

Online Appendix
The Effect of Advanced Placement Science on Students' Skills, Confidence and Stress
Conger, Kennedy, Long, McGhee

Appendix A: Supplemental Tables

Appendix Table 1: Balance Between Treatment and Control Group on Pre-Treatment Characteristics (Pre-Imputation)

Pre-Treatment Characteristic	(1)	(2)	(3)	(4)	(5)	(6)
	Full Sample			Survey Respondents		
	Non-Missing Observations	Control Group Mean	Difference Between Treated and Controls	Non-Missing Observations	Control Group Mean	Difference Between Treated and Controls
Age as of October of 11th Grade	1,789	16.6	-0.03 (0.02)	1,397	16.6	-0.01 (0.03)
Math Exam Score	1,534	0.41	0.06 (0.04)	1,212	0.47	0.04 (0.05)
Reading Exam Score	1,658	0.31	0.10 (0.03)	1,313	0.37	0.08 (0.03)
H.S. Grade Point Average	1,791	3.17	0.04 (0.03)	1,397	3.24	0.06 (0.03)
Female	1,793	0.59	0.00 (0.03)	1,398	0.61	-0.02 (0.04)
Asian American	1,793	0.12	0.02 (0.02)	1,398	0.12	0.03 (0.01)
Black	1,793	0.32	-0.02 (0.02)	1,398	0.27	0.00 (0.02)
Hispanic, Native American, or Multiracial	1,793	0.31	0.01 (0.02)	1,398	0.33	0.01 (0.02)
Disabled	1,796	0.02	0.00 (0.01)	1,398	0.01	0.00 (0.01)
Gifted	1,729	0.13	0.03 (0.02)	1,350	0.14	0.02 (0.02)
English Language Learner	1,729	0.05	0.01 (0.01)	1,350	0.04	0.01 (0.01)
Eligible for Free or Reduced-Price Lunch	1,380	0.55	0.01 (0.02)	1,105	0.51	0.01 (0.03)
Language Other than English Spoken at Home	1,623	0.36	0.02 (0.02)	1,277	0.36	0.01 (0.02)
Took Recommended Prerequisite Courses	1,791	0.79	0.01 (0.02)	1,397	0.79	0.02 (0.02)

Notes: Differences in columns (3) and (6) are conditional on school×cohort fixed effects. Robust standard errors clustered by school×cohort are in parentheses.

Appendix Table 2: Characteristics Associated With Completing the End of Year Survey

Pre-Treatment Characteristic	Coef.	S.E.
Age as of October of 11th Grade	-0.05	(0.08)
Math exam score	-0.10	(0.08)
Reading exam score	-0.04	(0.10)
H.S. Grade Point Average Before Year of Treatment	0.59	(0.09)
Female	-0.02	(0.07)
Black	0.20	(0.17)
Asian American	0.20	(0.12)
Hispanic, Native American, or Multiracial	0.19	(0.17)
Disabled	-0.84	(0.35)
Gifted	0.12	(0.13)
English Language Learner	-0.29	(0.20)
Eligible for Free or Reduced-Price lunch	0.08	(0.11)
Language Other than English Spoken at Home	0.07	(0.14)
Took Recommended Prerequisite Courses	0.20	(0.12)
School 1 x Cohort 2	0.45	(0.04)
School 2 x Cohort 1	(omitted)	
School 2 x Cohort 2	1.30	(0.05)
School 3 x Cohort 1	0.45	(0.09)
School 3 x Cohort 2	0.50	(0.06)
School 4 x Cohort 1	0.92	(0.12)
School 4 x Cohort 2	0.66	(0.15)
School 5 x Cohort 1	1.62	(0.12)
School 5 x Cohort 2	1.54	(0.11)
School 5 x Cohort 3	(omitted)	
School 6 x Cohort 1	1.63	(0.11)
School 6 x Cohort 2	1.85	(0.10)
School 6 x Cohort 3	1.64	(0.10)
School 7 x Cohort 1	0.64	(0.16)
School 8 x Cohort 1	0.71	(0.13)
School 8 x Cohort 2	0.97	(0.12)
School 9 x Cohort 1	0.21	(0.05)
School 9 x Cohort 2	(omitted)	

