Persistency in Teachers' Grading Biases and Effect on Longer Term Outcomes: University Admission Exams and Choice of Field of Study

Victor Lavy
University of Warwick

Rigissa Megalokonomou
University of Queensland

September 4, 2019

Introduction

- Girls outperform boys in school achievements in primary and secondary school.
- Smaller gaps in STEM, higher boys share in university degrees in these fields of study:
    - US 2014: women account for 57 percent of all BA degrees (NCES 2015).
    - STEM subjects: physical sciences/science technologies - 39%, computer/information sciences - 18%, engineering - 18%, computer engineering - 10 % (NCES 2015).
- These gaps determine gender occupational differences: 14% of engineers in US are women.

Introduction

- Debate of what shapes these gender differences is the focus of much recent research:
  - Biological gender differences determining gender cognitive differences (Witelson 1976, Waber 1976).
  - Social, psychological and environmental factors that influence this gap.
- Limited credible evidence for this debate.
  - Difficult to disentangle the impact of biological gender dissimilarities from environmental conditions.
  - It is difficult to measure stereotypes and prejudices and test their causal implications.

Objectives of this paper

- Measure high school teachers' gender grading biases and examine how persistent they are over time/classes.
- To do this we use a set of blind and non-blind exams for each student.
- Test whether teachers' gender biases during high school influence students' performance in university admission exams and student's choice of university field of study.
- Using data from a large sample of high schools in Greece (2003-2011).
- Use panel data on teachers, on average 15 classes per teacher.
- Measuring the bias 'out of sample' relieves concerns that the measure of gender bias may just pick up random (small sample) variation in the unobserved "quality" or "non-cognitive" skills of the boys vs. girls in a particular single class or any other class specific dynamics.
- Examine whether there is any association between teachers' gender bias and teachers' quality.

Contribution

- To our knowledge this is the first paper that establishes a believable causal connection between high school "culture" and the prevalence of gendered outcomes.
- We are also the first to study the effect of teacher gender bias on the choice of university study and other university related outcomes.
- This is the first paper that exploits a teachers' panel data and measures the bias "out of sample" to assess the impact of the persistence component in teachers' stereotypical biases.
- First paper that links teacher gender bias to teacher quality.

Main Findings

- Our results may be summarized with the following broad conclusions:
- *High persistency in teachers' stereotypical behavior:* teachers who are biased for one class are biased in the same way for other classes in the same, earlier or later academic years.
- The teachers' gender bias has a significant effect on:
  1) Students' subsequent performance.
  Boys (Girls) benefit from having a pro-boy (pro-girl) teacher, while for girls (boys) there is a negative effect on subsequent performance.
  2) Students' enrollment in a related field of study at the university.
  This effect is larger and statistically significant for girls and not different from zero for boys.
  3) Other educational decisions (such as university enrolment, quality of program they enrol).
- We also find that gender biases are more prevalent among low value added teachers, while the more effective teachers have an approximately neutral gender attitude.

How Teacher's Gender Biases Can Affect Students?

- Potential mechanisms that the **psychology and sociology literature** suggest:
- Teachers treat successes/failures of boys and girls differently by:
  - Encouraging boys to try harder and allowing girls to give up (Rebhorn and Miles 1999)
- Teachers give more attention to boys by:
  - addressing them more often in class,
  - giving them more time to respond,
  - providing them with more substantive feedback (Sadker and Sadker 1985)
- Teachers spend more time training girls in reading and less time in math, relative to boys (Leinhardt, Seewald and Engel 1979).
- Girls are less likely than boys to be advised/encouraged to take courses in math (NCES 1997).

Literature Review

- In economics, the systematic difference between blind and non-blind assessment across groups as a measure of discrimination was introduced by Blank (1991) and Goldin and Rouse (2000).
- Blank (1991) shows that the probability of papers being accepted by economic journals depends on authors' affiliation. Goldin and Rouse (2000) examine sex-biased hiring patterns in orchestras by comparing blind and non-blind auditions.
- In economics of education, the **blind and non-blind assessment** was used in Lavy (2008), Bjorn, Hoglin, and Johannesson (2011), Hanna and Linden (2012), Cornwell, Mustard, Van Parys (2013) and Burgess and Greaves (2013).
- However, two recent and related papers: Lavy and Sand (2017) and Terrier (2016) use this assessment to compute a **teacher's gender bias**.

Outline

Data

- We combine data from different sources:
- *Hellenic Ministry of Education:* student level information for 12th grade students in 135 schools and university admission related outcomes.
- *School Archives*:
  a) student level information for 11th graders in these 135 schools and
  b) teachers' panel information with class/subject assignment in 22 schools (including the gender of teachers and principals)
- Data for 11th and 12th graders: gender, year of birth, track of study in high school, absenteeism, repetition and drop out, test scores of the school and national exams in all subjects in 11th and 12th grade.
- Baseline sample: 11th grade students in 2003-2005 and 12th grade students in 2003-2011.

Institutional Information

- Teachers and students are randomly assigned to classes.
- Non-blind exam: the name and gender of students are revealed to teachers.
- Blind exam: the name and gender of students are hidden.
- The two exams have the same format, examine the same material and they are taken within a very short period of time.
- They are both high stake exams (different weights in the calculation of the university admission score).

Measuring Teacher Grading Bias

- Difference between boys' and girls' average gap between the non-blind (NB) and blind score (B):

$$Teacher_j GenderBiasinClass_c = Mean_c[\sum_{ic}(NBi - Bi|Male_i)]$$

$$-Mean_c[\sum_{ic}(NBi - Bi|Female_i)]$$

- Repeat this procedure for every class, subject and grade.
- Positive (negative) values indicate bias in favor of boys (girls) in this particular subject.
- Construct average bias of teacher based on all other classes (other students) except current class.
- High variation in teachers' discriminatory behaviour.
- This approach alleviates concern that our teacher bias measure picks up class level unobserved variation in boys' and girls' behavior or other gender differential non-cognitive characteristics.

