences. North Carolina implemented the Early College High School Initiative in response to low high school graduation rates. The goal of the initiative is to increase the number of students graduating from high school and who continue on and succeed in college. The study has three main goals: (1) Determine the impact of the model on selected student outcomes, including course-taking patterns, achievement, attitudes, and dropout and leaving rates; (2) Determine the extent to which outcomes differ by student characteristics; and (3) Examine the implementation of the model and the extent to which specific model components are associated with positive outcomes.

Schools participating in the study identify an eligible pool of student applicants. The research team then randomly assigns students to either the treatment group (attending the ECHS) or the control group (business as usual). The outcomes for students in the two groups are then tracked and compared. The study follows an intent-to-treat model in that once a student is assigned to ECHS, he or she remains in the treatment group regardless of whether he or she ends up enrolling in ECHS, or leaves ECHS.

In conducting early analyses on outcomes, which have been reported elsewhere (Edmunds et al., 2009), the research team found an impact on students' course-taking patterns. In particular, the research team found that ECHS significantly increased the percentage of students taking college preparatory mathematics courses, including Algebra I, Geometry, and Algebra II. **Table 1** reports differences in the percentage of students taking these college preparatory courses by treatment and control groups.

Given the larger proportion of treatment students taking college preparatory mathematics classes, a simple comparison of test scores between the two groups is no longer appropriate since the treatment and control groups are no longer comparable. The problem thus becomes a case of endogenous or program-related subgroups since the composition of the subgroup of interest (e.g., 9th grade Algebra I takers) is affected by the program (Schochet & Burghardt, 2007). This paper is designed to explore a possible approach to addressing this issue.

Purpose and Research Questions of the Study

The purpose of this paper is to investigate the methodological considerations of estimating impacts, particularly for program-related subgroups that are affected by issues of endogeneity. For reasons of sample size, we have chosen to focus on those students taking Algebra I in 9th grade although the proposed analyses could eventually be utilized with any subject in which course-taking patterns are different between the treatment and control students. We use the framework by Angrist, Imbens, & Rubin (1996) and Gennetian, Morris, & Bloom (2005) to categorize students who take Algebra I as: 1) Never takers; 2) Always-takers; 3) Defiers; and 4) Compliers. As **Figure 1** shows, "never takers" are students who never take Algebra I, whether they are in the treatment (ECHS) or not. "Always-takers" are students who, regardless of whether they are in ECHS, take Algebra I. "Compliers" are those who take Algebra I only if they are in ECHS; or in essence, they comply to the requirements of ECHS and take Algebra I. "Compliers" is an important subgroup since they are induced into taking Algebra I in 9th grade by the program and this may effect them adversely. "Defiers" are students who would not take Algebra I if they were assigned to ECHS but would take it if they were assigned to the control group. For the purposes of this paper, we assume that there are no "Defiers".

Following this framework, we investigate methods to estimate overall and program-related impacts of ECHS. Our research questions are as follows:

1. What is the overall effect of ECHS on passing Algebra I? This question focuses on the overall treatment effect, or the treatment effect of all three groups together: “Compliers”, “Always-takers”, and “Never takers”. This question is the focal point of the overall evaluation, and can easily be answered through the experimental design. Thus, it is not the main focus of this paper.
2. Of the group who take Algebra I in 9th grade, what is the effect of ECHS on passing Algebra I? This question focuses on estimating the treatment effect of “Compliers” and “Always-takers.”
3. Of the group who would have taken Algebra I in 9th grade regardless of ECHS, what is the effect of ECHS on passing Algebra I? This question focuses only on the “Always-takers.”

Note that ECHS can affect passing Algebra I through at least two pathways. First, it can induce ECHS students to take Algebra I and hence raise their pass-rates (Compliers). Second, it can directly affect passing Algebra I through better teaching at ECHS. The first two research questions take both pathways into account. The third question, on the other hand, isolates the latter pathway by employing Always-takers since they would not be induced by ECHS into taking Algebra I by definition. Hence, by comparing the combined effect on the Always-takers and Compliers, to the one on only the Always takers, we can get an idea of whether Compliers are adversely or positively affected by ECHS.[1] The methodological challenge is that Always-takers (in the ECHS group) are not observed and can only be identified using propensity score matching.

