

**AGE-BASED TREATMENT EFFECT HETEROGENEITY AND PROPENSITY SCORE MATCHING
EVIDENCE ON COLLEGE INTERVENTIONS*****Keyuan Cao**

New York University, College of Arts and Science, New York, NY 10003, USA

Received 19th January 2026; **Accepted** 26th February 2026; **Published online** 20th April 2026

Abstract

This paper revisits the experimental framework by examining the impact of academic support and financial incentives on college students' academic performance, with a particular focus on treatment effect heterogeneity by age. Using data from the Student Achievement and Retention Project (STAR), the study applies propensity score matching (PSM) to address residual covariate imbalance and improve causal inference. Average Treatment Effects (ATE) and Intent-to-Treat (ITT) estimates are obtained through OLS and 2SLS models, respectively, while quantile regression is employed to explore heterogeneity across the achievement distribution. The results suggest that none of the interventions, academic support (SSP), financial incentives (SFP), or their combination (SFSP), produce statistically significant improvements in GPA in the full sample. However, the combined intervention (SFSP) exhibits the largest economic magnitude and the most consistent positive effects across both short-term and long-term outcomes. Quantile regression reveals that treatment effects vary across the performance distribution, with mid- to high-achieving students benefiting more than those at the lower tail. Importantly, subsample analysis uncovers substantial heterogeneity by age. Elder (non-traditional) students respond more strongly to all interventions in both the short and long run, although these effects tend to decay over time. In contrast, younger students exhibit weaker immediate responses but more persistent gains. These findings highlight the importance of accounting for demographic heterogeneity in evaluating educational interventions and suggest that policy effectiveness may differ significantly across student populations.

Keywords: Causal Inference, Propensity Score Matching, Treatment Effects, Educational Interventions.

INTRODUCTION

Attrition in higher education remains a serious policy concern even in developed countries. As is reported by NCES (National Center for Education Statistics, 2025), more than a fifth of college freshmen drop out within one-year time. However, according to the human capital theory, education brings in productivity improvement and skill accumulation. It makes education crucial to lifetime achievement in micro level and to economic development in macro level (Becker, 1992). Indeed, Heckman et al. (2018) identify the causal effect of education on earnings, health, and life habits. It indicates that benefit from education extends beyond wage to a wider range of life outcomes. In addition, at the college level, Chetty et al. (2020) document that education serves as a potential solution to intergenerational transmission of inequality. Therefore, measurements to enhance retention in university is critical both to individuals and to the nation as a whole. In the seminal paper by Angrist et al. (2009), an experimental design is adopted to examine the effect of academic support and financial incentive on academic performance of college students. The research documents the increase of students' GPA following the randomized trial. In particular, the study illustrates the persistence of the treatment effect, which makes the intervention a promising tool to boost college study outcomes. Moreover, the analysis reveals the heterogeneity of the impact by gender groups and by students' ability. Aside from students' incentive, Duflo et al. (2011) utilize a field experiment to document that teachers' incentive is effective to raise test score of students through enhanced teacher devotion.

However, there are two dimensions of the analysis by Angrist et al. (2009) that deserve a future investigation. For one thing, the potential variation of the treatment effect by age groups remains unexamined. Elder and non-traditional students could respond different to the incentive and service when compared with younger peers. This is because age group could affect the time preference, opportunity cost, and study habits of students (Jacob, 2002). For the other, despite randomization, there remains observable imbalances across treatment group versus control group students. It could introduce bias in the treatment effect estimation.

The report addresses the problem by supplementing the analysis with age group separation and PSM (propensity score matching) approach. In this way, the report is expected to extend the original study with new insights regarding:

- Whether academic support and financial incentive enhance academic performance of college students?
- Whether the impact of intervention persists over time?
- Whether the impact of intervention varies by students' ability?
- Whether the impact of intervention is heterogenous among younger versus elder students?

DATA SET AND DESCRIPTIVE STATISTICS**Data source and variable selection**

The study adopts the same dataset as Angrist et al. (2009) obtained from the online replication package. It comes from the STAR (Student Achievement and Retention Project) records with a randomized trial performed in the satellite