Appendix Table 2 is continued on the next page

Appendix Table 2 (Continued): Characteristics Associated With Completing the End of Year Survey

Pre-Treatment Characteristic	Coef.	S.E.
School 10 x Cohort 1	0.06	(0.12)
School 11 x Cohort 1	1.37	(0.12)
School 12 x Cohort 1	0.37	(0.09)
School 13 x Cohort 1	1.16	(0.05)
School 13 x Cohort 2	1.41	(0.07)
School 14 x Cohort 1	-0.62	(0.12)
School 14 x Cohort 2	1.22	(0.11)
School 15 x Cohort 1	(omitted)	
School 15 x Cohort 2	(omitted)	
School 16 x Cohort 1	(omitted)	
School 16 x Cohort 2	0.95	(0.07)
School 16 x Cohort 3	1.25	(0.07)
School 17 x Cohort 1	0.93	(0.07)
School 17 x Cohort 2	0.41	(0.07)
School 18 x Cohort 1	0.99	(0.11)
School 19 x Cohort 1	0.12	(0.09)
School 19 x Cohort 2	0.17	(0.09)
School 20 x Cohort 1	0.42	(0.05)
School 20 x Cohort 2	1.16	(0.06)
School 20 x Cohort 3	1.15	(0.04)
School 21 x Cohort 1	0.78	(0.10)
School 21 x Cohort 2	1.26	(0.11)
School 22 x Cohort 1	0.49	(0.15)
School 22 x Cohort 2	1.00	(0.18)
School 23 x Cohort 1	(omitted)	
School 23 x Cohort 2	0.58	(0.07)
School 24 x Cohort 1	0.07	(0.07)
School 24 x Cohort 2	0.50	(0.06)
Constant	-1.18	(1.40)

Notes: With the exception of School 23 x Cohort 1, "(omitted)" signifies a school \times cohort that drops out of the regression due to a 100% response rate. The inverse probability weight is set to 1 for all students in these school-by-cohorts. For School 23 x Cohort 1, "(omitted)" signifies that this school-by-cohort drops out of the regression due to a 0% response rate as their surveys were lost. For this school-by-cohort, we estimate the probability of completing the end-of-year survey based on a regression excluding school-by-cohort fixed effects. Standard errors clustered by school \times cohort are in parentheses. The R^2 value from a linear probability model using the same covariates is 0.277 (computed using Harel's (2009) method for calculating R^2 estimates using multiply-imputed data).

Appendix Table 3: First Stage Impacts on AP Course Enrollment and Overall Course Enrollment (Not Weighted)

	(1)	(2)	(3)	(4)	(5)	(6)
	Full Sample			Survey Respondents		
	Control Group Mean	ITT	LATE	Control Group Mean	ITT	LATE
AP Treatment Course Enrollment	0.19	0.38 (0.05) [0.00]		0.22	0.40 (0.06) [0.00]	
Share of Credits During Study Year in:						
AP Science	0.03	0.04 (0.01) [0.00]	0.11 (0.01) [0.00]	0.03	0.04 (0.01) [0.00]	0.10 (0.01) [0.00]
All AP	0.13	0.04 (0.01) [0.00]	0.11 (0.02) [0.00]	0.15	0.04 (0.01) [0.00]	0.09 (0.02) [0.00]
Other Advanced Science	0.06	-0.01 (0.01) [0.24]	-0.02 (0.02) [0.23]	0.06	-0.01 (0.01) [0.13]	-0.03 (0.02) [0.12]
All Other Advanced	0.25	-0.01 (0.01) [0.23]	-0.03 (0.02) [0.23]	0.26	-0.01 (0.01) [0.21]	-0.03 (0.03) [0.22]
Regular Science	0.06	-0.01 (0.01) [0.24]	-0.02 (0.02) [0.20]	0.05	-0.01 (0.01) [0.32]	-0.02 (0.02) [0.27]
All Regular	0.62	-0.03 (0.01) [0.02]	-0.09 (0.03) [0.00]	0.60	-0.02 (0.01) [0.12]	-0.06 (0.03) [0.06]
Number of Observations		1,819			1,417	

