

USING PROPENSITY-SCORE MATCHING IN PROGRAM EVALUATION: LESSONS LEARNED IN A STATEWIDE EVALUATION OF A COLLEGE-PREP PROGRAM IN HAWAII

Malkeet Singh, Michelle Hodara, and Changhua Wang
Education Northwest



INTRODUCTION

In 2012, the Hawaii Board of Education adopted a policy that, beginning with the graduating class of 2016, all high school graduates must complete specified rigorous academic courses and work towards earning the Hawaii Board of Education Recognition Diploma (BOERD). This is known as the STEP UP program. Prior to this it was based on a voluntary basis, as students were given the opportunity at the end of their 8th or at the beginning of their 9th grade to pledge to work towards the BOERD. Students who had made the pledge to commit to the BOERD requirements were provided with college and career resources.

RESEARCH QUESTION

1. What is the relationship between students' participation in the Step Up program and their tenth grade performance on the Hawaii state assessment for math or reading?

POPULATION

The treatment group included Step Up scholars in three graduating classes: 2013, 2014, and 2015. The comparison group comprised the students who did not participate in the Step Up scholars program in the same graduating cohorts. Sample size was 7,601; 7,736; and 7,538 for the three cohorts, respectively.

RESULTS

To answer the overarching question, we calculated two estimates of the program's effects that complement each other. The average treatment effect on the treated (ATT) and the average treatment effect (ATE). The ATE is an estimate of the effect on the whole sample, while ATT is the effect that generalizes to the STEP UP scholars only. We used a list of variables to compute propensity scores (a probability score) for enrolling in the program.

Figure 1: Boxplots showing the spread of the estimated propensity scores in the treatment and comparison group