Descriptive Statistics for Different Measures of Teacher Bias in 11th Grade, Sample of 21 Schools

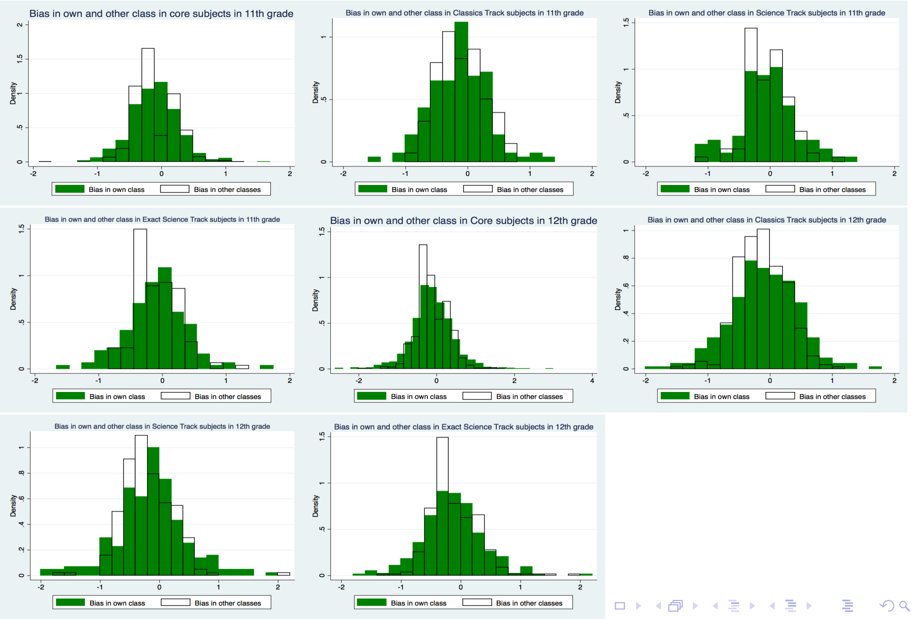
Variable	Prop. of Fem. Teachers	Teacher Bias measured in in other classes		Teacher Bias measured in in own class				Correlation between
		Mean	**(sd)**	**Mean**	**(sd)**	**(Diff.)**	**(se)**	(2)and(4)
Bias in	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
11th grade (2003-2005)								
Core subjects								
Modern Greek	0.71	-0.100	(0.369)	-0.097	(0.463)	-0.003	(0.004)	0.85
History	0.67	-0.110	(0.307)	-0.129	(0.378)	0.019	(0.004)	0.80
Algebra	0.39	-0.095	(0.236)	-0.106	(0.306)	0.010	(0.003)	0.78
Geometry	0.37	-0.102	(0.253)	-0.094	(0.303)	-0.008	(0.003)	0.78
Physics	0.45	-0.104	(0.269)	-0.094	(0.319)	-0.010	(0.003)	0.81
Classics Track								
Ancient Greek	0.63	-0.160	(0.355)	-0.152	(0.396)	-0.008	(0.007)	0.80
Philosophy	0.66	-0.061	(0.363)	-0.027	(0.415)	-0.034	(0.007)	0.79
Latin	0.69	-0.116	(0.284)	-0.087	(0.372)	-0.029	(0.007)	0.78
Science Track								
Mathematics	0.46	-0.074	(0.224)	-0.066	(0.326)	-0.008	(0.008)	0.72
Physics	0.41	-0.004	(0.246)	-0.018	(0.332)	0.013	(0.007)	0.75
Chemistry	0.37	-0.095	(0.323)	-0.077	(0.351)	-0.018	(0.006)	0.87
Exact Science Track								
Mathematics	0.32	-0.057	(0.258)	-0.080	(0.307)	0.024	(0.010)	0.86
Physics	0.39	-0.104	(0.243)	-0.105	(0.334)	0.0002	(0.005)	0.84
Technology and Computers	0.29	-0.223	(0.338)	-0.248	(0.397)	0.025	(0.007)	0.82

Notes: Negative bias means that the teacher is pro-girl. The means are weighted by number of students.

Variable Bias in	Prop. of Fem. Teachers (1)	Teacher Bias measured in in other classes Mean (2)	 (sd) (3)	Teacher Bias measured in in own class Mean (4)	 (sd) (5)	(Diff.) (6)	(se) (7)	Correlation between (2)and(4) (8)
12th grade (2003-2011)								
Core subjects								
Modern Greek	0.59	-0.045	(0.365)	-0.062	(0.500)	0.017	(0.004)	0.76
Biology	0.20	-0.112	(0.429)	-0.158	(0.667)	0.046	(0.007)	0.60
History	0.48	-0.113	(0.319)	-0.157	(0.409)	0.044	(0.005)	0.67
Mathematics	0.29	-0.128	(0.336)	-0.128	(0.512)	0.0003	(0.004)	0.72
Physics	0.020	-0.172	(0.276)	-0.193	(0.328)	0.021	(0.004)	0.69
Classics Track								
Ancient Greek	0.53	-0.060	(0.341)	-0.039	(0.396)	-0.021	(0.003)	0.81
Latin	0.64	-0.101	(0.300)	-0.080	(0.388)	-0.021	(0.003)	0.77
Literature	0.57	-0.108	(0.352)	-0.069	(0.502)	-0.040	(0.004)	0.80
History	0.58	-0.150	(0.292)	-0.178	(0.370)	0.029	(0.003)	0.75
Science Track								
Biology	0.25	-0.141	(0.373)	-0.089	(0.587)	-0.052	(0.005)	0.73
Mathematics	0.13	-0.195	(0.381)	-0.203	(0.511)	0.008	(0.005)	0.68
Physics	0.20	-0.231	(0.283)	-0.268	(0.462)	0.037	(0.004)	0.74
Chemistry	0.19	-0.169	(0.408)	-0.147	(0.529)	-0.022	(0.005)	0.71
Exact Science Track								
Mathematics	0.27	-0.126	(0.284)	-0.138	(0.327)	0.012	(0.003)	0.76
Physics	0.21	-0.193	(0.279)	-0.184	(0.345)	-0.009	(0.003)	0.69
Business Administration	0.58	-0.134	(0.313)	-0.150	(0.401)	0.016	(0.003)	0.73
Computers	0.35	-0.182	(0.273)	-0.191	(0.373)	0.008	(0.003)	0.68
Optional								
Economics	0.56	-0.108	(0.307)	-0.065	(0.440)	-0.044	(0.004)	0.76