Notes: The ITT results in the AP Treatment Course Enrollment row are found by estimating Equation (2), and subsequent ITT and LATE results are found by variants of estimating Equation (1). Course-taking information collected from student transcripts. Control Group Mean uses the full control group for the first outcome (i.e., AP Treatment Course Enrollment) and those control group members who complied with their assignment (i.e., those who did not take the AP Treatment Course) for the subsequent outcomes. Standard errors clustered by school×cohort are in parentheses and *p*-values are in brackets.

Appendix Table 4: Treatment Contrast (Composite Variables) (Not Weighted)

	(1)	(2)	(3)
	Control Group Complier Mean	ITT	LATE
Course was Intellectually Challenging	-0.34	0.33 (0.10) [0.00]	0.81 (0.23) [0.00]
Inquiry-Based Classroom Activities	-0.08	0.14 (0.07) [0.04]	0.35 (0.16) [0.03]
Integrated Use of Technology	-0.14	0.13 (0.08) [0.09]	0.33 (0.18) [0.07]
Number of Observations		1,417	

Notes: To construct these composite variables, we first converted the values on each component variable (e.g., strongly agree, agree, neutral, disagree, or strongly disagree) so that the highest category was set to 1.0, the lowest to 0.0, and the remaining categories evenly spaced between 0.0 and 1.0. We then averaged and standardized these converted values. Appendix Table 5 provides the list of component variables. Standard errors clustered by school×cohort are in parentheses and *p*-values are in brackets.

Appendix Table 5: Treatment Contrast (Component Variables)

	(1)	(2)	(3)
	Control Group Complier Mean	ITT	LATE
Course was Intellectually Challenging (1)	0.68	0.15 (0.03) [0.00]	0.37 (0.07) [0.00]
Teacher Set High Standards (1)	0.74	0.07 (0.03) [0.02]	0.19 (0.07) [0.01]
Students Were Driven to Succeed (1)	0.56	0.09 (0.04) [0.02]	0.24 (0.10) [0.01]
Often Participate in Hands-on Learning (2)	0.52	0.10 (0.04) [0.03]	0.25 (0.11) [0.03]
Often Design Own Projects or Experiments (2)	0.24	0.07 (0.03) [0.05]	0.18 (0.09) [0.05]
Often Work Independently (2)	0.59	-0.01 (0.03) [0.83]	-0.02 (0.08) [0.82]
Often Work in Small Groups (2)	0.54	0.07 (0.03) [0.03]	0.17 (0.07) [0.02]
Often Present What You Learned (2)	0.32	-0.02 (0.04) [0.60]	-0.05 (0.10) [0.58]
Often Apply Knowledge to Solve New Problem (2)	0.48	0.03 (0.03) [0.29]	0.08 (0.08) [0.29]
Often Read Science Non-Textbook or Magazine (2)	0.11	0.00 (0.03) [0.98]	0.00 (0.06) [0.98]
Technology Used to:			
Conduct Interactive Simulations (3)	0.26	0.02 (0.03) [0.57]	0.05 (0.08) [0.55]
Create Graphical Presentations of Information (3)	0.27	0.02 (0.03) [0.48]	0.05 (0.07) [0.44]
Develop Collaborative Projects and/or Group Presentations (3)	0.34	0.02 (0.03) [0.45]	0.06 (0.08) [0.42]
Practice Concepts (3)	0.28	0.03 (0.03) [0.38]	0.07 (0.08) [0.36]
Course was Interesting	0.74	0.02 (0.03) [0.53]	0.05 (0.07) [0.52]
Often Listen to Teacher Lecture	0.75	0.01 (0.02) [0.73]	0.02 (0.06) [0.71]
Often Receive Homework	0.47	0.15 (0.04) [0.00]	0.39 (0.10) [0.00]
Often Take Multiple Choice Test or Quizzes	0.74	0.03 (0.03) [0.25]	0.08 (0.07) [0.22]
Number of Observations		1,417	