Setting/Subjects

This paper uses data from six ECHS sites in North Carolina, all of which used random assignment to identify students. The sample used to estimate the results reported here is composed of 706 ninth grade students randomly assigned to the ECHS or control group (412 treatment and 294 control) in 6 sites between 2006 and 2008. **Table 2** presents the baseline characteristics of the full sample as a whole and broken down by treatment status and Algebra I taking. There were no statistically significant differences in baseline characteristics between the treatment and control group. Four characteristics were statistically significantly different between students who took Algebra I and those who did not (first generation college bound, disability, and 8th grade math and reading scores).

Intervention: Early College High School and the Algebra I Subgroup

Early College High Schools are small autonomous high schools, located on the campuses of community colleges or universities. Targeted at students who are underrepresented in college, these schools are designed to provide students with a high school diploma and two years of transferable college credit in four or five years. A core component of the ECHS model is placing all students on a college preparatory track of study, which, in mathematics, includes Algebra I or

[1] The direct examination of whether Compliers are adversely affected by ECHS requires knowing what the outcomes of the control group Compliers would be if they were assigned to ECHS and hence took up Algebra I.

a higher math in 9^{th} grade. Completing Algebra I by the end of 9^{th} grade is essentially required if a student is to be considered on track-for college. Although Algebra I is required for graduation in North Carolina, there is no requirement that it be taken by the end of 9^{th} grade. Thus, Algebra I (and higher mathematics) course-taking can be considered a particularly sensitive indicator for the impact of the ECHS.

Research Design

Using data from the study briefly described in the introduction, this paper address the three research questions presented above to examine the effect of ECHS on passing Algebra I. The first two research questions are addressed within the experimental framework whereas for the third research, we employ a quasi-experimental method, propensity score matching, to identify the Always-takers in the ECHS group.

Data Collection and Analysis

Our analyses are based on administrative data, collected by the North Carolina Department of Public Instruction (NCDPI), and merged and de-identified by the North Carolina Education Research Center (NCERDC) at Duke University and include students' demographic and socio-economic characteristics, course-taking patterns, and results of end-of course examinations.

To address the three research questions stated above, we break the treatment and control students into three groups by Algebra I course-taking: A-A', B-B', and C-C'. As seen in **Figure 2**, students in groups A and A' would take Algebra I regardless of school assignment (Always-takers). Groups B and B' represents students who would take Algebra I only if they participated in ECHS (Compliers). Finally, C and C' denote the students who would not take Algebra I regardless of the treatment assignment (Never takers). Note that only the distinction between algebra-takers and non-takers (A'+B' vs. C' and A vs. B+C) is observed. In order to explain our approach, we use the following notation:

N(G) = number of students in group G (G= A, A', B, B', C, and C'). Sometimes it is convenient to combine groups; for example, the total in groups A and B is N(A+B).

P(G)= proportion of students in each subgroup. For example, P(A) = N(A)/N(A+B+C)

R(G) = number of students who passed Algebra I in group G.

X(G) = proportion of students of students who passed Algebra I in group G. X(A), for example, equals R(A)/N(A).

<u>Addressing RQ 1: Average Treatment Effect</u>. To address the first research question, we estimated the effect of ECHS on every student regardless of whether he or she took Algebra I. This effect is called the average treatment effect (ATE). In this estimation, we assumed that students who had not taken Algebra I could not have passed Algebra I if they had been tested (i.e. R(C') = R(B) = R(C) = 0). ATE is then the difference between the pass-rates in the whole treatment and control groups:

$$(1)\quad ATE = \frac{R(A'+B'+C')}{N(A'+B'+C')} - \frac{R(A+B+C)}{N(A+B+C)} = \frac{R(A'+B')}{N(A'+B'+C')} - \frac{R(A)}{N(A+B+C)}$$

Variance of this estimate is derived in the appendix and it equals:

$$(2)\ Var(ATE) = \left[\frac{N(A'+B')}{N(A'+B'+C')}\right]^2 \frac{X(A'+B')[1-X(A'+B')]}{N(A'+B')} + \left[\frac{N(A)}{N(A+B+C)}\right]^2 \frac{X(A)[1-X(A)]}{N(A)}$$

Addressing RQ 2: Average Treatment Effect on the Treated (ATT). To address the second research question, we focus on the effect of ECHS on students who actually took Algebra I, hence never-takers (group C and C') are not used in this analysis. In particular, we use a Bloom-type adjustment for no-shows by adjusting the ATE by the Algebra I take-up rate in the ECHS group, which relies on the assumption that students in group B would not have passed Algebra I. Under this assumption, ATT and its variance are:

$$(3)\quad ATT = \frac{ATE}{P(A'+B')} \text{ and } Var(ATT) = \frac{Var(ATE)}{[P(A'+B)']^2}$$