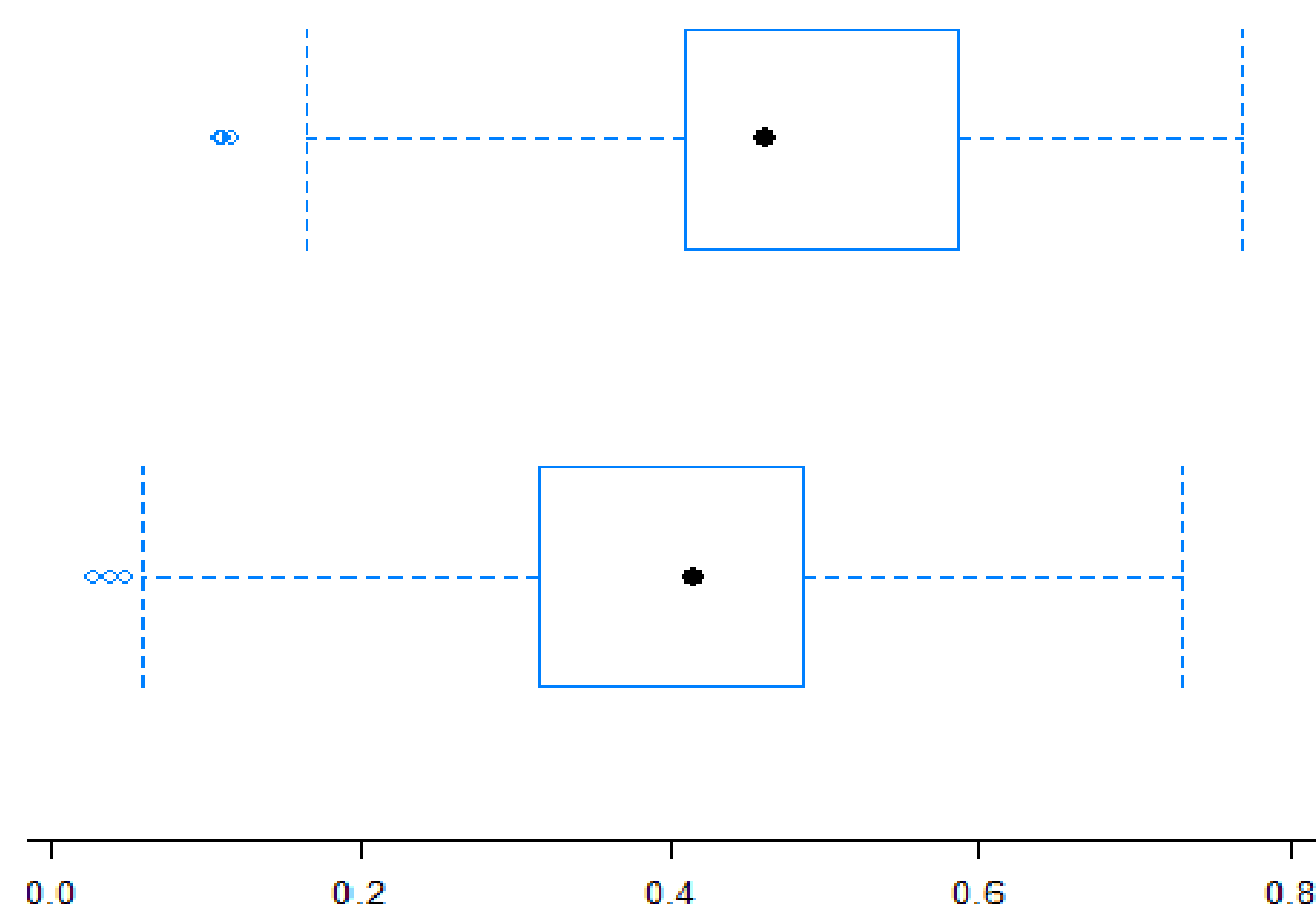


Table 1. Unweighted and weighted means of pretreatment variables show that differences in the pretreatment variables were reduced to almost zero

	Unweighted means			Weighted means		
	Treatment	Control	Effect size	Treatment	Control	Effect size
SES	0.433	0.492	0.118	0.465	0.466	0.002
ELL	0.110	0.106	0.013	0.105	0.107	0.007
SPED	0.060	0.112	0.183	0.086	0.089	0.008
Ethnicity A	0.004	0.005	0.013	0.004	0.004	0.002
Ethnicity B	0.009	0.010	0.009	0.010	0.010	0.001
Ethnicity C	0.053	0.023	0.132	0.036	0.036	0.000
Ethnicity D	0.303	0.241	0.135	0.270	0.269	0.002
Ethnicity E	0.043	0.058	0.076	0.051	0.051	0.003
Ethnicity F	0.179	0.250	0.184	0.219	0.219	0.001
Ethnicity G	0.151	0.098	0.147	0.121	0.121	0.001
Ethnicity H	0.016	0.011	0.036	0.013	0.013	0.003
Ethnicity I	0.013	0.022	0.082	0.017	0.018	0.003
Ethnicity J	0.024	0.031	0.044	0.028	0.028	0.001
Ethnicity K	0.022	0.040	0.125	0.032	0.032	0.001
Ethnicity L	0.062	0.095	0.136	0.080	0.080	0.000
Ethnicity M	0.111	0.108	0.010	0.109	0.109	0.001
Ethnicity N	0.011	0.104	0.027	0.009	0.009	0.001
Gender	0.426	0.565	0.279	0.504	0.504	0.001

The |effect size| represents the absolute value of the standardized effect size defined as the treatment group mean divided by the treatment group standard deviation minus the control group mean divided by the treatment group standard deviation.