Notes: (1) denotes component of Academically Challenging Curriculum. (2) denotes component of Inquiry- Based Classroom Activities. (3) denotes component of Integrated Use of Technology. All outcomes are binary: =1 if strongly agree or agree versus =0 if neutral, disagree, or strongly disagree; or =1 if very frequently or frequently versus =0 if occasionally, rarely, or never). Results are weighted by the inverse probability of completing the survey. Standard errors clustered by school×cohort are in parentheses and *p*-values are in brackets.

Appendix Table 6: AP Course Impact on Science Skill, STEM Interest, Confidence, Stress, and Grades (Not Weighted)

	(1)	(2)	(3)
	Control Group Complier Mean	ITT	LATE
Science Skill	-0.13	0.08 (0.06) [0.17]	0.20 (0.14) [0.16]
STEM Interest	0.61	0.04 (0.02) [0.08]	0.10 (0.06) [0.08]
Confidence in College Science	0.92	-0.04 (0.02) [0.08]	-0.10 (0.05) [0.04]
Stress	0.12	0.07 (0.03) [0.02]	0.18 (0.07) [0.01]
Grades in Science Courses	2.80	-0.12 (0.07) [0.08]	-0.29 (0.16) [0.07]
Grades in Other Courses	3.14	-0.07 (0.02) [0.00]	-0.18 (0.06) [0.00]
Number of Observations	1,819 for grades; 1,417 for other outcomes		

Notes: Science skill has been standardized to have a mean of 0 and SD of 1 for the full sample of participating students. STEM interest = 1 if high or some interest in pursuing a STEM degree, or =0 if no interest. Confidence in college science = 1 if extremely or somewhat confident in ability to complete a college science course, or =0 if somewhat not confident or not at all confident. Stress = 1 if most recent science course had strong negative or negative impact on physical or emotional health, or =0 if strong positive impact, positive impact, or no impact. Grades in science and other courses are obtained from student transcripts and measure grades during the study year. Results, with the exception of grades during study year, are weighted by the inverse probability of completing the survey. Standard errors clustered by school \times cohort are in parentheses and p -values are in brackets.

Appendix B: Heterogeneity Analyses

B.1. Heterogeneity in Impacts by Students and Teachers

We test for heterogeneity in treatment effects along several dimensions. Here we offer the theoretical motivation for each of these analyses.

Prior Academic Preparation: Skeptics of AP worry that the College Board’s focus on expanding access has resulted in too many insufficiently-prepared students taking the challenging courses and performing poorly as a result (Bowie 2013; Dougherty and Mellor 2009; Smith et al. 2017; Tierney 2012). To test this hypothesis, we focus on three measures of prior academic preparation: 1) whether the student took the recommended prerequisite courses; 2) whether the student tested above or below the median math score prior to enrollment; and 3) whether the student tests above or below median on their predicted outcome based on pre-treatment characteristics. To identify students on this last dimension, we first regress each outcome on all pre-treatment covariates using data from the sample of control group compliers, and then applying the estimated coefficients to predict the outcome for all participants. For this analysis, we use the leave-out-one method recommended by Abadie, Chingos, and West (2018) to correct for the bias caused by endogenous stratification.

Teacher Preparation: Teachers who have some experience with the class may be more skilled, leading to larger treatment impacts. We examine this by estimating heterogeneity in treatment effects according to: 1) whether the AP Science teacher previously taught an AP course, and 2) whether the student was in their school’s first cohort (i.e., the first administration of the new AP course).