Addressing RQ 3: Average Treatment Effect on the Always-takers (ATA). To address the third research question, we estimate the ECHS effect on the Always-takers, which is challenging as Always-takers in the treatment group are not observed. One way to address this issue is to match each control student who took Algebra I (Always-takers since there are no Defiers) with a similar treatment student who also took the course via propensity scoring.2 We identified the Always-takers in the treatment group empirically through the following steps:

1. Modeling Algebra I taking: In this step, we developed a logistic regression model to predict the probability of Algebra I taking. This model employs the baseline characteristics in Table 1 and is estimated using only the control students (Group A, B, and C) in order to eliminate the effect of ECHS on Algebra I taking.
2. Estimating Propensity Scores: Using the estimated model in Step 1, we then predicted propensity scores for all students who took Algebra I (groups A', B', and A). Other students' propensity scores were not calculated as they were not part of the matching.
3. Propensity Score Matching: We implemented one-to-one matching with replacement to match each Algebra I taker in the control group with the most similar (i.e., closest propensity score) Algebra I taker in the treatment group.3 We conducted the matching separately within each cohort of students in each site.
4. Checking the quality of matches: We tested whether the matching characteristics were balanced across matched control and treatment students using t-tests (Dehejia & Wahba, 2002) as well as standardized differences to supplement the t-tests which are sensitive to sample sizes (Morgan & Harding, 2006; Morgan & Winship, 2007).

Let A'' denote the Always-takers in the treatment group (matched treatment students) and W_i'' and W_i indicate whether treatment and control members of the ith matched pair passed Algebra I. Then, the ATA and its variance can be calculated as follows:

[2] Several recent studies used propensity score matching when estimating impacts for program-related subgroups (Peck, 2003; Schochet & Burghardt, 2007).

[3] There are various matching algorithms such as one-to-one and one-to-many matching, interval matching, and kernel matching (Heckman, Ichimura, & Todd, 1997; Morgan & Harding, 2006; Caliendo & Kopeinig, 2008). For simplicity and illustrative purposes we implemented one-to-one matching with replacement which allows the potential matches (here treatment group Algebra I takers) to be used in the matching more than once.

(4) $$ATA = \frac{\sum_i (W_i^{"} - W_i)}{N(A)} = \frac{\sum_i Z_i}{N(A)} \text{ and } Var(ATA) = \frac{\sum_i (Z_i - ATA)^2}{N(A)}$$

Results

Table 3 presents the number of ECHS and control students taking and passing Algebra I. Using these students, the calculated *ATE* was 0.18 (standard error = 0.06). Plugging these estimates into Equation 3, the *ATT* was 0.19 (standard error = 0.07).

We then proceeded with the identification of the Always-takers. **Table 4** exhibits the estimated model of Algebra I taking in the control group.[4] Being first generation college bound, having a disability, and eighth grade math test scores appear to be strong predictors of Algebra I taking. Using this estimated model, we then calculated propensity scores of Algebra I takers, a histogram of which are presented in **Figure 3**. **Figure 3** shows that the distribution of the propensity scores in the treatment and control group is somewhat different. Using the predicted propensity scores, we then matched each control Algebra I taker with a treatment Algebra I taker. **Figures 4 and 5** presents the histogram of the propensity scores of the control and matched treatment groups and the propensity scores of matched pairs of control and treatment students, which suggest that the matching process worked quite well.

Table 5 presents the results from more rigorous checks of the quality of these matches, showing that none of the treatment vs. control differences of the matching characteristics, including eight grade math test scores (which was significant before matching) were statistically significant at $p<0.05$ after matching. In addition, the standardized differences, presented in **Table 6,** show that none of these are larger than 0.15, further suggesting that matching was successful. Using the identified Always-takers and Equation 4, we estimated *ATA* to be 0.01 (standard error = 0.45).

Discussion

This paper focuses on estimating impacts of program-related subgroups (Schochet & Burghardt, 2007) using Angrist, Imbens, & Rubin (1996)'s framework of making causal inferences. As noted, ECHS can affect passing Algebra 1 through at least two pathways: inducing students to take Algebra I and through better teaching. *ATE* and *ATT* take both pathways into account whereas *ATA* pertains to the second pathway only. The three treatment effect estimates vary to a certain degree: *ATE* is 0.18, *ATT* is 0.19, and most notably *ATA* is 0.01. These estimates suggest that ECHS has the same impact on pass-rates for the Always-takers as the traditional high school. Where the ECHS are making difference is by increasing the number of students who are taking these courses. Our findings further justify our use of these three different methods to study the effect of ECHS on Algebra I passing.