Notes: Negative bias means that the teacher is pro-girl. The means are weighted by number of students.

Distributions of teacher biases in own and other classes



Descriptive Statistics for 11th and 12th Grade Teachers, Sample of 21 Schools

Variable	Mean	Std. Dev.	Min.	Max.
Panel A: 11th grade				
Number of classes taught by a teacher	13.12	10.06	1	58
Number of classes/subjects taught by a teacher	4.66	3.02	1	18
Number of different subjects taught by a teacher	1.90	0.92	1	5
Number of different classes taught by a teacher	1.98	1.22	1	7
Number of years a teacher teaches by year	2.87	1.78	1	9
Panel B: 12th grade				
Number of classes taught by a teacher	18.88	15.22	1	73
Number of classes/subjects taught by a teacher	2.65	1.60	1	9
Number of different subjects taught by a teacher	2.58	1.92	1	18
Number of different classes taught by a teacher	2.17	1.42	1	8
Number of years a teacher teaches	4.39	2.34	1	9

Notes: Panel A includes all teachers who teach core or track subjects in 11th. Panel B includes all teachers who teach core or track subjects in 11th.

Correlations Between Different Measures of Teacher Bias

Dependent Variable: Current year own teacher bias				
	11th grade		12th grade	
Bias measured in other classes in same year	0.813	0.801	0.612	0.606
	(0.035)***	(0.037)***	(0.037)***	(0.038)***
Sample Size	818	818	1,279	1,279
Bias measured in other classes in any year	0.720	0.704	0.731	0.723
	(0.052)***	(0.049)***	(0.033)***	(0.033)***
Sample Size	844	844	1,895	1,895
Subjects FE	✓	✓	✓	✓
Year FE	✓	✓	✓	✓
School FE		✓		✓

Notes: The sample includes all teachers who teach core and track subjects. Standard errors are clustered by school and are reported in parentheses. *,**,*** denotes significance at the 10%, 5% and 1% level respectively.

Correlations between Biases in the own class (Core subjects, 11th grade)

Dependent Variable: Teacher bias in own subject				
	Different Teachers		Same Teachers	
	(1)	(2)	(3)	(4)
Teacher bias in other subject	0.034	0.034	0.719	0.717
	(0.028)	(0.028)	(0.058)***	(0.058)***
N	1,032	1,032	127	127
Subject FE	X	✓	X	✓

Note: The table includes stacked observations for the teacher bias in each subject. The (OLS) estimated coefficients in columns 1-2 is the between biases measures of different teachers who instruct students from the same class two different subjects and the (OLS) estimated coefficients in columns 3-4 is the between biases measures of same teachers who instruct students from the same class in two subjects.

▸ Core by subject ▸ Classics ▸ Science ▸ Exact Science ▸ 12th Grade

Outline

Methodology and Estimation Framework

- Estimate effect of teachers' gender biases in 11th grade on 12th grade national exams test scores:

$$Y_{icjt} = \alpha + \tau_c + \theta_j + \lambda_t + \gamma X_{icjt} + \pi TB_{cj} + \phi_{cj} + \psi_{icjt}$$

  Y_{icjt}: outcome of student i, school or class c, subject j and year t;
  X_{icjt}: student characteristics: track, score in national exam;
  τ_c: is a high school or class fixed effect;
  θ_j: is a subject fixed effect;
  λ_t: is a year fixed effect;
  TB_{cj}: is the teachers' bias
  ϕ_{cj}: is a school and subject specific random element;
  ψ_{icjt}: is an individual random element
  π: captures the effect of teacher's biases on academic outcomes.

- For identification, we rely on the random assignments of teachers and students to classes in high schools in Greece.

Methodology and Estimation Framework

- In our regression we also use:
- **Empirical Bayes estimates** to cope with estimation error arising from sampling variation (when small samples, a few observations can have a large impact).
  - The basic idea is to multiply a noisy estimate of each teacher bias by an estimate of its reliability.
  - Lefgren (2005) showed that using the empirical Bayes estimates as an explanatory variable in a regression yields point estimates that are unaffected by the attenuation bias that would result from using standard OLS estimates.
- **Two-step Bootstrapping Method** because the bias is a generated regressor.
  - Pagan (1984) and Murphy, Topel. (1985): two-step estimation methods yield inconsistent estimates of standard errors in the second-stage regression when they fail to account for the presence of a generated regressor.

