## Online Appendix

Not For Publication

Table A1: Descriptive Statistics, Full Sample

|  | Mean | Std.Dev | Min | Max |
| :---: | :---: | :---: | :---: | :---: |
|  | (1) | (2) | (3) | (4) |
| I. Student Characteristics |  |  |  |  |
| $11^{\text {th }}$ Grade |  |  |  |  |
| Female | 0.563 | 0.496 | 0 | 1 |
| Total Absences (in Hours) | 50.851 | 27.425 | 1 | 450 |
| Prop. of Students in Classics | 0.366 | 0.482 | 0 | 1 |
| Prop. of Students in Science | 0.280 | 0.449 | 0 | 1 |
| Prop. of Students in Exact Science | 0.344 | 0.475 | 0 | 1 |
| GPA $11^{\text {th }}$ Grade | 72.335 | 14.159 | 0 | 100 |
| $12^{\text {th }}$ Grade |  |  |  |  |
| Total Absences (in Hours) | 73.443 | 30.752 | 1 | 208 |
| Prop. of Students in Classics | 0.369 | 0.483 | 0 | 1 |
| Prop. of Students in Science | 0.159 | 0.366 | 0 | 1 |
| Prop. of Students in Exact Science | 0.463 | 0.499 | 0 | 1 |
| GPA $12{ }^{\text {th }}$ Grade | 76.976 | 12.526 | 44 | 100 |

## II. School Characteristics

| Private School | 0.037 | 0.190 | 0 | 1 |
| :--- | :---: | :---: | :---: | :---: |
| Experimental School | 0.044 | 0.207 | 0 | 1 |
| Public School | 0.919 | 0.275 | 0 | 1 |
| Urban | 0.896 | 0.306 | 0 | 1 |
| Postcode Income (in 2009 Euro) | 22455.029 | 7945.331 | 11784.5 | 66521.38 |


| II. University Enrollment Characteristics |  |  |  |  |
| :--- | :---: | :---: | :---: | :---: |
| National Exams Average Score | 64.987 | 20.178 | 10.35 | 99.3 |
| Retake the National Exams | 0.113 | 0.317 | 0 | 1 |
| Number of University Departments in Preference List | 25.014 | 22.071 | 1 | 257 |
| Rank in Preference List of the Actual University Attended | 8.399 | 10.616 | 1 | 242 |
| Enrollment in University or Vocational Schooling | 0.817 | 0.386 | 0 | 1 |
| Exact Science Department | 0.153 | 0.360 | 0 | 1 |
| Science Apartment | 0.042 | 0.201 | 0 | 1 |
| Humanities Department | 0.193 | 0.395 | 0 | 1 |
| Social Science Department | 0.220 | 0.414 | 0 | 1 |
| Vocational Schooling | 0.209 | 0.407 | 0 | 1 |

Note: All statistics reported include students who graduate from high school between 2003 and 2011. Total absences are measured in hours per year. GPA11 and GPA12 include the average over the school exam scores in the first and second term, in $11^{\text {th }}$ and $12^{\text {th }}$ grade. The full sample of schools is used. There are three types of schools in the sample: public, private, and experimental schools. Experimental are public schools and school admission is based on a lottery for the sample years. A school is located in an urban area if the area has more than 20,000 inhabitants. Postcode income is expressed in 2009 Euro.

Table A2: Random Assignment: Balancing Test of Student Characteristics and Teacher Characteristics by Student Gender

|  | Girls |  |  | Boys |  |  |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: |
|  | (1) | (2) | (3) | (4) | (5) | (6) |
|  | Student Prior Performance | Student Age | $\begin{aligned} & \text { Class } \\ & \text { Size } \end{aligned}$ | Student Prior Performance | Student Age | $\begin{aligned} & \text { Class } \\ & \text { Size } \end{aligned}$ |
| Female Teacher ( $=1$ if Female) | $\begin{gathered} 0.058 \\ (0.039) \\ \hline \end{gathered}$ | $\begin{gathered} 0.017 \\ (0.012) \\ \hline \end{gathered}$ | $\begin{gathered} -0.258^{*} \\ (0.143) \\ \hline \end{gathered}$ | $\begin{gathered} 0.029 \\ (0.042) \\ \hline \end{gathered}$ | $\begin{aligned} & -0.014 \\ & (0.013) \\ & \hline \end{aligned}$ | $\begin{array}{r} -0.128 \\ (0.193) \\ \hline \end{array}$ |
| Obs. | 10,351 | 10,351 | 10,351 | 7,612 | 7,612 | 7,612 |
| Teacher Previous Year Quality | $\begin{gathered} 0.100 \\ (0.147) \\ \hline \end{gathered}$ | $\begin{aligned} & -0.054 \\ & (0.050) \\ & \hline \end{aligned}$ | $\begin{gathered} 0.726 \\ (0.615) \\ \hline \end{gathered}$ | $\begin{gathered} 0.228 \\ (0.177) \\ \hline \end{gathered}$ | $\begin{aligned} & -0.020 \\ & (0.042) \\ & \hline \end{aligned}$ | $\begin{array}{r} 1.146 \\ (0.954) \\ \hline \end{array}$ |
| Obs. | 10,349 | 10,349 | 10,349 | 7,612 | 7,612 | 7,612 |
| Teacher Experience | $\begin{gathered} 0.006 \\ (0.004) \end{gathered}$ | $\begin{aligned} & -0.002^{*} \\ & (0.001) \end{aligned}$ | $\begin{gathered} 0.008 \\ (0.014) \end{gathered}$ | $\begin{aligned} & -0.000 \\ & (0.003) \end{aligned}$ | $\begin{gathered} 0.000 \\ (0.001) \end{gathered}$ | $\begin{aligned} & -0.020 \\ & (0.022) \end{aligned}$ |
| Obs. | 10,351 | 10,351 | 10,351 | 7,612 | 7,612 | 7,612 |
| Year FE | Yes | Yes | Yes | Yes | Yes | Yes |
| Grade FE | Yes | Yes | Yes | Yes | Yes | Yes |
| Class FE | Yes | Yes | Yes | Yes | Yes | Yes |