[4] This model also include second and third powers of eight grade test scores and their interaction with the African-American indicator variable since inclusion of these terms was found to improve the balance of the matching characteristics across the matched groups by trial and error.

Appendix A: References

Adelman, C. (1999). Answers in the Toolbox: Academic Intensity, Attendance Patterns, and Bachelor's Degree Attainment (Vol. PLLI 1999-8021). Washington D.C.: U.S. Government Printing Office.

Angrist, J., Imbens, G., & Rubin, D. (1996). Identification of causal effects using instrumental variables. *Journal of the American Statistical Association, 91*(434), 444-455.

Caliendo, M., & Kopeining, S. (2008). Some practical guidance for the implementation of propensity score matching. *Journal of Economic Surveys, 22*(1), 31-72.

Choy, S. P., Horn, L. J., Nunez, A.-M., & Chen, X. (2000). Transition to college: What helps at-risk students and students whose parents did not attend college. *New Directions for Institutional Research, 107*, 45-63.

Dehejia, R., & Wahba, S. (2002). Propensity score matching methods for non-experimental causal studies. *Review of Economics and Statistics, 84*, 151-161.

Edmunds, J., Arshavsky, N., Bernstein, L., Unlu, F., Glennie, E., Willse, J., and Dallas, A. (2009). The Study of the Efficacy of North Carolina’s Learn and Earn Early College High School Model – Summary of Early Results. Poster presented at the Annual Research Conference of the Institute of Education Sciences, U.S. Department of Education, Washington, D.C.

Gennetian, L., Morris, J., Bos, J., & Bloom, H. (2005). Coupling the nonexperimental technique of instrumental variables with experimental data to learn how programs create impacts. In H. Bloom (Ed.), *Moving to the Next Level: Combining Experimental and Non-Experimental Methods to Advance Employment Policy Research*. New York: Russell Sage Foundation.

Greene, J. P. (2002). High School Graduation Rates in the United States. New York: Manhattan Institute for Policy Research.

Heckman, J. J., Ichimura, H., & Todd, P. (1997). Matching as an econometric evaluation estimator: Evidence from evaluation a job training programme. *Review of Economic Studies, 64*, 605-654.

Horn, L., & Nunez, A. (2000). Mapping the Road to College: First-Generation Students' Math Track, Planning Strategies, and Context of Support (Vol. NCES 2000-153). Washington D.C.: U.S. Government Printing Office.

House, J. D. (1993). The relationship between academic self-concept and school withdrawal. *Journal of Social Psychology, 133*, 125-127.

Lan, W., & Lanthier, R. (2003). Changes in students' academic performance and perceptions of school and self before dropping out of school. *Journal of Education for Students Placed at Risk, 8*(3), 309-332.

Lee, V. E., & Burkam, D. T. (2003). Dropping out of high school: The role of school organization and structure. *American Educational Research Journal, 40*(2), 353-393.

Morgan, S. L., & Harding, D. J. (2006). Matching estimators of casual effects: Prospects and pitfalls in theory and practice. *Sociological Methods and Research, 35*, 3-60.

Morgan, S. L., & Winship, C. (2007). *Counterfactuals and causal inference: Methods and principles in social research*. New York: Cambridge University Press.

North Carolina New Schools Project. (December, 2007). Design Principals for High School Innovation Projects.

Peck, L. R. (2003). Subgroup analysis in social experiments: Measuring program impacts based on post-treatment choice. *American Journal of Evaluation, 2*(24), 157-187.

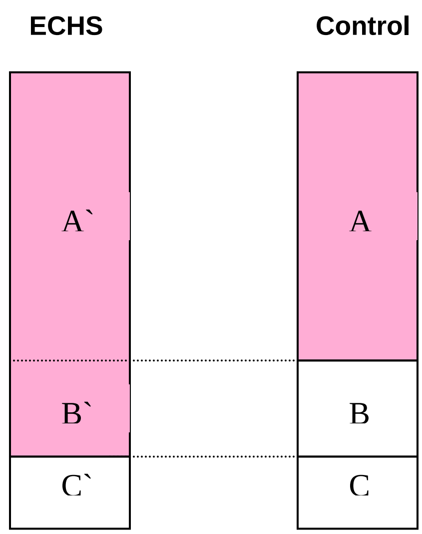
Schochet, P. Z., & Burghardt, J. (2007). Using propensity scoring to estimate program-related subgroup impacts in experimental program evaluations. *Evaluation Review, 31*, 95-120.