Outline

Outline

Effect of 11th Grade Teacher Bias on Blind Score in 12th Grade, Sample of 21 Schools

	BOYS			GIRLS		
	(1)	(2)	(3)	(4)	(5)	(6)
Core Subjects						
	0.047	0.065	0.103	-0.100	-0.085	-0.111
	(0.040)	(0.034)*	(0.037)***	(0.041)**	(0.039)**	(0.043)***
Sample Size	9,406	9,406	9,406	11,844	11,844	11,844
Classics Subjects						
	0.067	0.088	0.067	-0.086	-0.063	-0.067
	(0.089)	(0.064)	(0.063)	(0.046)*	(0.042)	(0.046)
Sample Size	1,817	1,817	1,817	7,080	7,080	7,080
Science Subjects						
	0.036	0.059	0.128	-0.094	-0.082	-0.088
	(0.069)	(0.056)	(0.066)*	(0.047)**	(0.045)*	(0.047)*
Sample Size	3,236	3,236	3,236	3,386	3,386	3,386
Exact Science Subjects						
	0.012	0.009	0.051	-0.107	-0.102	-0.154
	(0.048)	(0.043)	(0.046)	(0.066)	(0.061)*	(0.067)**
Sample Size	5,703	5,703	5,703	3,364	3,364	3,364
Subjects FE	✓	✓	✓	✓	✓	✓
Year FE	✓	✓	✓	✓	✓	✓
School FE		✓	✓			✓
Class FE			✓			✓

Notes: The dependent variable is the 12th grade blind score. The teacher bias is measured in all other classes. The datasets for the core subjects and each track subjects include stacked observations for each subject/exam. Each row presents estimates from separate OLS regressions. All specifications include the students' blind score as a control. The second panel "Classics Subjects" includes relevant exams from the core and the classics track. The third panel "Science Subjects" includes relevant exams from the core and the science track. The forth panel "Exact Science Subjects" includes relevant exams from the core and the exact science track. Standard errors are clustered by class and are reported in parentheses. All scores are standardised z-scores. *,**,*** denotes significance at the 10%,5% and 1% level respectively.

▸ Placebo:Ressufling Biases across Teachers

Effect of 11^{th} and 12^{th} Grade Gender Biases on Students Attendance

Table: Effect of 11^{th} and 12^{th} Grade Gender Biases (measured in all other classes) on Students Total, Excused and Unexcused Absences in 11^{th} and 12^{th} Grade

	Dependent Variable: Total, Excused and Unexcused Absences											
	11^{th} Grade						12^{th} Grade					
	Boys Total Absences	Girls Total Absences	Boys Excused	Girls Excused	Boys Unexcused	Girls Unexcused	Boys Total Absences	Girls Total Absences	Boys Excused	Girls Excused	Boys Unexcused	Girls Unexcused
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
Bias in core subjects	-0.259	0.299	0.078	-0.316	-0.337	0.615	0.361	0.214	0.969	-0.630	-0.551	0.641
	(0.102)**	(0.130)**	(0.058)	(0.099)***	(0.098)***	(0.096)***	(0.405)	(0.358)	(0.382)***	(0.384)*	(0.253)*	(0.254)**
Sample Size	5,842	7,443	5,842	7,443	5,842	7,443	4,834	5,769	4,533	5,433	4,815	5,729
Bias in classics subjects	-0.297	-0.237	0.123	-0.286	-0.420	0.048	0.319	-0.089	0.954	-0.887	-0.777	0.621
	(0.254)	(0.175)	(0.156)	(0.135)*	(0.195)**	(0.144)	(0.208)	(0.181)	(0.353)***	(0.305)**	(0.221)***	(0.225)***
Sample Size	2,804	4,776	2,804	4,776	2,804	4,776	2,528	3,835	2,385	3,638	2,517	3,816
Bias in science subjects	-0.363	0.223	0.100	-0.163	-0.463	0.387	0.830	0.736	1.381	-0.477	-0.438	0.987
	(0.140)***	(0.094)**	(0.089)	(0.072)**	(0.114)**	(0.092)***	(0.647)	(0.529)	(0.611)*	(0.618)	(0.393)	(0.390)***
Sample Size	4,504	5,492	4,504	5,492	4,504	5,492	2,881	3,340	2,657	3,098	2,869	3,306
Bias in exact science subjects	-0.206	0.225	0.101	-0.211	-0.308	0.436	0.605	0.462	0.861	-0.717	-0.227	0.970
	(0.131)	(0.104)**	(0.080)	(0.085)**	(0.114)***	(0.097)***	(0.459)	(0.454)	(0.429)	(0.529)	(0.278)	(0.327)***
Sample Size	4,071	4,812	4,071	4,812	4,071	4,812	3,996	3,824	3,745	3,568	3,980	3,791
Subjects FE	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Year FE	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Class FE	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓

This table presents estimates for the effects of the bias (measured in all other classes) in the related subjects on students' different types of attendance (in hours). The estimation is based on the sample of 21 schools. The outcome variable is the total number of absences in a year (in hours), the excused number of absences in a year (in hours), and the unexcused number of absences in a year (in hours). The estimates are presented separately for the 11^{th} and 12^{th} grade. All estimates have been calculated using an empirical Bayes estimation strategy. All standard errors (reported in parentheses) are calculated using a two-step bootstrapping technique and are clustered at the class level. In the first panel all core subjects are used. The second panel includes only classics subjects. The third panel includes only science subjects. The forth panel includes only exact science subjects. The scores are standardized z-scores. *,**, and *** denote significance at the 10%, 5%, and 1% level.

Outline

Effect of 11^{th} and 12^{th} Grade Gender Biases on Enrollment in Post-Secondary Schooling

Table: Effect of 11^{th} and 12^{th} Grade Gender Biases (measured in all other classes) on Enrollment in Post-Secondary Schooling

Dependent Variable: Dummy variable for Enrollment Status in University				
	11^{th} grade		12^{th} grade	
	Boys	Girls	Boys	Girls
	(1)	(2)	(3)	(4)
Bias core subjects	0.020	-0.047	0.036	-0.028
	(0.005)***	(0.005)***	(0.008)***	(0.008)***
Sample Size	6,845	8,640	5,699	6,861
Bias in classics subjects	0.046	-0.014	0.050	-0.037
	(0.008)***	(0.008)*	(0.015)***	(0.014)***
Sample Size	3,288	5,598	2,998	4,675
Bias in science subjects	0.021	-0.028	0.039	-0.021
	(0.006)***	(0.005)***	(0.011)***	(0.010)*
Sample Size	5,219	6,357	3,376	3,950
Bias in exact science subjects	0.015	-0.032	0.030	-0.026
	(0.007)**	(0.006)***	(0.008)***	(0.010)**
Sample Size	4,795	5,587	4,786	4,495
Subjects FE	✓	✓	✓	✓
Year FE	✓	✓	✓	✓
Class FE	✓	✓	✓	✓