Race and Gender: Female, Black, and Latino students in the United States are far less likely to pursue math-intensive STEM degrees and occupations than their male, White and Asian counterparts (Chen 2009; National Center for Education Statistics (NCES) 2011; Quinn and Cooc 2015; Turner and Bowen 1999). Prior research attributes these disparities to low confidence among girls in mathematics and other cultural barriers that reduce female and minority students’ feelings of belongingness in their science courses (Bench et al. 2015; Ellis et al. 2016; Fryer and Levitt 2004; Kahn and Ginther 2017; Kurth et al. 2002; Leslie et al. 2015; Litzler et al. 2014). AP science courses were restructured in 2012 to address some of these barriers. Instead of a heavy reliance on rote memorization and direct learning, AP science courses are supposed to rely more on students leading the inquiry, developing hypotheses and designing their own experiments. AP science instructors are also encouraged to emphasize depth over breadth and learn different modes of instruction in order to reach students with different learning styles (College Board 2011a, b). Science experts posit that these approaches will be particularly effective at drawing in students whose interest in STEM is often suppressed by more traditional pedagogy (National Research Council 2012). Correspondingly, we test whether treatment effects differ according to: 1) whether the student is from an underrepresented minority group (Latino, Black, Native American) or not (White, Asian); and 2) whether the student is female or not female.

To estimate heterogeneity in the LATE, we implement the following fully-interacted specification:

$$(1) Y_{ij} = \alpha_j + \text{Indicator}_i \text{AP}_{ij} \beta_1 + (1 - \text{Indicator}_i) \text{AP}_{ij} \beta_2 + \mathbf{X}_i \gamma + \text{Indicator}_i \mathbf{X}_i \lambda + \varepsilon_{ij},$$

where Indicator_i equals 1 if the student is in a subgroup of interest (e.g., above median predicted science skill). We test for equality of β_1 and β_2 to assess whether there is evidence of

heterogeneity in the effects of taking AP. Equation (3) contains two endogenous variables: $Indicator_i AP_{ij}$ and $(1 - Indicator_i) AP_{ij}$. We follow the strategy for dealing with endogenous interaction terms suggested by Wooldridge (2010). The first stage regressions place each of our endogenous variables on the left-hand side of the equation:

$$(2) Indicator_i AP_{ij} = \varphi_j + \rho_1 Indicator_i Offered_{ij} + \sigma_1 (1 - Indicator_i) Offered_{ij} + \mathbf{X}_i \tau_1 + Indicator_i \mathbf{X}_i \psi_1 + \varepsilon_{ij},$$

$$(3) (1 - Indicator_i) AP_{ij} = \zeta_j + \rho_2 Indicator_i Offered_{ij} + \sigma_2 (1 - Indicator_i) Offered_{ij} + \mathbf{X}_i \tau_2 + Indicator_i \mathbf{X}_i \psi_2 + \varepsilon_{ij},$$

where φ_j and ζ_j are school by cohort fixed effects. Both $Indicator_i Offered_{ij}$ and $(1 - Indicator_i) Offered_{ij}$ serve as instruments in this first stage. We estimate our second stage, Equation (3), by replacing $Indicator_i AP_{ij}$ and $(1 - Indicator_i) AP_{ij}$ with the fitted values generated in Equations (4) and (5).

The results from all of these analyses are show in the table below. As explained in the text (and shown in the Table), these results are too noisy to draw meaningful conclusions.

Appendix Table 6: Heterogeneity in AP Impact (LATE)