Appendix B: Tables and Figures

Figure 1: Methodological Framework for Estimating Program-related Subgroups

		D_i (0) = Control	
D_i (1) = ECHS	Algebra I Taking	0	1
	0	Never Takers (Does not take Algebra I, regardless of ECHS)	Defiers (Takes Algebra I only if assigned to Control)
	1	Compliers (Takes Algebra I only if assigned to ECHS)	Always Takers (Takes Algebra I regardless of ECHS)

Figure 2: Distribution of Students by ECHS/Control Status and Algebra I Course-Taking



Key:

A'/A Always-takers: Always take Algebra I (even in the absence of ECHS)

B'/B Compliers: Take Algebra I only if in the ECHS group

C'/C Never-takers: Never take Algebra I (even if assigned to the ECHS group)

Figure 3: Propensity Score Before Matching

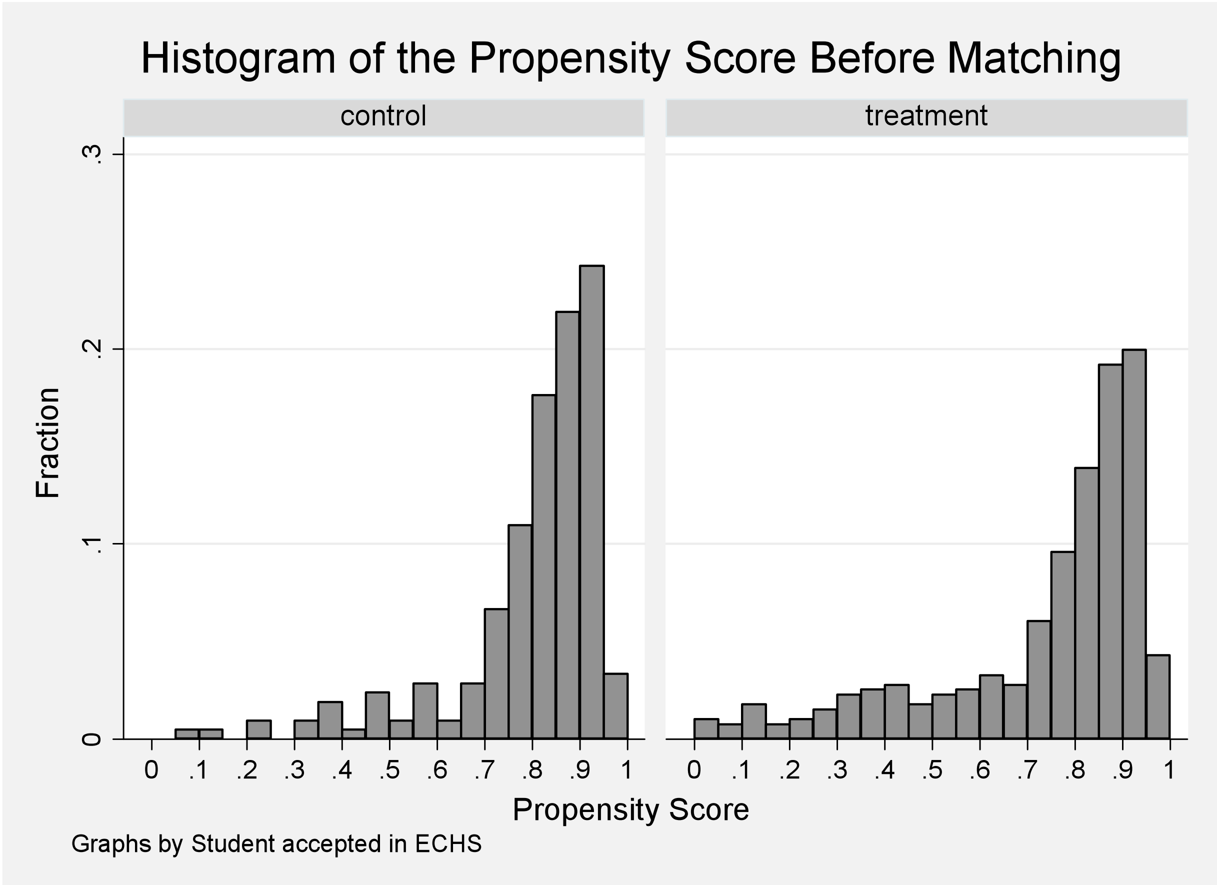


Figure 4: Propensity Score After Matching