***Corresponding Author: Keyuan Cao**

New York University, College of Arts and Science, New York, NY 10003, USA

campus of a Canadian university. It is a cross-sectional dataset with 1656 undergraduate students registered for 2005-fall semester. Specifically, students in the experiment are randomly allocated to one of four groups. The first treatment group SSP (student support program) offers access to peer advising and facilitated study groups. The second treatment group SFP (student fellowship program) provides students with merit-based cash awards conditional on meeting GPA targets given high school grade quartile. The third treatment group SFSP combines the treatment of SSP and SFP. Besides, there is a control group of students with no additional intervention beyond the standard supporting programs offered by the university. The two outcome variables to measure academic performance are GPA of year 1 and year 2, which reflect the immediate treatment effect and the persistence of treatment effect respectively. The key independent variables are the indicator for treatment status. In particular, the study considers the three treatments separately and obtains both treatment assignment and actual treatment participant as variables of interest. The control variables are the same as with Angrist et al. (2009). It contains student level features and parental level characteristics to exclude confounding factors for the analysis of study outcomes. In particular, the report separates students into elder-aged and younger-aged groups by the median age (18 years old) for the analysis of age-based treatment effect heterogeneity.

Descriptive statistics (with matched sample)

Table 1 reports the descriptive statistics with the matched sample. Compared with table 1 in the original paper, the step of PSM better ensures the balance of treatment sample with the control sample. It makes possible a more valid estimation of the program effect through the enforcement of a common support on control variables.

RESEARCH METHOD

Propensity score matching and balance check

While random assignment sets a credible basis for causal inference, in a finite sample with imbalanced covariates between treatment group and treatment group, biased estimation is still a possible threat to internal validity. The study supplements the experiment with PSM method discussed in Rosenbaum and Rubin (1983). In particular, the propensity score is the conditional probability of receiving any treatment, which is estimated with the logit regression using pre-treatment covariates. The study implements one-to-one nearest neighbor matching with replacement based on the estimated propensity score. Moreover, the sample is restricted to the region of common support where the distribution of propensity score of treated and controlled students overlaps. Balance check is performed with the standardized bias to assess the success of PSM in reducing covariate imbalance. The obtained matching weight is adopted for all subsequent analysis to obtain estimation that is robust to residual selection on observed features.

ATE (average treatment effect estimation) and ITT (intent to treat effect)

The study estimates the ATE with the OLS regression model:

$$y = \beta_0 + \beta_1 SSP + \beta_2 SFP + \beta_3 SFSP + \theta ControlVariables + error$$

Furthermore, considering the issue of non-participation, the ITT is estimated with 2SLS regression, where the treatment allocation serves as the instrumental variable for actual participation status. The associated second-stage regression model is:

$$y = \beta_0 + \beta_1 SSP^* + \beta_2 SFP^* + \beta_3 SFSP^* + \theta ControlVariables + error$$

where the superscript indicates actual participation into the three treatments.

Quantile regression and heterogeneity by students' ability

To examine the possible variation of treatment effect across the ability distribution of students, the study performs quantile regression for the ATE analysis. In particular, the quantile points are selected following Angrist et al. (2009) to be 10th, 25th, 50th, 75th, and 90th of GPA percentile points.

Subsample analysis and heterogeneity by students' age groups

Finally, in order to detect the treatment effect heterogeneity by age groups, the study split the sample into younger and elder students. Then ATE and ITT are re-estimated to examine whether non-traditional students have a different response to incentive and service treatment. It explores an extended aspect of treatment effect not discussed in the original paper.

RESEARCH FINDINGS

ATE and ITT estimation

Table 2 reports the estimation for ATE and ITT. For immediate treatment effect, table 2 indicates a change of first year GPA by 0.02, -0.019, and 0.122 respectively with the treatment allocation of SSP, SFP, and SFSP. The associated ITT estimation is of a slightly larger magnitude as 0.036, -0.022, and 0.155 respectively with the actual participation of SSP, SFP, and SFSP. Statistically, none of the estimation is significant at 5% level. Economically, the combination of academic support with monetary incentive is shown to have the most potential to boost academic performance. The associated increase of first year GPA is about one sixth of the sample standard deviation. However, the long-term effect is slightly different from the instant effect. On the one hand, for both ATE and ITT estimation, the treatment effect on the second year GPA is of a larger economic magnitude for SSP and SFSP. On the other hand, while SFP is found to have a negative immediate effect, the long-term impact shifts to be positive. Nonetheless, similar as the immediate effect estimation, the strongest treatment effect lies with SFSP.

Quantile regression and heterogeneity analysis

Table 3a and table 3b report the quantile regression for first year and second year GPA respectively. Similar to table 2, none of the regression indicates a statistically significant treatment effect on students' academic performance. However, the quantile regression offers additional information regarding the distribution of treatment effect across students' achievement. It shows that in the first year, students of 25th, 50th, and 90th quantile points benefit from SSP and SFSP, while only students from 25th and 50th benefit from SFP. In the second year, while bottom students fail to benefit from

SSP and top students fail to benefit from SFP, all other cases treatment effect is positive. In particular, SFSP is shown to be promising in promoting long-term academic performance for median to upper median students.