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)	(15)	(16)	(17)	(18)	(19)	(20)	(21)
	Student at or Above Median Predicted Outcome			Student Took Recommended Prerequisite Courses			Student at or Above Median Math Exam Score			AP Science Teacher Previously Taught an AP Course			Cohort 1 (First Administration of Course)			Underrep. Minority			Female		
Outcome	Yes	No	Diff.	Yes	No	Diff.	Yes	No	Diff.	Yes	No	Diff.	Yes	No	Diff.	Yes	No	Diff.	Yes	No	Diff.
Science Skill	0.07	0.59	-0.52	0.34	-0.09	0.43	0.12	0.39	-0.27	0.06	0.36	-0.30	0.22	0.24	-0.01	0.30	0.08	0.22	0.23	0.31	-0.09
	0.18	0.36	0.40	(0.19)	(0.24)	(0.31)	(0.19)	(0.20)	(0.24)	(0.20)	(0.24)	(0.31)	(0.30)	(0.16)	(0.34)	0.17	0.21	0.24	0.22	0.19	0.27
	[0.70]	[0.10]	[0.19]	[0.08]	[0.70]	[0.16]	[0.51]	[0.05]	[0.27]	[0.75]	[0.13]	[0.34]	[0.45]	[0.13]	[0.97]	[0.08]	[0.71]	[0.37]	[0.31]	[0.10]	[0.74]
STEM Interest	-0.18	0.43	-0.61	0.14	-0.01	0.15	-0.03	0.26	-0.28	0.09	0.11	-0.02	0.13	0.08	0.05	0.22	-0.02	0.24	0.16	0.07	0.09
	0.11	0.12	0.19	(0.08)	(0.09)	(0.12)	(0.10)	(0.13)	(0.19)	(0.07)	(0.12)	(0.14)	(0.16)	(0.06)	(0.17)	0.11	0.09	0.14	0.11	0.09	0.15
	[0.11]	[0.00]	[0.00]	[0.10]	[0.91]	[0.21]	[0.79]	[0.05]	[0.13]	[0.20]	[0.34]	[0.86]	[0.42]	[0.18]	[0.76]	[0.04]	[0.84]	[0.09]	[0.15]	[0.44]	[0.55]
Confidence in College Science	-0.20	-0.02	-0.18	-0.10	-0.11	0.00	-0.11	-0.11	0.00	-0.12	-0.08	-0.03	-0.05	-0.15	0.11	-0.05	-0.14	0.09	-0.16	-0.06	-0.10
	0.07	0.09	0.11	(0.07)	(0.05)	(0.09)	(0.08)	(0.09)	(0.12)	(0.07)	(0.07)	(0.10)	(0.09)	(0.06)	(0.11)	0.07	0.07	0.09	0.08	0.06	0.10
	[0.00]	[0.86]	[0.11]	[0.12]	[0.03]	[0.97]	[0.15]	[0.24]	[0.99]	[0.09]	[0.25]	[0.73]	[0.62]	[0.01]	[0.33]	[0.50]	[0.06]	[0.33]	[0.06]	[0.30]	[0.33]
Stress	0.19	0.29	-0.10	0.21	0.01	0.20	0.20	0.20	0.00	0.22	0.11	0.12	0.18	0.15	0.03	0.12	0.24	-0.12	0.25	0.11	0.14
	0.09	0.17	0.20	(0.08)	(0.08)	(0.09)	(0.09)	(0.09)	(0.12)	(0.08)	(0.10)	(0.13)	(0.08)	(0.10)	(0.13)	0.08	0.09	0.10	0.10	0.08	0.11
	[0.04]	[0.09]	[0.61]	[0.00]	[0.92]	[0.03]	[0.03]	[0.03]	[1.00]	[0.00]	[0.31]	[0.36]	[0.03]	[0.12]	[0.84]	[0.14]	[0.00]	[0.21]	[0.01]	[0.18]	[0.21]
Grades in Science Courses	-0.46	-0.17	-0.29	-0.26	-0.42	0.16	-0.32	-0.38	0.06	-0.52	-0.10	-0.42	-0.14	-0.45	0.31	-0.23	-0.39	0.16	-0.31	-0.29	-0.02
	0.12	0.27	0.29	(0.20)	(0.19)	(0.25)	(0.19)	(0.20)	(0.22)	(0.19)	(0.23)	(0.29)	(0.22)	(0.21)	(0.30)	0.19	0.20	0.22	0.20	0.20	0.24
	[0.00]	[0.54]	[0.31]	[0.19]	[0.03]	[0.53]	[0.10]	[0.05]	[0.77]	[0.01]	[0.65]	[0.15]	[0.53]	[0.03]	[0.30]	[0.23]	[0.05]	[0.46]	[0.11]	[0.15]	[0.93]
Grades in Other Courses	-0.09	-0.26	0.17	-0.17	-0.24	0.07	-0.22	-0.17	-0.06	-0.14	-0.22	0.08	-0.18	-0.19	0.01	-0.18	-0.21	0.03	-0.25	-0.16	-0.09
	0.08	0.09	0.12	(0.07)	(0.12)	(0.15)	(0.09)	(0.09)	(0.14)	(0.09)	(0.09)	(0.12)	(0.10)	(0.06)	(0.12)	0.08	0.11	0.14	0.08	0.10	0.12
	[0.28]	[0.01]	[0.18]	[0.02]	[0.06]	[0.63]	[0.01]	[0.07]	[0.67]	[0.11]	[0.02]	[0.53]	[0.09]	[0.00]	[0.94]	[0.02]	[0.06]	[0.84]	[0.00]	[0.10]	[0.48]