Notes: The table presents the estimated effects from separate regressions of each of the student pre-assignment characteristics and prior test scores on teacher gender ( $1=$ female), teacher previous year quality, and teacher experience. Each estimate in this table is generated from a different regression. Student test scores are standardized and have a zero mean and a standard deviation of one. All regressions condition on subject fixed effects, year fixed effects, grade fixed effects, and class fixed effects. Robust standard errors clustered at the class level are reported in parentheses.

Table A3: Random Assignment: Balancing Test of Teacher Characteristics and Student Characteristics by Student Gender

|  | Girls BoysFemale Teacher |  | Girls Boys <br> Previous TVA |  | Girls BoysTeacher Experience |  |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: |
|  | (1) | (2) | (3) | (4) | (5) | (6) |
| Lagged Test Score | $\begin{gathered} 0.006 \\ (0.004) \end{gathered}$ | $\begin{gathered} 0.003 \\ (0.004) \end{gathered}$ | $\begin{gathered} 0.000 \\ (0.001) \end{gathered}$ | $\begin{gathered} 0.001 \\ (0.001) \end{gathered}$ | $\begin{gathered} 0.062 \\ (0.040) \end{gathered}$ | $\begin{aligned} & \hline-0.007 \\ & (0.043) \end{aligned}$ |
| Class Size | $\begin{aligned} & -0.010^{*} \\ & (0.006) \end{aligned}$ | $\begin{aligned} & -0.004 \\ & (0.007) \end{aligned}$ | $\begin{gathered} 0.001 \\ (0.001) \end{gathered}$ | $\begin{gathered} 0.002 \\ (0.002) \end{gathered}$ | $\begin{gathered} 0.037 \\ (0.071) \end{gathered}$ | $\begin{aligned} & -0.094 \\ & (0.102) \end{aligned}$ |
| Age | $\begin{gathered} 0.006 \\ (0.004) \end{gathered}$ | $\begin{aligned} & -0.003 \\ & (0.002) \end{aligned}$ | $\begin{aligned} & -0.001 \\ & (0.001) \end{aligned}$ | $\begin{gathered} 0.000 \\ (0.000) \end{gathered}$ | $\begin{gathered} -0.082^{*} \\ (0.049) \end{gathered}$ | $\begin{aligned} & -0.006 \\ & (0.030) \end{aligned}$ |
| Observations | 10,351 | 7,612 | 10,349 | 7,612 | 10,351 | 7,612 |
| Grade FE | Yes | Yes | Yes | Yes | Yes | Yes |
| Year FE | Yes | Yes | Yes | Yes | Yes | Yes |
| Class FE | Yes | Yes | Yes | Yes | Yes | Yes |
| Mean of Y Variable | 0.35 | 0.32 | -0.00 | -0.00 | 9.92 | 10.00 |
| F-statistics | 3.43 | 1.12 | 1.03 | 1.05 | 1.21 | 0.30 |
| P -value of the F-model | 0.02 | 0.34 | 0.38 | 0.37 | 0.31 | 0.83 |

Notes: All estimates in each column are generated from the same regression. The table reports OLS estimates from separate regressions of each of the remaining teacher characteristics on all student preassignment characteristics and class size, separately by student gender. Estimated effects for female students are shown in column (1) and for male students in column (2). A student's pre-assignment characteristics are prior test scores, and age. The remaining teacher characteristics are the following: a teacher's gender ( $1=$ female), a teacher previous year quality (measured by previous year's value added), and experience (measured by how many times a teacher teaches in the whole sample period 2003-2011). All regressions include class fixed effects, year fixed effects and grade fixed effects. Robust standard errors clustered at the class level are reported in parentheses.