Notes: The outcome variable is the post-secondary enrollment status (1 if enrolled, 0 otherwise). In these regressions, we also control for the blind performance a student gets in each grade (11^{th} grade for columns 1-2 and 12^{th} grade for columns 3-4). Standard errors are clustered by class and are reported in parentheses. The datasets for the core subjects and each track subjects include stacked observations for each subject/exam. Each row presents estimates from a separate regression using an empirical Bayes estimation strategy, for 11^{th} (columns 1-2) and 12^{th} (columns 3-4) grade separately. The empirical Bayes shrinkage factor is the ratio of signal variance to signal plus noise variance. We assume that there is a sampling error problem and the measure of teacher gender bias consists of an error component. Estimating teachers' effects on students' weighted difference between "non-blind" and "blind" scores enables us to separate between the signal and the noise variance. The empirical Bayes estimate for each teacher is a weighted average of the teacher estimated effect and the mean of teacher estimates, where the weight is the empirical Bayes shrinkage factor. Standard errors are clustered using a two-step bootstrapping method. In the first stage, a random sample with replacement is drawn from each class by gender and the corresponding OLS coefficients are obtained. In the second stage, the effect of these new teachers' gender bias measures in 11^{th} grades on students' performance in 12^{th} grade are estimated and the coefficients are stored. This process of two-step bootstrap sampling and estimation is repeated 1,000 times. The standard deviations in the sample of 1,000 observations of coefficients estimates from the second stage regression are the bootstrap standard errors of the point estimates of these coefficients. All specifications include the students' blind score as a control. All scores are standardized z-scores. The first panel "Core Subjects" includes all core subjects. *, **, and *** denote significance at the 10%, 5%, and 1% level.

Descriptive Statistics by University Field of Studies 2003-2011

Field of studies	Mean Enrolment: Girls				Boys				Difference	
	Mean (1)	(sd) (2)	**Mean** (3)	(sd) (4)	**Mean** (5)	(sd) (6)	**Mean** (7)	(sd) (8)	(se) (1)-(5)	(se) (3)-(7)
Exact Science	0.099	(0.298)	0.121	(0.326)	0.223	(0.416)	0.273	(0.445)	-0.125 (0.003)***	0.152 (0.003)***
Science	0.046	(0.209)	0.056	(0.230)	0.037	(0.188)	0.045	(0.207)	0.009 (0.002)***	0.011 (0.002)***
Social Science	0.227	(0.419)	0.278	(0.448)	0.213	(0.409)	0.260	(0.438)	0.014 (0.003)***	0.018 (0.004)***
Humanities	0.273	(0.445)	0.334	(0.472)	0.088	(0.284)	0.108	(0.310)	0.184 (0.003)***	0.226 (0.004)***
Vocational-non academic studies	0.172	(0.377)	0.211	(0.408)	0.258	(0.437)	0.314	(0.464)	-0.086 (0.003)***	-0.104 (0.004)***
Not enrolled in post-secondary studies	0.184	(0.387)			0.181	(0.385)			0.003 (0.003)	

Notes: The sample includes 30,740 female students and 21,496 male students. Columns (3) and (7) refer only to enrollment in university studies. Humanities include the departments of Liberal Arts, Physcology, Journalism, Philosophy, Education, Greek Language, History, Foreign Languages, Home Economics and Law. Social Science includes the departments of Economics, Statistics, Business and Management, Accounting, Political and European studies. Exact Science includes the departments of Mathematics, Engineering, Physics and Computer Science. Science includes the departments of Biology, Chemistry, Medicine, Pharmacy, Veterinary Studies and Dentistry. Vocational-non academic studies include students who enrol in technical education institutes and agricultural studies.

Modelling Choice of University Field of Study

- We model the choice of students in a linear regression.
- We stack the four possible choices as the dependent variable for each student against the teachers' bias in each of the four areas of university studies.
- The dependent variable is a 0/1 indicator, assuming the value of 1 for the observed field of study and a value 0 for the other three possible choices.
- We estimate simple linear probability models since a probit or logit models will yield similar estimates given that we use very large samples.
- We estimate three different specifications:
  -the benchmark includes a year and major fixed effects and national exam score in 11th
  -a second specification includes also high school fixed effect
  -in a third specification we replace the latter with a high school class fixed effect.
  -standard errors are clustered at the class level.