Number of Observations

1,819 for grades; 1,417 for other outcomes

Notes: Science skill has been standardized to have a mean of 0 and SD of 1 for the full sample of participating students. STEM interest = 1 if high or some interest in pursuing a STEM degree, or = 0 if no interest. Confidence in college science = 1 if extremely or somewhat confident in ability to complete a college science course, or = 0 if somewhat not confident or not at all confident. Stress = 1 if most recent science course had strong negative or negative impact on physical or emotional health, or = 0 if strong positive impact, positive impact, or no impact. Grades in science and other courses are obtained from student transcripts and measure grades during the study year. Results, with the exception of grades during study year, are weighted by the inverse probability of completing the survey. Standard errors clustered by School×Cohort are in parentheses and p-values are in brackets

B.2. Quantile Regressions

We conduct quantile regressions for our three continuous outcomes (i.e., science skill, grades in sciences, and grades in other classes), conditional on \mathbf{X}_i , to further examine the prospect of heterogeneous results. These results are shown in Figure 1.

We have tested the joint null that all of the estimated quantile treatment effects are equal and we are unable to reject this null hypothesis with p -values above 0.9 in each case.

Supplemental Figure 1: Conditional Quantile Regression Results

Figure 1a. Science Skill

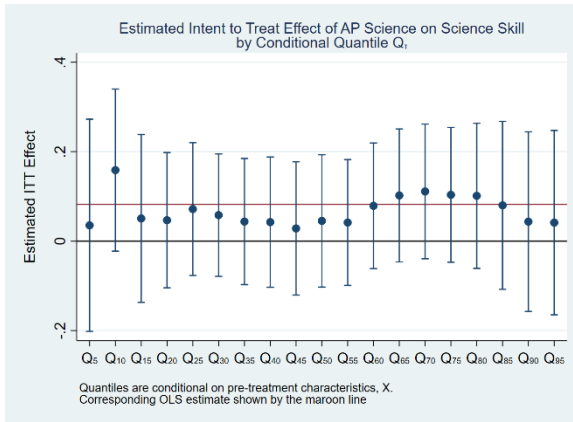


Figure 1b. Grades in Science

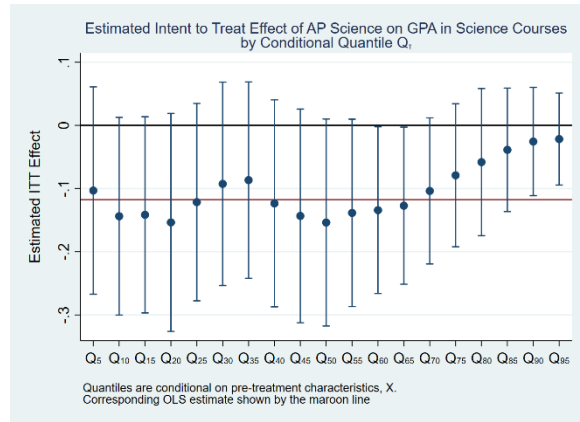
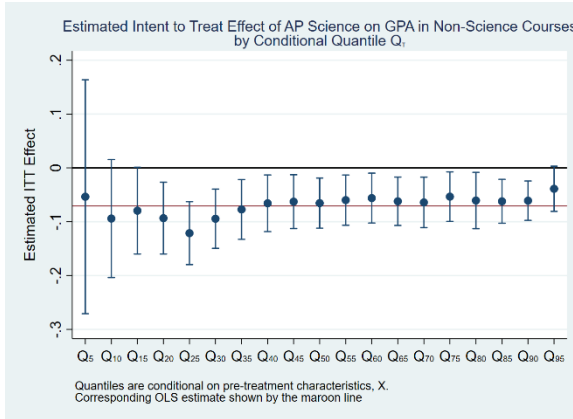