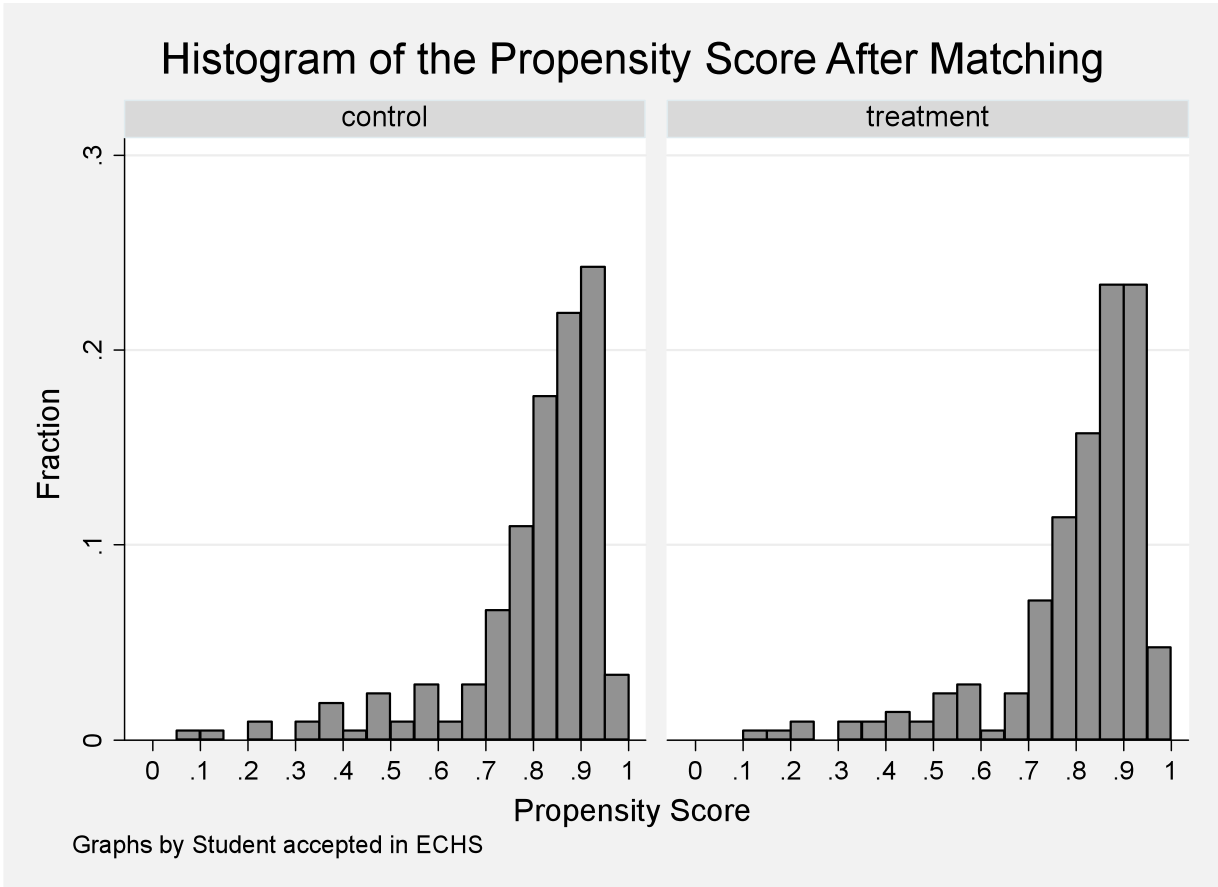


Figure 5: Propensity Score of Matched Pairs

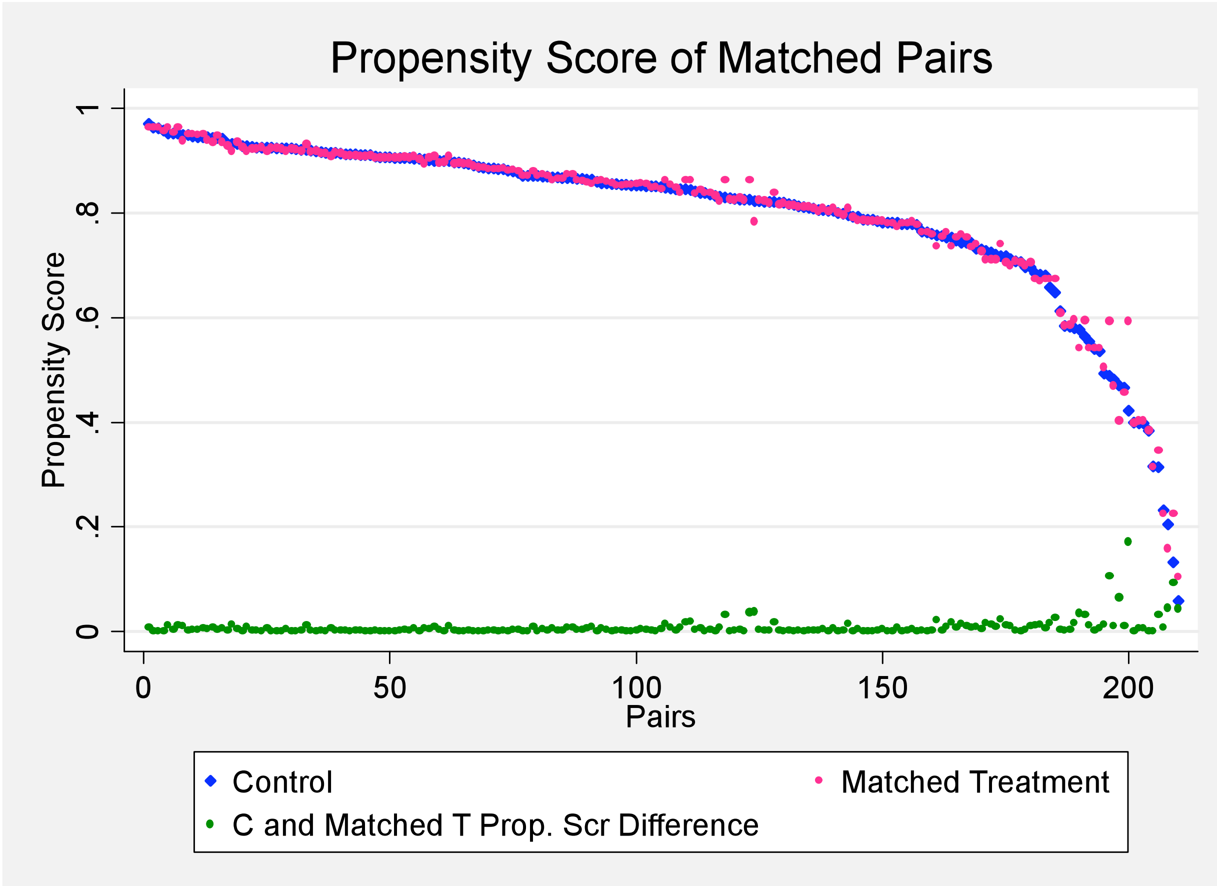


Table 1: 9th Grade Math Course Take-Up Rates

	Whole Sample	ECHS Group	Control Group	ECHS vs. Ctrl. Difference	P-Value
Algebra I	85.84%	96.12%	71.43%	24.69%	<0.001*
Algebra II	7.37%	11.41%	1.70%	9.71%	<0.001*
Geometry	27.05%	30.10%	22.79%	7.31%	0.031*

Notes: Statistically significant differences (at the $p<0.05$ level) are denoted by *.

Table 2: Descriptive Statistics of ECHS sample overall and by Algebra I subgroup

	Overall (N=706)		Treatment (N=412)		Control (N=294)		Algebra Takers (N=606)		Algebra Non-Takers (N=100)	
Variable	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD
% African American	21.67	41.23	21.60	41.20	21.77	41.34	21.29	40.97	24.00	42.92
% Hispanic	5.52	22.86	5.83	23.45	5.10	22.04	5.94	23.66	3.00	17.14
% Male	38.30	48.61	37.96	48.53	38.78	48.81	37.85	48.50	41.00	49.43
%First Generation College Bound	45.01	49.25	43.57	49.03	47.02	49.57	42.13*	48.83	62.45	48.41
% Free/Reduced Priced Lunch Elig.	44.01	49.22	43.65	48.99	44.52	49.61	43.30	49.18	48.32	49.47
% Disabled	3.81	18.83	3.70	18.75	3.96	18.99	3.15*	17.44	7.76	25.48
% Gifted	12.05	32.37	12.04	32.38	12.07	32.41	12.93	33.50	6.72	23.86
8th grade Math Score	0.00	0.98	0.00	0.97	0.00	0.99	0.07*	0.94	-0.41	1.09
8th grade Reading Score	0.00	0.98	0.03	0.97	-0.05	0.99	0.05*	0.94	-0.32	1.16

Notes: * denotes characteristics that are statistically significant between the two groups at the $p<0.05$ level. 8th grade math and reading scores are standardized so that mean = 0, SD = 1.