Table A4: Effect of $12^{\text {th }}$ Grade Teacher Bias on National Exam Score in $11^{\text {th }}$ Grade

|  | Boys |  |  |  | Girls |  |  |  |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
|  | (1) | (2) | (3) | (4) | (5) | (6) | (7) | (8) |
| Panel A: All Subjects |  |  |  |  |  |  |  |  |
|  | 0.189 | -0.032 | -0.016 | -0.072 | -0.022 | -0.111 | 0.171 | 0.262 |
|  | (0.260) | (0.234) | (0.248) | (0.358) | (0.223) | (0.196) | (0.210) | (0.325) |
| Sample Size | 15,932 | 15,932 | 15,932 | 15,908 | 18,712 | 18,712 | 18,712 | 18,670 |

## Panel B: Core Subjects

Sample Size

| 0.368 | 0.190 | -0.024 | -0.095 |
| :---: | :---: | :---: | :---: |
| $(0.260)$ | $(0.234)$ | $(0.248)$ | $(0.358)$ |
|  |  |  |  |
| 5,438 | 5,438 | 5,438 | 5,434 |


| 0.160 | 0.052 | -0.262 | 0.274 |
| :--- | :--- | :--- | :--- |
| $(0.223)$ | $(0.196)$ | $(0.210)$ | $(0.325)$ |
|  |  |  |  |
| 6,763 | 6,763 | 6,763 | 6,756 |

## Panel C: Classics Subjects

Sample Size

| 0.100 | -0.040 | -0.035 | -1.011 | -0.346 | -0.436 | -0.219 | 0.174 |
| :--- | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
| $(0.445)$ | $(0.401)$ | $(0.374)$ | $(1.042)$ | $(0.277)$ | $(0.249)$ | $(0.256)$ | $(0.467)$ |
|  |  |  |  |  |  |  |  |
| 2,517 | 2,517 | 2,517 | 2,389 | 8,760 | 8,760 | 8,760 | 8,745 |

## Panel D: Science Subjects

Panel E: Exact Science Subjects

|  | $\begin{aligned} & 0.061 \\ & (0.279) \end{aligned}$ | $\begin{gathered} -0.082 \\ (0.263) \end{gathered}$ | $\begin{aligned} & 0.028 \\ & (0.288) \end{aligned}$ | $\begin{gathered} -0.314 \\ (0.403) \end{gathered}$ | $\begin{aligned} & 0.201 \\ & (0.295) \end{aligned}$ | $\begin{aligned} & 0.016 \\ & (0.264) \end{aligned}$ | $\begin{aligned} & 0.492 \\ & (0.269) \end{aligned}$ | $\begin{aligned} & 0.892 \\ & (0.605) \end{aligned}$ |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
| Sample Size | 9,606 | 9,606 | 9,606 | 9,548 | 5,374 | 5,374 | 5,374 | 5,246 |
| Subject FE | $\checkmark$ | $\checkmark$ | $\checkmark$ |  | $\checkmark$ | $\checkmark$ | $\checkmark$ |  |
| Year FE | $\checkmark$ | $\checkmark$ | $\checkmark$ |  | $\checkmark$ | $\checkmark$ | $\checkmark$ |  |
| School FE |  | $\checkmark$ |  |  |  | $\checkmark$ |  |  |
| Class FE |  |  | $\checkmark$ |  |  |  | $\checkmark$ |  |
| School-by-Subject-by-Year FE |  |  |  | $\checkmark$ |  |  |  | $\checkmark$ |

Notes: The datasets for the core subjects and each of the track subjects include stacked observations for each subject/exam. The estimation is based on the sample of 21 schools. Each row presents estimates from a separate regression using an empirical Bayes estimation strategy. The standard errors are reported in parentheses and are corrected using a two-step bootstrapping method. This process of two-step bootstrap sampling and estimation is repeated 1,000 times. All specifications include the teacher's gender as control. A student's $11^{\text {th }}$ grade subject-specific national exam performance is used as an outcome. The variable of interest is the teacher gender bias in $12^{\text {th }}$ grade. The first panel includes all subjects. The second panel includes all core subjects. Panel C includes all classics track subjects. Panel D includes all science track subjects. Panel E includes all exact science track subjects. Standard errors are also clustered by class.

Table A5: Effect of $12^{\text {th }}$ Grade Teacher Bias on School Exam Score in $11^{\text {th }}$ Grade