Subsample comparison and heterogeneity analysis

Table 4a and table 4b repeat the analysis in table 2, while separating the sample by age groups. It reveals important findings regarding the heterogeneity of treatment effect by age of students. Among younger students, short-term impact of SSP and SFP is negative, while the influence reverses in sign for the second year. Meanwhile, SFSP has a positive effect in the short-term, and the influence gets larger in the second year. Among elder students, the treatment effect is positive for all three measurements in both short-term and long-term. The economic magnitude is much larger in contrast with younger students, which indicates that non-traditional students benefit more from the experiment. However, it is worth noting that while the effect accumulates for younger students by time, it decays among elder students. It shows that although elder students are more responsive to the intervention, the effect does not persist as with the younger sample.

Conclusion

Utilizing the cross-sectional data by Angrist et al. (2009), the report replicates the analysis while extending it with PSM and age heterogeneity analysis. Similar to the article, the study finds that in the full sample, no statistically meaningful improvement of students' GPA following academic support or financial incentive or their combination. Nonetheless, among the three choices, the combined treatment shows the most potential to enhance students' performance in both short-term and long-term.

APPENDIX

Table 1 Descriptive statistics

Variable	Control Mean (wtd)	Control SD (wtd)	SSP vs Control	SSP SE	SFP vs Control	SFP SE	SFSP vs Control	SFSP SE	F-stat	p-value
Courses enrolled as of fall 2005	4.989	0.889	-0.008	0.081	-0.059	0.078	-0.11	0.102	0.508	0.677
Female	0.58	0.494	0.002	0.046	0.028	0.045	0.015	0.055	0.151	0.929
High school GPA	78.96	4.074	-0.112	0.397	-0.023	0.387	-0.288	0.469	0.143	0.934
Age	18.264	0.577	-0.043	0.055	-0.03	0.052	0.064	0.067	0.882	0.45
Mother tongue is English	0.749	0.434	-0.023	0.04	-0.025	0.04	-0.016	0.048	0.172	0.915
Lives at home	0.801	0.4	-0.032	0.038	0.017	0.036	0.018	0.043	0.61	0.609
At first choice school	0.216	0.412	0.059	0.04	0.075	0.04	0.077	0.049	1.687	0.168
Plans to work while in school	0.781	0.414	0.027	0.037	-0.066	0.04	0.021	0.045	1.915	0.125
Mother a HS graduate	0.825	0.38	0.064	0.033	0.025	0.035	0.011	0.042	1.42	0.235
Mother a college graduate	0.312	0.464	0.106	0.044	0.02	0.043	-0.019	0.051	2.491	0.059
Father a HS graduate	0.838	0.369	0.027	0.033	0.003	0.034	0.007	0.041	0.266	0.85
Father a college graduate	0.422	0.495	0.054	0.046	0.031	0.045	0.009	0.055	0.514	0.673
Rarely puts off studying	0.229	0.421	0.007	0.04	-0.004	0.04	0.073	0.051	0.859	0.462
Never puts off studying	0.033	0.18	0	0.015	0.009	0.016	-0.016	0.015	0.684	0.562
Wants more than a BA	0.519	0.5	0.097	0.046	0.014	0.046	0.111	0.054	2.48	0.06
Intends to finish in 4 years	0.829	0.377	-0.012	0.035	0.003	0.034	-0.062	0.046	0.724	0.538

Table 2 ATE and ITT estimation

	ATE GPA Year1	ATE GPA Year2	ITT GPA Year1	ITT GPA Year2
Control Mean	1.769	2.028	1.769	2.028
Control SD	0.888	0.844	0.888	0.844
SSP	0.02	0.078	0.036	0.144
(SE)	0.083	0.092	0.15	0.171
SFP	-0.019	0.034	-0.022	0.039
(SE)	0.078	0.086	0.09	0.099
SFSP	0.122	0.123	0.155	0.157
(SE)	0.101	0.107	0.129	0.136
All control variables	Y	Y	Y	Y
N	897	784	897	784

The quantile regression documents the variation of the treatment effect by achievement distribution. However, the pattern indicates a possible extension of study gap since bottom students are found to benefit least from the program. Most importantly, the study detects the heterogeneity of treatment effect by age groups. It shows that elder-aged students benefit considerably more from the treatment, while their progress is less persistent in contrast with younger-aged students.