Figure 1c. Grades in Other (Non-Science)



References

- Abadie, A., M. M. Chingos, and M. R. West. (2018). Endogenous stratification in randomized experiments, *Review of Economics and Statistics*, posted online on March 16, https://doi.org/10.1162/REST_a_00732.
- Bench, S. W., H. C. Lench, J. Liew, K. Miner, and S. A. Flores (2015). Gender gaps in overestimation of math performance. *Sex Roles* 72(11-12), 536–546.
- Bennett, J., S. Hogarth, F. Lubben, B. Campbell, and A. Robinson. (2010). Talking science: The research evidence on the use of small group discussions in science teaching. *International Journal of Science Education*, 32(1), 69-95.
- Bowie, L. (2013). Maryland schools have been leader in Advanced Placement, but results are mixed. *The Baltimore Sun*.
- Chen, X. (2009). Students who study Science, Technology, Engineering, and Mathematics (STEM) in postsecondary education. Stats in brief. NCES 2009-161. *National Center for Education Statistics*.
- Dougherty, C. and L. Mellor (2009). Preparation matters. *National Center for Educational Achievement*.
- Ellis, J., B. K. Fosdick, and C. Rasmussen (2016). Women 1.5 times more likely to leave stem pipeline after calculus compared to men: lack of mathematical confidence a potential culprit. *PLOS ONE* 11(7), e0157447.
- Fryer, R. G. and S. D. Levitt (2004). Understanding the black-white test score gap in the first two years of school. *The Review of Economics and Statistics* 86(2), 447–464.
- Harel, O. 2009. “The estimation of R-squared and adjusted R-squared in incomplete data sets using multiple imputation.” *Journal of Applied Statistics* 36 (10): 1109–1118.
- Kahn, S. and D. Ginther (2017). Women and STEM. *National Bureau of Economic Research*, Working Paper No. 23525.
- Kurth, L. A., C. W. Anderson, and A. S. Palincsar (2002). The case of Carla: Dilemmas of helping all students to understand science. *Science Education* 86(3), 287–313.
- Leslie, S. J., A. Cimpian, M. Meyer, and E. Freeland (2015). Expectations of brilliance underlie gender distributions across academic disciplines. *Science* 347(6219), 262–265.
- Litzler, E., C. C. Samuelson, and J. A. Lorah (2014). Breaking it down: Engineering student STEM confidence at the intersection of race/ethnicity and gender. *Research in Higher Education* 55(8), 810–832.
- National Center for Education Statistics (NCES) (2011). The nation’s report card: Science 2012. *Institute of Education Sciences, U.S. Department of Education*.
- National Research Council (2012). A framework for science education: Preliminary public draft. Washington, DC: Committee on Conceptual Framework for New Science Education Standards, Board on Science Education, *National Research Council*.
- Quinn, D. M. and N. Cooc (2015). Science achievement gaps by gender and race/ethnicity in elementary and middle school: Trends and predictors. *Educational Researcher* 44(6), 336–346.
- Tierney, J. (2012). AP classes are a scam. *The Atlantic*. October 13.
- Turner, S. E. and W. G. Bowen (1999). Choice of major: The changing(unchanging) gender gap. *Industrial and Labor Relations Review*. *Industrial and Labor Relations Review* 52(2), 289–313.
- Wooldridge, J.M. (2010). *Econometric analysis of cross section and panel data*. 2nd Edition. MIT Press.