|  | Boys |  |  |  | Girls |  |  |  |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
|  | (1) | (2) | (3) | (4) | (5) | (6) | (7) | (8) |
| Panel A: All Subjects |  |  |  |  |  |  |  |  |
|  | 0.368 | 0.166 | 0.015 | -0.130 | 0.112 | 0.067 | 0.025 | -0.014 |
|  | (0.169) | (0.131) | (0.134) | (0.236) | (0.122) | (0.111) | (0.125) | (0.185) |
| Sample Size | 16,126 | 16,126 | 16,126 | 16,102 | 18,955 | 18,955 | 18,955 | 18,913 |
| Panel B: Core Subjects |  |  |  |  |  |  |  |  |
|  | 0.172 | 0.042 | -0.184 | -0.158 | -0.016 | -0.021 | -0.280 | -0.028 |
|  | (0.208) | (0.182) | (0.179) | (0.242) | (0.164) | (0.153) | (0.175) | (0.188) |
| Sample Size | 5,438 | 5,438 | 5,438 | 5,434 | 6,763 | 6,763 | 6,763 | 6,756 |
| Panel C: Classics Subjects |  |  |  |  |  |  |  |  |
|  | 0.137 | 0.001 | -0.091 | -0.757 | -0.081 | -0.078 | -0.029 | 0.062 |
|  | (0.273) | (0.229) | (0.217) | (0.737) | (0.168) | (0.158) | (0.174) | (0.284) |
| Sample Size | 2,559 | 2,559 | 2,559 | 2,431 | 8,882 | 8,882 | 8,882 | 8,867 |
| Panel D: Science Subjects |  |  |  |  |  |  |  |  |
|  | 0.513 | 0.382 | 0.042 | 0.360 | 0.133 | 0.094 | -0.087 | -0.354 |
|  | (0.226) | (0.219) | (0.254) | (0.527) | (0.169) | (0.162) | (0.193) | (0.445) |
| Sample Size | 3,680 | 3,680 | 3,680 | 3,595 | 4,453 | 4,453 | 4,453 | 4,351 |
| Panel E: Exact Science Subjects |  |  |  |  |  |  |  |  |
|  | 0.279 | 0.123 | 0.048 | -0.397 | 0.192 | 0.039 | 0.058 | -0.263 |
|  | (0.187) | (0.148) | (0.150) | (0.276) | (0.165) | (0.131) | (0.135) | (0.338) |
| Sample Size | 9,685 | 9,685 | 9,685 | 9,627 | 5,412 | 5,412 | 5,412 | 5,284 |
| Subject FE | $\checkmark$ | $\checkmark$ | $\checkmark$ |  | $\checkmark$ | $\checkmark$ | $\checkmark$ |  |
| Year FE | $\checkmark$ | $\checkmark$ | $\checkmark$ |  | $\checkmark$ | $\checkmark$ | $\checkmark$ |  |
| School FE |  | $\checkmark$ |  |  |  | $\checkmark$ |  |  |
| Class FE |  |  | $\checkmark$ |  |  |  | $\checkmark$ |  |
| School-by-Subject-by-Year FE |  |  |  | $\checkmark$ |  |  |  | $\checkmark$ |

Notes: The datasets for the core subjects and each of the track subjects include stacked observations for each subject/exam. The estimation is based on the sample of 21 schools. Each row presents estimates from a separate regression using an empirical Bayes estimation strategy. The standard errors are reported in parentheses and are corrected using a two-step bootstrapping method. This process of two-step bootstrap sampling and estimation is repeated 1,000 times. All specifications include the teacher's gender as control. A student's $11^{\text {th }}$ grade second-semester school exam performance is used as an outcome. The variable of interest is the teacher gender bias in $12^{t h}$ grade. The first panel includes all subjects. The second panel includes all core subjects. Panel C includes all classics track subjects. Panel D includes all science track subjects. Panel E includes all exact science track subjects. Standard errors are also clustered by class.

Table A6: Effect of $11^{\text {th }}$ Grade Teacher Bias on National Test Scores in $12^{\text {th }}$ Grade for STEM and non-STEM subjects

Dependent Variable: Test score in $12^{\text {th }}$ grade national exams

|  | Boys |  |  |  | Girls |  |  |  |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
|  | (1) | (2) | (3) | (4) | (5) | (6) | (7) | (8) |
| Panel A: STEM Subjects |  |  |  |  |  |  |  |  |
|  | 0.162 | 0.120 | 0.178 | 0.113 | 0.045 | -0.017 | -0.085 | -0.011 |
|  | (0.044) | (0.041) | (0.044) | (0.062) | (0.038) | (0.037) | (0.042) | (0.052) |
| Sample Size | 9,000 | 9,000 | 9,000 | 8,992 | 9,425 | 9,425 | 9,425 | 9,410 |
| Panel B: Non-STEM Subjects |  |  |  |  |  |  |  |  |
|  | 0.033 | 0.135 | 0.197 | 0.117 | -0.159 | -0.025 | -0.030 | -0.203 |
|  | (0.060) | (0.055) | (0.066) | (0.068) | (0.046) | (0.044) | (0.048) | (0.062) |
| Sample Size | 3,274 | 3,274 | 3,274 | 3,248 | 5,588 | 5,588 | 5,588 | 5,586 |
| Subject FE | $\checkmark$ | $\checkmark$ | $\checkmark$ |  | $\checkmark$ | $\checkmark$ | $\checkmark$ |  |
| Year FE | $\checkmark$ | $\checkmark$ | $\checkmark$ |  | $\checkmark$ | $\checkmark$ | $\checkmark$ |  |
| School FE |  | $\checkmark$ |  |  |  | $\checkmark$ |  |  |
| Class FE |  |  | $\checkmark$ |  |  |  | $\checkmark$ |  |
| School-by-Subject-by-Year FE |  |  |  | $\checkmark$ |  |  |  | $\checkmark$ |

Notes: As non-STEM subjects we have grouped together all non-STEM subjects in the core (modern Greek and history) and all subjects in the classics track. As STEM subjects we have grouped together all STEM subjects in the core (Physics, Algebra and Geometry) and all subjects in the science and exact science tracks. The estimation is based on the sample of 21 schools. Each row presents estimates from a separate regression using an empirical Bayes estimation strategy. The standard errors are also corrected using a two-step bootstrapping method that we describe in the text. This process of two-step bootstrap sampling and estimation is repeated 1,000 times. All specifications include the students' $11^{\text {th }}$ second-semester school exam performance and the teacher's gender as controls. All scores are standardized z-scores. STEM subjects include algebra, geometry, physics from the core, and all science and exact science track subjects. Non-STEM subjects include history, and modern Greek from the core, and all classics track subjects. Standard errors are clustered by class.