Effect of Teacher's Bias on the Choice of University Field by Gender

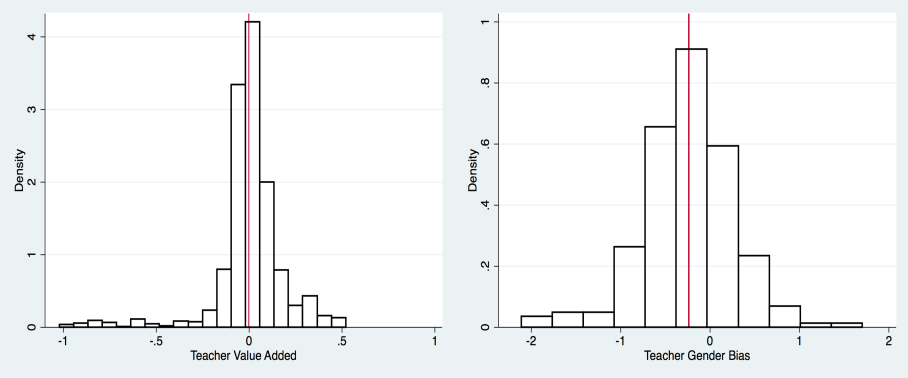
	Dependent Variable: Dummy variable for the choice of University study					
	(1)	(2)	(3)	(4)	(5)	(6)
	BOYS			GIRLS		
	Stack 11th and 12th grades & Grade FE. (2003-2005)					
	0.004 (0.016)	0.006 (0.016)	-0.002 (0.021)	-0.027 (0.013)**	-0.031 (0.014)**	-0.032 (0.018)*
Sample Size	7,916	7,916	7,916	9,957	9,957	9,957
	12th grade (2003-2011)					
	-0.026 (0.016)*	-0.026 (0.016)	-0.033 (0.022)	-0.032 (0.015)**	-0.038 (0.015)**	-0.038 (0.022)*
Sample Size	5,209	5,209	5,209	6,646	6,646	6,646
Year FE	✓	✓	✓	✓	✓	✓
Major FE	✓	✓	✓	✓	✓	✓
Track FE	✓	✓	✓	✓	✓	✓
School FE		✓			✓	
Class FE			✓			✓

Notes: The datasets include stacked observations for each field's related subject/exam. Each row presents estimates from separate OLS regressions. Standard errors are clustered by class and are reported in parentheses. The dependent variable is the choice to study in Social Science, Science, Exact Science or Humanities departments. The scores are standardised and have a zero mean and a standard deviation of one. *,**,*** denotes significance at the 10%,5% and 1% level respectively.

Outline

- We explore the relationship between teacher gender biases and teacher quality.
- We use teachers' value-added (TVA) as a measure of their quality (Chetty et al., 2014 a,b).
- We construct TVA for teachers in the sample of 21 schools using the data for the 2003-2005 period.
- We use students' panel data in 10th-12th grades.
- The teacher gender bias measure we use in this section is the average teacher gender bias overall classes the teacher taught during 2006-2011.
- We restrict the analysis to this period in order to avoid an overlap between the period in which we measure TVA and the period we use to estimate the correlation between teachers' gender bias and TVA.
- This restriction is not a limitation at all because of the high persistency in teachers' biased behavior across classes and years.

Histogram of Teacher Value Added Measure and Average Teacher Bias



Notes: The top panel presents the distribution of the TVA measure, which is weighted by the number of students in the school-year-grade-subject-class year cell. To derive these value added measures we pool the 11^{th} and 12^{th} grade data for the years 2003-2005. We use 10^{th} and 11^{th} grade performance as a prior measure of performance. We follow closely the value added procedure described in [**?**]. This sample includes only students who have non-missing baseline controls to estimate the VA model. TVA is estimated using the baseline control vector, which includes: lagged own-subject scores, student-level characteristics including age, gender, a dummy for being born in the first quarter of the birth year, dummies for whether students expressed a special interest in classics, science or exact science (indicated by the track they have chosen), class size, school-grade enrollment, income as well as school, year, and subject dummies. When prior test scores are missing, we set the prior score equal to 0 and include an indicator for missing data. Student data are from the administrative records of 21 schools in Greece. The structure of the dataset is one observation per teacher-year-grade-subject-class combination. The bottom panel presents the distribution of the average teacher bias measured in all other classes across subjects and classes. To derive a teacher's bias we calculate the average bias a teacher exhibits in all classes they taught in the later years, in particular between 2006 and 2011.

Table: Comparisons of Mean Teacher Value Added for Pro-Boys, Neutral, and Pro-Girls Teachers

	(1)	(2)	(3)
		Panel A	
	Neutral /(sd)	Pro-Boy /(sd)	Difference /(se)
Teacher Value Added	0.053	-0.037	0.090
	(0.132)	(0.222)	(0.032)***
N	58	101	159
		Panel B	
	Neutral /(sd)	Pro-Girl /(sd)	Difference /(se)
Teacher Value Added	0.053	-0.049	0.102
	(0.132)	(0.235)	(0.032)***
N	58	259	317

Notes: We pool data on test scores for 11^{th} and 12^{th} grades for the period 2003-2005. The TVA measures are derived following the procedure described in Chetty et al (2014). Pro-boy teacher exhibit bias larger than or equal to 0.10. Pro-girl teachers exhibit a bias smaller than or equal to -0.10. We define as neutral teachers who exhibit bias that is larger than or equal to -0.10 and smaller than 0.10. The teacher bias measures are derived as the average bias across subjects, grades and classes a teacher exhibits in the 2006-2011 sample. Our sample includes only students who have non-missing baseline controls to estimate the VA model. Our baseline VA model controls for a rich set of student demographics and other variables, as well as teacher, class, and school level variables. In particular, our baseline VA model controls for a student's gender, age, a dummy whether the student is born in the first quarter of a calendar year, his/her lagged performance in the same subject, class size, school-level-grade enrollment, a dummy that takes the value of 1 if the teacher is female and 0 otherwise, how many classes each teacher taught in the sample (our proxy for a teacher's experience), students' average performance in all other classes taught by the same teacher in the sample and neighborhood income. When the prior test score is missing, we set the prior score equal to 0 and include an indicator for missing data.