REFERENCES

- Becker, G.S. 1992. Human capital and the economy. *Proceedings of the American philosophical society*, 136(1), 85-92.
- Chetty, R., Friedman, J. N., Saez, E., Turner, N., & Yagan, D. (2020). Income segregation and intergenerational mobility across colleges in the United States. *The Quarterly Journal of Economics*, 135(3), 1567-1633.
- Duflo, E., Hanna, R., & Ryan, S. P. (2012). Incentives work: Getting teachers to come to school. *American economic review*, 102(4), 1241-1278.
- Jacob, B. A. (2002). Where the boys aren't: Non-cognitive skills, returns to school and the gender gap in higher education. *Economics of Education review*, 21(6), 589-598.
- Heckman, J. J., Humphries, J. E., & Veramendi, G. (2018). Returns to education: The causal effects of education on earnings, health, and smoking. *Journal of Political Economy*, 126(S1), S197-S246.
- National Center for Education Statistics. (2025). Digest of Education Statistics. Available online from: https://nces.ed.gov/programs/digest/current_tables.asp [Accessed on Feb. 12th, 2026]
- Rosenbaum, P. R., & Rubin, D. B. (1983). The central role of the propensity score in observational studies for causal effects. *Biometrika*, 70(1), 41-55.

Table 3a. Quantile regression for year 1 GPA

	(1)	(2)	(3)	(4)	(5)	(6)
	OLS	0.1	0.25	0.5	0.75	0.9
ssp	0.020 (0.083)	-0.240 (0.153)	0.080 (0.133)	0.110 (0.117)	-0.080 (0.114)	0.160 (0.128)
sfp	-0.019 (0.078)	-0.030 (0.151)	0.050 (0.096)	0.140 (0.121)	-0.120 (0.124)	-0.050 (0.125)
sfsp	0.122 (0.101)	0.010 (0.183)	0.160 (0.172)	0.110 (0.111)	-0.070 (0.187)	0.170 (0.277)
_cons	1.782*** (0.054)	0.670*** (0.103)	1.140*** (0.072)	1.740*** (0.081)	2.480*** (0.084)	2.930*** (0.085)
Control Variables	Y	Y	Y	Y	Y	Y
N	897	897	897	897	897	897

Standard errors in parentheses

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

Table 3b. Quantile regression for year 2 GPA

	(1)	(2)	(3)	(4)	(5)	(6)
	OLS	0.1	0.25	0.5	0.75	0.9
ssp	0.078 (0.092)	-0.180 (0.304)	0.120 (0.167)	0.080 (0.102)	0.100 (0.107)	0.060 (0.138)
sfp	0.034 (0.086)	0.050 (0.281)	0.030 (0.147)	0.030 (0.099)	0.030 (0.082)	-0.050 (0.092)
sfsp	0.123 (0.107)	0.060 (0.334)	0.070 (0.144)	0.130 (0.116)	0.180 (0.122)	0.040 (0.146)
_cons	2.012*** (0.058)	0.750*** (0.209)	1.500*** (0.072)	2.100*** (0.060)	2.630*** (0.058)	3.160*** (0.060)
Control Variables	Y	Y	Y	Y	Y	Y
N	784	784	784	784	784	784

Standard errors in parentheses

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

Table 4a. ATE and ITT estimation for younger group

	ATE GPA Year1	ATE GPA Year2	ITT GPA Year1	ITT GPA Year2
Control Mean	1.85	2.068	1.85	2.068
Control SD	0.841	0.825	0.841	0.825
SSP	-0.05	0.074	-0.086	0.128
(SE)	0.091	0.099	0.155	0.171
SFP	-0.083	0.007	-0.095	0.007
(SE)	0.087	0.094	0.1	0.107
SFSP	0.075	0.09	0.097	0.118
(SE)	0.114	0.119	0.15	0.154
All control variables	Y	Y	Y	Y
N	688	614	688	614

Table 4b. ATE and ITT estimation for elder group

	ATE GPA Year1	ATE GPA Year2	ITT GPA Year1	ITT GPA Year2
Control Mean	1.526	1.898	1.526	1.898
Control SD	0.979	0.897	0.979	0.897
SSP	0.218	0.021	0.514	0.059
(SE)	0.191	0.228	0.471	0.642
SFP	0.178	0.112	0.208	0.132
(SE)	0.164	0.203	0.193	0.241
SFSP	0.296	0.247	0.353	0.303
(SE)	0.205	0.236	0.244	0.286
All control variables	Y	Y	Y	Y
N	209	170	209	170