Table A7: Descriptive Statistics for the Teacher Bias by Teacher Gender, Sample of 21 Schools

|  | Male Teachers |  | Female Teachers |  | Diff. | s.e. |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: |
|  | Mean | Std.Dev | Mean | Std.Dev |  |  |
|  | (1) | (2) | (3) | (4) | (5) | (6) |
| Panel A: $11^{\text {th }}$ Grade |  |  |  |  |  |  |
| I. Core Subjects |  |  |  |  |  |  |
| Algebra | -0.155 | 0.228 | -0.073 | 0.307 | -0.081 | 0.045 |
| Geometry | -0.212 | 0.215 | -0.076 | 0.308 | -0.136 | 0.046 |
| History | -0.185 | 0.354 | -0.080 | 0.380 | -0.105 | 0.061 |
| Modern Greek | -0.150 | 0.402 | -0.060 | 0.303 | -0.090 | 0.065 |
| Physics | -0.149 | 0.223 | -0.163 | 0.315 | 0.014 | 0.044 |
| II. Classics Track |  |  |  |  |  |  |
| Ancient Greek | -0.216 | 0.440 | -0.305 | 0.428 | 0.089 | 0.134 |
| Latin | -0.137 | 0.365 | -0.183 | 0.419 | 0.046 | 0.119 |
| Philosophy | -0.170 | $0.389$ | -0.153 | $0.411$ | -0.018 | 0.124 |
| III. Science Track |  |  |  |  |  |  |
| Chemistry | -0.154 | 0.203 | -0.079 | 0.306 | -0.075 | 0.100 |
| Mathematics | -0.148 | 0.175 | -0.006 | 0.291 | -0.143 | 0.084 |
| Physics | -0.035 | 0.242 | -0.039 | 0.302 | 0.004 | 0.094 |
| IV. Exact Science Track |  |  |  |  |  |  |
| Mathematics | -0.149 | 0.236 | -0.045 | 0.286 | -0.104 | 0.091 |
| Physics | -0.154 | 0.404 | -0.144 | 0.239 | -0.010 | 0.102 |
| Technology and Computers | -0.318 | 0.191 | -0.250 | 0.385 | -0.068 | 0.118 |
| Panel B: $12^{\text {th }}$ Grade |  |  |  |  |  |  |
| I. Core Subjects |  |  |  |  |  |  |
| Biology | -0.035 | 0.526 | -0.211 | 0.487 | 0.175 | 0.070 |
| History | -0.238 | 0.366 | -0.203 | 0.367 | -0.035 | 0.064 |
| Mathematics | -0.159 | 0.281 | -0.153 | 0.363 | -0.007 | 0.047 |
| Modern Greek | -0.153 | 0.351 | -0.132 | 0.292 | -0.022 | 0.040 |
| Physics | -0.156 | 0.235 | -0.187 | 0.356 | 0.031 | 0.055 |
| II. Classics Track |  |  |  |  |  |  |
| Ancient Greek | -0.154 | 0.455 | -0.186 | 0.394 | 0.031 | 0.088 |
| History | -0.192 | 0.360 | -0.244 | 0.381 | 0.053 | 0.080 |
| Latin | -0.192 | 0.453 | -0.196 | 0.397 | 0.004 | 0.097 |
| Literature | -0.181 | 0.408 | -0.251 | 0.382 | 0.070 | 0.087 |
| III. Science Track |  |  |  |  |  |  |
| Biology | 0.068 | 0.296 | -0.287 | 0.409 | 0.355 | 0.155 |
| Chemistry | -0.254 | 0.582 | -0.181 | 0.347 | -0.073 | 0.163 |
| Mathematics | -0.222 | 0.218 | -0.261 | 0.421 | 0.039 | 0.139 |
| Physics | -0.274 | 0.239 | -0.309 | 0.262 | 0.035 | 0.094 |
| IV. Exact Science Track |  |  |  |  |  |  |
| Business Administration | -0.167 | 0.459 | -0.130 | 0.516 | -0.037 | 0.100 |
| Computer Science | -0.153 | 0.418 | -0.209 | 0.526 | 0.056 | 0.106 |
| Mathematics | -0.061 | 0.383 | -0.191 | 0.358 | 0.131 | 0.081 |
| Physics | -0.129 | 0.311 | -0.191 | 0.308 | 0.062 | 0.070 |
| V. Optional |  |  |  |  |  |  |
| Economics | -0.122 | 0.329 | -0.041 | 0.457 | -0.081 | 0.049 |

This table presents the means, standard deviations and differences of teacher biases measured in all other classes by teacher gender for the sample of 21 schools. These teacher biases by teacher gender are presented for each subject in $11^{\text {th }}$ and $12^{\text {th }}$ grades. A teacher sample is used here. A negative bias means that the teacher is pro-girl. The baseline sample is $11^{\text {th }}$ grade students in 2003-2005 and $12^{\text {th }}$ grade students in 2003-2011.