Table 3: Number of Students Passing Algebra by Treatment Subgroups

Group	Number of Students (N(G))	Number of Students Who Passed Algebra (X(G))
A' and B'	396	329
C'	16	0
Treatment= A' + B' + C'	412	329
A	210	183
B and C	84	0
Control = A + B + C	294	183

Table 4: Logistic Regression Modeling Algebra I Taking in the Control Group

Variable	Odds Ratio	Std. Err.	Z	P|z|	95% Confidence Interval	
African American	1.187	0.703	0.29	0.772	0.372	3.789
Hispanic	2.684	2.153	1.23	0.219	0.557	12.929
Male	0.784	0.252	-0.76	0.447	0.418	1.470
First Generation College Bound	0.480	0.161	-2.19	0.028	0.249	0.925
Free/Reduced Priced Lunch Eligible	1.152	0.404	0.40	0.686	0.580	2.291
Disabled	0.132	0.095	-2.81	0.005	0.032	0.542
Gifted	1.667	1.143	0.75	0.456	0.435	6.390
8^{th} grade Math Score	2.649	0.939	2.75	0.006	1.323	5.307
8^{th} grade Reading Score	0.853	0.180	-0.75	0.452	0.564	1.290
Covariates Imputed	0.101	0.055	-4.25	0.000	0.035	0.291
8^{th} gr. Math Score Square	0.645	0.104	-2.73	0.006	0.471	0.884
8^{th} gr. Math Score Third Power	0.968	0.092	-0.34	0.734	0.804	1.166
8^{th} grade Math Score* African American	0.900	0.916	-0.10	0.918	0.123	6.614
8^{th} gr. Math Score Square * African American	1.151	1.674	0.10	0.923	0.067	19.915
8^{th} gr. Math Score Third Power * African American	1.154	0.853	0.19	0.846	0.271	4.912

Notes: 8^{th} grade math and reading scores are standardized so that mean = 0, SD = 1.

Table 5: Balance of the Control and Treatment Groups Before and After Matching

		Mean				t-test	
Variable	Sample	Control	Treatment	% Difference	% Reduct Difference	T	p>|t|
% African American	Unmatched	20	21.97	-4.8		-0.56	0.574
	Matched	20	25.71	-14.0	-190.1	-1.39	0.164
% Hispanic	Unmatched	5.71	6.06	-1.5		-0.17	0.864
	Matched	5.71	2.86	12.1	-725.0	1.45	0.149
% Male	Unmatched	36.67	38.48	-3.7		-0.44	0.662
	Matched	36.67	40	-6.9	-83.8	-0.70	0.484
%First Generation College Bound	Unmatched	40.38	.43.06	-5.5		-0.64	0.522
	Matched	40.38	43.5	-6.4	-16.6	-0.65	0.515
% Free/ Reduced Priced Lunch Eligible	Unmatched	41.64	44.18	-5.2		-0.61	0.545
	Matched	41.64	46.88	-10.6	-105.9	-1.08	0.280
% Disabled	Unmatched	1.92	3.81	-11.3		-1.27	0.206
	Matched	1.92	0.95	5.8	48.5	0.84	0.403
% Gifted	Unmatched	14.82	11.93	8.5		1.01	0.313
	Matched	14.82	13.81	3.0	65.1	0.30	0.768
8^{th} grade Math Score	Unmatched	0.19	0.001	20.3		2.35	0.019
	Matched	0.19	0.17	2.6	87.0	0.29	0.770
8^{th} grade Reading Score	Unmatched	0.1	0.03	7.3		0.85	0.395
	Matched	0.1	0.15	-6.2	15.4	-0.66	0.510

Notes: 8^{th} grade math and reading scores are standardized so that mean = 0, SD = 1.

Table 6: Balance in Matching Characteristics

Variable	Control Mean	Treatment Mean	|Difference|	Standardized Difference
% African American	20.00	25.71	5.71	0.14
% Hispanic	5.71	2.86	2.86	0.14
% Male	36.67	40.00	3.33	0.07
%First Generation College Bound	40.38	43.50	3.12	0.06
% Free/Reduced Priced Lunch Eligible	41.64	46.88	5.24	0.11
% Disabled	1.92	0.95	0.97	0.08
% Gifted	14.82	13.81	1.01	0.03
8^{th} grade Math Score	0.19	0.17	0.02	0.03
8^{th} grade Reading Score	0.10	0.15	0.06	0.06

Notes: 8^{th} grade math and reading scores are standardized so that mean = 0, SD = 1.