Table: Correlations Between Teacher Bias And Teacher Quality (TVA) for Pro-Girl and Pro-Boy Teachers (Spline Variables)

	Dependent Variable: Teacher Quality (Measured by TVA)			
	(1)	(2)	(3)	(4)
Spline for Pro-Boys Teachers	-0.113	-0.113	-0.112	-0.112
	(0.041)***	(0.040)***	(0.040)***	(0.043)**
Spline for Pro-Girls Teachers	0.049	0.049	0.050	0.050
	(0.026)*	(0.027)*	(0.028)*	(0.029)*
Female Teacher		-0.006	-0.007	-0.007
		(0.018)	(0.018)	(0.019)
Class Size			-0.002	-0.002
			(0.003)	(0.003)
Teacher Experience				-0.0001
				(0.001)
Year FE	✓	✓	✓	✓
School FE	✓	✓	✓	✓
Grade FE	✓	✓	✓	✓
Observations	418	418	418	418

Notes: The "Spline for Pro-Girl Teachers" takes the actual negative teacher bias values, and the value of zero for the positive ones. The "Spline for Pro-Boy Teachers" takes the actual positive teacher bias values, and the value zero for the negative ones. The teacher gender bias measures the average bias a teacher exhibits in different subjects and classes in the 2006-2011 sample. We include the two splines simultaneously in the regression. The outcome variable is the teacher value added derived using the 2003-2005 sample and is described in details in the text. "Teacher experience" measures the different combination of classes and subjects a teacher has taught in 11^{th} and 12^{th} grades in the sample period 2003-2011. Standard errors are clustered by school and year and are reported in parentheses. *, **, and *** denote significance at the 10%, 5%, and 1% level.

Table: Correlations Between Teacher Gender Bias And Teacher Quality (Measured by TVA)

Dependent Variable: Teacher Quality (Measured by TVA)				
	(1)	(2)	(3)	(4)
Pro-Girl Teacher Dummy	-0.035	-0.036	-0.037	-0.037
	(0.017)**	(0.017)**	(0.018)**	(0.018)**
Pro-Boy Teacher Dummy	-0.031	-0.032	-0.032	-0.031
	(0.018)*	(0.018)*	(0.018)*	(0.018)*
Female Teacher		-0.007	-0.008	-0.007
		(0.018)	(0.018)	(0.019)
Class Size			-0.003	-0.003
			(0.003)	(0.003)
Teacher Experience				0.001
				(0.001)
Year FE	✓	✓	✓	✓
School FE	✓	✓	✓	✓
Grade FE	✓	✓	✓	✓
Observations	418	418	418	418

Notes: The "Pro-Girl Teacher Dummy" takes the value of one if the teacher exhibits a bias that is smaller than or equal to -0.10. The "Pro-Boy Teacher Dummy" takes the value of one if the teacher exhibits a bias that is above 0.10. We define as neutral teachers those who have a bias that is larger than or equal to -0.10 and smaller than 0.10. The omitted category in the regression is neutral teachers. The teacher bias is calculated in the sample period of 2006-2011. The outcome variable is the TVA derived using the 2003-2005 sample and described in the text. "Teacher experience" measures the different combination of classes and subjects a teacher has taught in 11^{th} and 12^{th} grades in the sample period 2003-2011. Standard errors are clustered by school and year and are reported in parentheses. *, **, and *** denote significance at the 10%, 5%, and 1% level.

Outline

Conclusion

- This the first study to present evidence on the effect of teachers' grading biases using persistent measures of stereotypical behavior in schools.
- Using panel data on teachers' class assignment history we find high correlation in teachers biases across their different classes over time.
- In this paper we extend the analysis beyond test scores and show that teachers' biases in specific courses lower the likelihood that students enroll in a related field of study at the university.
- Our results suggest that less effective teachers can harm their students twice, by being a bad teacher and by discriminating against one of the genders
- Implications: there might be a scope for training teachers to be aware of gender stereotypes that lead to teachers' classroom differential behavior towards students by gender and to grading biases that have long term consequences.

Correlations between Biases by Same and Different teachers, by subject, Core, 11th grade

Dependent Variable: Teacher bias in own subject, Classics Track				
	Different Teachers		Same Teachers	
Teacher bias	Modern Greek	in Geometry	Modern Greek	Geometry
	(1)	(2)	(3)	(4)
History	0.057		0.856	
	(0.088)		(0.139)***	
N	119		36	
Algebra		0.020		0.831
		(0.128)		(0.057)***
N		63		85

Note: The table presents estimated correlation coefficient of teachers' biases measures by subjects of instruction in Classics Track The (OLS) estimated coefficients in columns 1-2 is the between biases measures of the same teachers who instruct students fron the same class two subjects and the (OLS) estimated coefficients in columns 3-4 is the between biases measures of different teacher who instruct students from the same class in two subjects. Each estimate comes from a separate OLS regression. Standard error are reported in parentheses. *,**, and *** denotes significance at the 10%, 5%, and 1% level, respectively.

Go Back

Correlations between Biases by Same and Different teachers, Classics, 11th grade

Dependent Variable: Teacher bias in own subject, Classics Track				
	Different Teachers		Same Teachers	
	Anc.Greek	Philosophy	Anc.Greek	Philosophy
	(1)	(2)	(3)	(4)
Teacher bias	-0.025	-0.024	0.706	0.714
	(0.100)	(0.100)	(0.112)***	(0.110)***
N	89	89	28	28
Subject FE	X	✓	X	✓

Note: The table presents estimated correlation coefficient of teachers' biases measures by subjects of instruction in Classics Track. The (OLS) estimated coefficients in columns 1-2 is the between biases measures of the same teachers who instruct students from the same class two subjects and the (OLS) estimated coefficients in columns 3-4 is the between biases measures of different teachers who instruct students from the same class in two subjects. Each estimate comes from a separate OLS regression. Standard errors are reported in parentheses. *,**, and *** denotes significance at the 10%, 5%, and 1% level, respectively.

Go Back

Dependent Variable: Teacher bias in own subject, Science				
	Different Teachers		Same Teachers	
	(1)	(2)	(3)	(4)
Teacher bias	0.470	0.470	0.919	1.101
	(0.090)***	(0.092)***	(0.112)***	(0.090)***
N	93	93	8	8
Subject FE	X	✓	X	✓

Note: The table includes stacked observations for the teacher bias in each subject in the Science Track. The (OLS) estimated coefficients in columns 1-2 is the between biases measures of the same teachers who instruct students from the same class two subjects and the (OLS) estimated coefficients in columns 3-4 is the between biases measures of different teachers who instruct students from the same class in two subjects. Each estimate comes from a separate OLS regression. Standard errors are reported in parentheses. *,**, and *** denotes significance at the 10%, 5%, and 1% level, respectively.