Table A8: Heterogeneity in the Effect of $11^{\text {th }}$ Grade Gender Bias on Blind $12^{\text {th }}$ Grade Score by the Gender of the Teacher

| Dependent Variable: Blind score in $12^{\text {th }}$ grade national exams |  |  |  |  |  |  |  |  |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
|  | Boys |  |  |  | Girls |  |  |  |
|  | (1) | (2) | (3) | (4) | (5) | (6) | (7) | (8) |
| All Subjects |  |  |  |  |  |  |  |  |
| Bias | $\begin{gathered} 0.092 \\ (0.046) \end{gathered}$ | $\begin{gathered} 0.056 \\ (0.045) \end{gathered}$ | $\begin{gathered} 0.054 \\ (0.045) \end{gathered}$ | $\begin{gathered} 0.059 \\ (0.063) \end{gathered}$ | $\begin{gathered} 0.005 \\ (0.043) \end{gathered}$ | $\begin{aligned} & -0.027 \\ & (0.041) \end{aligned}$ | $\begin{aligned} & -0.116 \\ & (0.041) \end{aligned}$ | $\begin{gathered} 0.052 \\ (0.060) \end{gathered}$ |
| Bias $\times$ Female Teacher | $\begin{aligned} & -0.044 \\ & (0.065) \end{aligned}$ | $\begin{gathered} 0.016 \\ (0.064) \end{gathered}$ | $\begin{gathered} 0.026 \\ (0.061) \end{gathered}$ | $\begin{gathered} 0.069 \\ (0.091) \end{gathered}$ | $\begin{aligned} & -0.124 \\ & (0.062) \end{aligned}$ | $\begin{aligned} & -0.071 \\ & (0.061) \end{aligned}$ | $\begin{gathered} 0.042 \\ (0.059) \end{gathered}$ | $\begin{aligned} & -0.241 \\ & (0.088) \end{aligned}$ |
| Female Teacher | $\begin{gathered} -0.018 \\ (0.013) \end{gathered}$ | $\begin{aligned} & -0.006 \\ & (0.013) \end{aligned}$ | $\begin{aligned} & -0.003 \\ & (0.012) \end{aligned}$ | $\begin{gathered} -0.009 \\ (0.027) \end{gathered}$ | $\begin{aligned} & -0.019 \\ & (0.012) \end{aligned}$ | $\begin{gathered} 0.002 \\ (0.012) \end{gathered}$ | $\begin{aligned} & -0.015 \\ & (0.011) \end{aligned}$ | $\begin{gathered} 0.028 \\ (0.025) \end{gathered}$ |
| Sample Size | 18,503 | 18,503 | 18,503 | 18,474 | 21,119 | 21,119 | 21,119 | 21,111 |
| Core Subjects |  |  |  |  |  |  |  |  |
| Bias | $\begin{gathered} 0.085 \\ (0.051) \end{gathered}$ | $\begin{gathered} 0.032 \\ (0.053) \end{gathered}$ | $\begin{gathered} 0.030 \\ (0.053) \end{gathered}$ | $\begin{gathered} 0.060 \\ (0.063) \end{gathered}$ | $\begin{gathered} 0.063 \\ (0.049) \end{gathered}$ | $\begin{aligned} & -0.006 \\ & (0.050) \end{aligned}$ | $\begin{aligned} & -0.110 \\ & (0.050) \end{aligned}$ | $\begin{gathered} 0.065 \\ (0.060) \end{gathered}$ |
| Bias $\times$ Female Teacher | $\begin{gathered} 0.013 \\ (0.075) \end{gathered}$ | $\begin{gathered} 0.104 \\ (0.077) \end{gathered}$ | $\begin{gathered} 0.073 \\ (0.076) \end{gathered}$ | $\begin{gathered} 0.082 \\ (0.091) \end{gathered}$ | $\begin{aligned} & -0.191 \\ & (0.072) \end{aligned}$ | $\begin{aligned} & -0.093 \\ & (0.074) \end{aligned}$ | $\begin{gathered} 0.023 \\ (0.077) \end{gathered}$ | $\begin{aligned} & -0.233 \\ & (0.088) \end{aligned}$ |
| Female Teacher | $\begin{gathered} -0.040 \\ (0.017) \end{gathered}$ | $\begin{aligned} & -0.032 \\ & (0.017) \end{aligned}$ | $\begin{aligned} & -0.027 \\ & (0.015) \end{aligned}$ | $\begin{aligned} & -0.010 \\ & (0.028) \end{aligned}$ | $\begin{aligned} & -0.026 \\ & (0.014) \end{aligned}$ | $\begin{aligned} & -0.007 \\ & (0.014) \end{aligned}$ | $\begin{aligned} & -0.022 \\ & (0.013) \end{aligned}$ | $\begin{gathered} 0.027 \\ (0.025) \end{gathered}$ |
| Sample Size | 12,407 | 12,407 | 12,407 | 12,407 | 14,979 | 14,979 | 14,979 | 14,977 |
| Classics Subjects |  |  |  |  |  |  |  |  |
| Bias | $\begin{aligned} & -0.205 \\ & (0.120) \end{aligned}$ | $\begin{aligned} & -0.003 \\ & (0.108) \end{aligned}$ | $\begin{aligned} & -0.066 \\ & (0.095) \end{aligned}$ | $\begin{aligned} & -0.029 \\ & (0.149) \end{aligned}$ | $\begin{aligned} & -0.265 \\ & (0.092) \end{aligned}$ | $\begin{aligned} & -0.015 \\ & (0.084) \end{aligned}$ | $\begin{gathered} -0.142 \\ (0.086) \end{gathered}$ | $\begin{aligned} & -0.089 \\ & (0.143) \end{aligned}$ |
| Bias $\times$ Female Teacher | $\begin{gathered} 0.280 \\ (0.138) \end{gathered}$ | $\begin{gathered} 0.123 \\ (0.125) \end{gathered}$ | $\begin{gathered} 0.246 \\ (0.111) \end{gathered}$ | $\begin{gathered} 0.177 \\ (0.178) \end{gathered}$ | $\begin{aligned} & 0.111 \\ & (0.109) \end{aligned}$ | $\begin{aligned} & -0.098 \\ & (0.100) \end{aligned}$ | $\begin{gathered} 0.129 \\ (0.099) \end{gathered}$ | $\begin{aligned} & -0.164 \\ & (0.168) \end{aligned}$ |