Figure 2: Standardized effect size in pretreatment variables after weighting

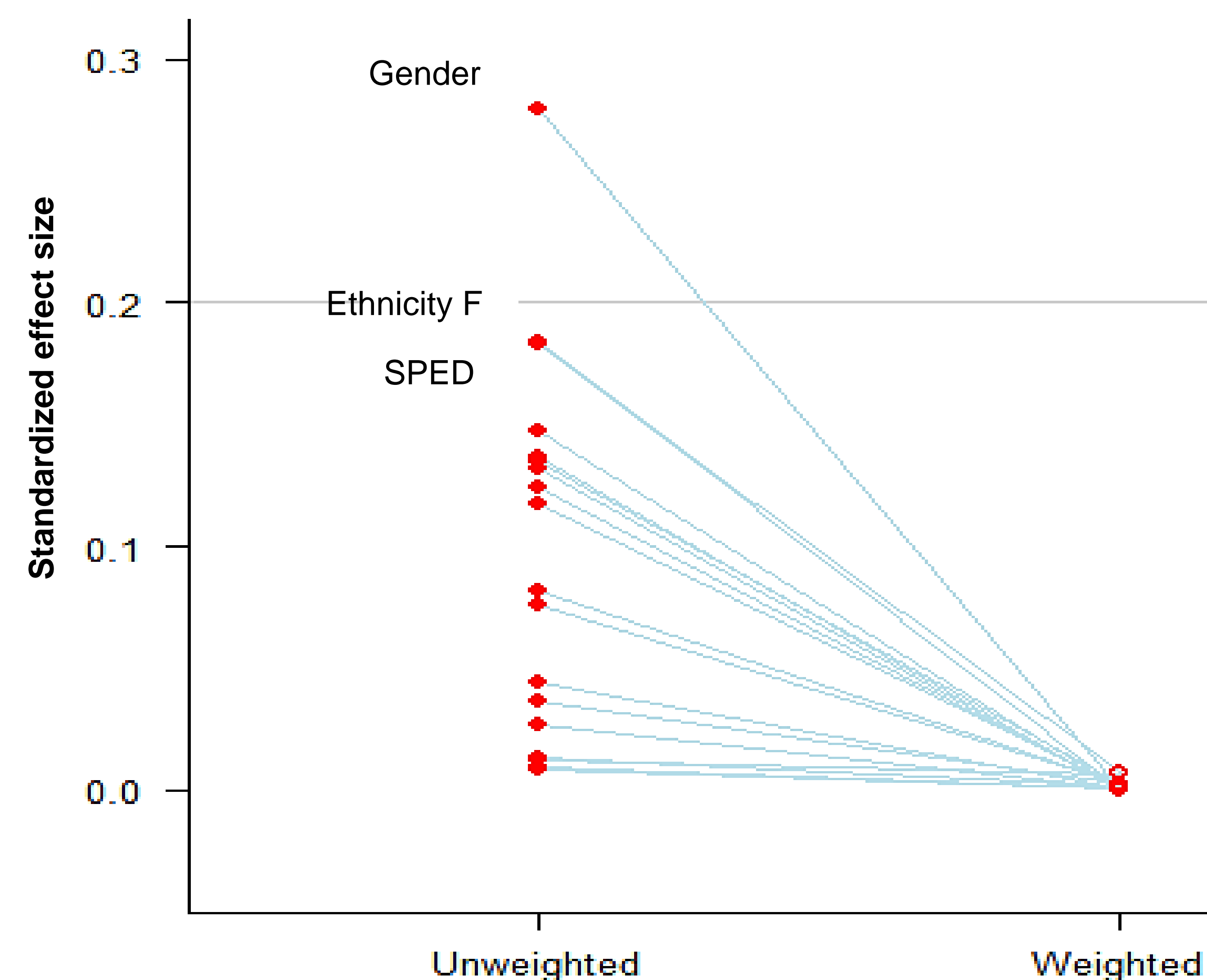


Table 2. Average treatment effect on the treated

	Estimated regression coefficient of STEP UP (S.E.)	95% confidence limits	Effect size, Cohen's d (SE)	p-value
Cohort 2013				
HSA 10 th Read	1.42 (0.41)	0.61 2.22	0.09 (0.025)	0.001
HSA 10 th Math	2.88 (0.50)	1.91 3.86	0.14 (0.025)	0.000
Cohort 2014				
HSA 10 th Read	1.02 (0.38)	0.27 1.77	0.07 (0.025)	0.008
HSA 10 th Math	2.42 (0.48)	1.47 3.36	0.12 (0.026)	0.000
Cohort 2015				
HSA 10 th Read	0.71 (0.40)	-0.07 1.49	0.04 (0.025)	0.076
HSA 10 th Math	3.70 (0.50)	2.72 4.67	0.18 (0.025)	0.000

Note: 2013 cohort ($n = 2906$ STEP UP and $n = 3748$ non-STEP UP), 2014 ($n = 3212$ STEP UP and $n = 3436$ non-STEP UP), and 2015 cohort ($n = 3320$ STEP UP and $n = 3190$ non-STEP UP).

Table 3. Average treatment effects

	Estimated regression coefficient of STEP UP (S.E.)	95% confidence limits	Effect size, Cohen's d (SE)	p-value
Cohort 2013				
HSA 10 th Read	1.45 (0.41)	0.64 2.26	0.09 (0.024)	0.004
HSA 10 th Math	3.07 (0.48)	2.14 4.01	0.16 (0.024)	0.000
Cohort 2014				
HSA 10 th Read	1.23 (0.38)	0.48 1.99	0.08 (0.024)	0.001
HSA 10 th Math	2.62 (0.47)	1.70 3.55	0.13 (0.024)	0.000
Cohort 2015				
HSA 10 th Read	0.71 (0.39)	-0.05 1.47	0.04(0.024)	0.068
HSA 10 th Math	3.81 (0.48)	2.86 4.75	0.19 (0.024)	0.000

Note: 2013 ($n = 2647$ STEP UP and $n = 4528$ non-STEP UP), 2014 ($n = 2863$ STEP UP and $n = 4310$ non-STEP UP), and 2015 cohort ($n = 3007$ STEP UP and $n = 3986$ non-STEP UP)

We statistically controlled for all covariates and also included prior achievement scores at grades 3, 5, and 8.

CONCLUSION

A positive effect was detected for Step Up scholars on reading or math outcomes at grade 10 for both the ATE and ATT estimates. We also conducted sensitivity analysis by removing the predictors used for estimating the propensity scores from the main outcome analysis model. We found no substantive difference in the estimates for both the ATEs, and ATTs, and it did not alter the conclusion of the statistical significance of our main analysis.