Go Back

Dependent Variable: Teacher bias in own subject, Core				
	Different Teachers		Same Teachers	
Teacher bias	Modern Greek	Physics	Modern Greek	Physics
	(1)	(2)	(3)	(4)
History	0.108	0.029	0.608	
	(0.155)	(0.081)	(0.136)***	
N	82	132	52	
Biology	0.054	0.094		0.225
	(0.045)	(0.070)		(0.110)*
N	224	114		21

Note: The table presents the estimated correlation coefficient of teachers' biases measures by subjects of instruction. The estimates in each row in columns 1-4 are the correlation coefficients between bias measures using the sample of all teachers (same or different teachers for each two subjects), from separate OLS regressions. The estimated coefficients in each row in columns 5-8 are similar to those in columns 1-4, but high-school fixed effects are included in the regressions. *,**,*** denotes significance at the 10%, 5% and 1% level respectively. The (OLS) estimated coefficients in columns 9-12 is the between biases measures of the same teachers who instruct students from the same class two subjects and the (OLS) estimated coefficients in columns 13-16 is the between biases measures of different teachers who instruct students from the same class in two subjects. Each estimate comes from a separate OLS regression. Standard errors are reported in parentheses.

Correlations between Biases by Same and Different teachers, Core, 12th grade

Dependent Variable: Teacher bias in own subject				
	Different Teachers		Same Teachers	
	(1)	(2)	(3)	(4)
Teacher bias	0.119	0.126	0.536	0.527
	(0.033)***	(0.034)***	(0.094)***	(0.095)***
N	1,198	1,198	87	87
Subject FE	X	✓	X	✓

Note: The table includes stacked observations for the teacher bias in each subject in the Exact Science Track. The (OLS) estimated coefficients in columns 1-2 is the between biases measures of the same teachers who instruct students from the same class two subjects and the (OLS) estimated coefficients in columns 3-4 is the between biases measures of different teachers who instruct students from the same class in two subjects. Each estimate comes from a separate OLS regression. Standard errors are reported in parentheses. *,**, and *** denotes significance at the 10%, 5%, and 1% level, respectively.

Go Back

Table: Placebo Estimation: Randomly Shuffle Biases Across Teachers of the Same Subjects Within Schools

Dependent Variable: Subsequent Blind score in 12^{th} grade	Boys			Girls		
	(1)	(2)	(3)	(4)	(5)	(6)
Core Subjects						
	-0.010	-0.022	-0.004	0.004	-0.004	0.019
	(0.027)	(0.027)	(0.025)	(0.022)	(0.022)	(0.021)
Sample Size	8,217	8,217	8,217	10,347	10,347	10,347
Classics Track Subjects						
	-0.111	-0.093	0.026	-0.048	0.063	0.030
	(0.083)	(0.084)	(0.090)	(0.043)	(0.050)	(0.068)
Sample Size	1,391	1,391	1,391	5,443	5,443	5,443
Science Track Subjects						
	0.062	0.001	-0.045	0.103	0.062	0.010
	(0.058)	(0.056)	(0.051)	(0.052)	(0.049)	(0.042)
Sample Size	4,492	4,492	4,492	4,932	4,932	4,932
Exact Science Track Subjects						
	0.011	-0.010	0.005	-0.109	-0.075	-0.078
	(0.042)	(0.043)	(0.039)	(0.055)	(0.054)	(0.052)
Sample Size	5,119	5,119	5,119	3,0130	3,030	3,030
Subjects FE	✓	✓	✓	✓	✓	✓
Year FE	✓	✓	✓	✓	✓	✓
School FE		✓			✓	
Class FE			✓			✓

Notes: We randomly reshuffle teacher biases within schools across teachers who teach the same subjects. The estimation is based on the sample of 21 schools. The datasets for the core subjects and each track subjects include stacked observations for each subject/exam. Each row presents estimates from a separate regression using an empirical Bayes estimation strategy, for 11^{th} (columns 1-2) and 12^{th} (columns 3-4) grade separately. The empirical Bayes shrinkage factor is the ratio of signal variance to signal plus noise variance. We assume that there is a sampling error problem and the measure of teacher gender bias consists of an error component. Estimating teachers' effects on students' weighted difference between "non-blind" and "blind" scores enables us to separate between the signal and the noise variance. The empirical Bayes estimate for each teacher is a weighted average of the teacher estimated effect and the mean of teacher estimates, where the weight is the empirical Bayes shrinkage factor. Standard errors are clustered using a two-step bootstrapping method. In the first stage, a random sample with replacement is drawn from each class by gender and the corresponding OLS coefficients are obtained. In the second stage, the effect of these new teachers' gender bias measures in 11^{th} grades on students' performance in 12^{th} grade are estimated and the coefficients are stored. This process of two-step bootstrap sampling and estimation is repeated 1,000 times. The standard deviations in the sample of 1,000 observations of coefficients estimates from the second stage regression are the bootstrap standard errors of the point estimates of these coefficients. All specifications include the students' blind score as a control. All scores are standardized z-scores. The first panel "Core Subjects" includes all core subjects. The second panel "Classics Subjects" includes relevant exams from the core (history and modern Greek) and all the classics track subjects. The third panel "Science Subjects" includes relevant exams from the core (Algebra, Geometry and physics) and all the science track subjects. The forth panel "Exact Science Subjects" includes relevant exams from the core (Algebra, Geometry and physics) and all the exact science track subjects. *, **, and *** denote significance at the 10%, 5%, and 1% level.