| Female Teacher | $\begin{gathered} 0.010 \\ (0.027) \end{gathered}$ | $\begin{gathered} 0.057 \\ (0.027) \end{gathered}$ | $\begin{gathered} 0.003 \\ (0.022) \end{gathered}$ | $\begin{gathered} 0.037 \\ (0.045) \end{gathered}$ | $\begin{aligned} & -0.004 \\ & (0.022) \end{aligned}$ | $\begin{gathered} 0.035 \\ (0.021) \end{gathered}$ | $\begin{aligned} & -0.011 \\ & (0.018) \end{aligned}$ | $\begin{aligned} & -0.025 \\ & (0.039) \end{aligned}$ |
| Sample Size | 3,679 | 3,679 | 3,679 | 3,679 | 7,226 | 7,226 | 7,226 | 7,224 |
| Science Subjects |  |  |  |  |  |  |  |  |
| Bias | $\begin{gathered} 0.123 \\ (0.052) \end{gathered}$ | $\begin{gathered} 0.060 \\ (0.053) \end{gathered}$ | $\begin{gathered} 0.105 \\ (0.054) \end{gathered}$ | $\begin{gathered} 0.069 \\ (0.068) \end{gathered}$ | $\begin{gathered} 0.132 \\ (0.048) \end{gathered}$ | $\begin{gathered} 0.047 \\ (0.049) \end{gathered}$ | $\begin{aligned} & -0.094 \\ & (0.050) \end{aligned}$ | $\begin{gathered} 0.117 \\ (0.067) \end{gathered}$ |
| Bias $\times$ Female Teacher | $\begin{gathered} 0.006 \\ (0.085) \end{gathered}$ | $\begin{gathered} 0.100 \\ (0.088) \end{gathered}$ | $\begin{gathered} 0.045 \\ (0.094) \end{gathered}$ | $\begin{gathered} 0.047 \\ (0.118) \end{gathered}$ | $\begin{aligned} & -0.160 \\ & (0.088) \end{aligned}$ | $\begin{aligned} & -0.064 \\ & (0.087) \end{aligned}$ | $\begin{gathered} 0.010 \\ (0.093) \end{gathered}$ | $\begin{aligned} & -0.162 \\ & (0.115) \end{aligned}$ |
| Female Teacher | $\begin{gathered} -0.040 \\ (0.019) \end{gathered}$ | $\begin{aligned} & -0.033 \\ & (0.020) \end{aligned}$ | $\begin{gathered} -0.007 \\ (0.020) \end{gathered}$ | $\begin{aligned} & -0.030 \\ & (0.035) \end{aligned}$ | $\begin{aligned} & -0.021 \\ & (0.017) \end{aligned}$ | $\begin{gathered} 0.000 \\ (0.017) \end{gathered}$ | $\begin{aligned} & -0.023 \\ & (0.018) \end{aligned}$ | $\begin{gathered} 0.067 \\ (0.031) \end{gathered}$ |
| Sample Size | 10,994 | 10,994 | 10,994 | 10,992 | 11,794 | 11,794 | 11,794 | 11,792 |
| Exact Science Subjects |  |  |  |  |  |  |  |  |
| Bias | $\begin{gathered} 0.127 \\ (0.053) \end{gathered}$ | $\begin{gathered} 0.072 \\ (0.052) \end{gathered}$ | $\begin{gathered} 0.084 \\ (0.055) \end{gathered}$ | $\begin{gathered} 0.069 \\ (0.068) \end{gathered}$ | $\begin{gathered} 0.082 \\ (0.051) \end{gathered}$ | $\begin{gathered} 0.023 \\ (0.052) \end{gathered}$ | $\begin{aligned} & -0.103 \\ & (0.053) \end{aligned}$ | $\begin{gathered} 0.115 \\ (0.067) \end{gathered}$ |
| Bias $\times$ Female Teacher | $\begin{aligned} & -0.126 \\ & (0.083) \end{aligned}$ | $\begin{aligned} & -0.051 \\ & (0.083) \end{aligned}$ | $\begin{gathered} 0.005 \\ (0.090) \end{gathered}$ | $\begin{gathered} 0.049 \\ (0.118) \end{gathered}$ | $\begin{aligned} & -0.124 \\ & (0.088) \end{aligned}$ | $\begin{aligned} & -0.055 \\ & (0.086) \end{aligned}$ | $\begin{aligned} & 0.018 \\ & (0.090) \end{aligned}$ | $\begin{aligned} & -0.161 \\ & (0.115) \end{aligned}$ |
| Female Teacher | $\begin{aligned} & -0.052 \\ & (0.017) \end{aligned}$ | $\begin{gathered} -0.037 \\ (0.018) \end{gathered}$ | $\begin{aligned} & -0.023 \\ & (0.017) \end{aligned}$ | $\begin{aligned} & -0.031 \\ & (0.035) \end{aligned}$ | $\begin{aligned} & -0.036 \\ & (0.017) \end{aligned}$ | $\begin{gathered} 0.014 \\ (0.017) \end{gathered}$ | $\begin{aligned} & -0.023 \\ & (0.018) \end{aligned}$ | $\begin{gathered} 0.068 \\ (0.031) \end{gathered}$ |
| Sample Size | 13,113 | 13,113 | 13,113 | 13,113 | 12,011 | 12,011 | 12,011 | 12,007 |
| $\begin{aligned} & \text { Subject FE } \\ & \text { Year FE } \end{aligned}$ School FE | $\checkmark$ | $\checkmark$ $\checkmark$ $\checkmark$ | $\checkmark$ |  |  | $\checkmark$ $\checkmark$ $\checkmark$ | $\checkmark \checkmark$ |  |
| Class FE <br> School-by-Subject-by-Year FE |  |  | $\checkmark$ | $\checkmark$ |  |  | $\checkmark$ | $\checkmark$ |

Notes: Each row presents estimates from a separate regression using an empirical Bayes estimation strategy. The standard errors are corrected using a two-step bootstrapping method. The process of two-step bootstrap sampling and estimation is repeated 1,000 times. All specifications include the students' $11^{\text {th }}$ grade first-semester 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 fourth panel "Exact Science Subjects" includes relevant exams from the core (algebra, geometry and physics) and all the exact science track subjects. All specifications include the students' $11^{\text {th }}$ grade first-semester test scores as controls.

Table A9: Heterogeneity in the Effect of $11^{\text {th }}$ Grade Gender Bias on Blind $12^{\text {th }}$ Grade Score by the Gender of the Principal

| Dependent Variable: Blind score in $12^{\text {th }}$ grade national exams |  |  |
| :---: | :---: | :---: |
|  | Boys | Girls |
|  | (1) | (2) |
| All Subjects |  |  |
| Bias | $\begin{gathered} 0.051 \\ (0.058) \end{gathered}$ | $\begin{aligned} & -0.075 \\ & (0.047) \end{aligned}$ |
| Bias $\times$ Female Principal | $\begin{gathered} 0.048 \\ (0.076) \end{gathered}$ | $\begin{gathered} 0.074 \\ (0.068) \end{gathered}$ |
| Female Principal | $\underset{(0.026)}{0.037}$ | $\begin{gathered} 0.031 \\ (0.022) \end{gathered}$ |
| Sample Size | 12,881 | 15,376 |
| Core Subjects |  |  |
| Bias | $\begin{gathered} 0.047 \\ (0.069) \end{gathered}$ | $\begin{gathered} -0.036 \\ (0.057) \end{gathered}$ |
| Bias $\times$ Female Principal | $\begin{gathered} 0.121 \\ (0.091) \end{gathered}$ | $\begin{gathered} 0.079 \\ (0.079) \end{gathered}$ |
| Female Principal | $\begin{gathered} 0.013 \\ (0.026) \end{gathered}$ | $\begin{aligned} & -0.001 \\ & (0.021) \end{aligned}$ |
| Sample Size | 8,805 | 11,074 |
| Classics Subjects |  |  |
| Bias | $\begin{aligned} & -0.007 \\ & (0.089) \end{aligned}$ | $\begin{gathered} -0.173 \\ (0.063) \end{gathered}$ |
| Bias $\times$ Female Principal | $\begin{gathered} 0.123 \\ (0.139) \end{gathered}$ | $\begin{gathered} 0.058 \\ (0.108) \end{gathered}$ |
| Female Principal | $\begin{gathered} 0.104 \\ (0.043) \end{gathered}$ | $\begin{gathered} 0.107 \\ (0.034) \end{gathered}$ |
| Sample Size | 2,689 | 5,297 |
| Science Subjects |  |  |
| Bias | $\begin{gathered} 0.112 \\ (0.079) \end{gathered}$ | $\underset{(0.73)}{0.147}$ |
| Bias $\times$ Female Principal | $\begin{gathered} 0.039 \\ (0.098) \end{gathered}$ | $\begin{aligned} & -0.024 \\ & (0.098) \end{aligned}$ |
| Female Principal | $\begin{gathered} -0.012 \\ (0.027) \end{gathered}$ | $\begin{aligned} & -0.067 \\ & (0.023) \end{aligned}$ |
| Sample Size | 7,633 | 8,561 |
| Exact Science Subjects |  |  |
| Bias | $\begin{gathered} 0.143 \\ (0.077) \end{gathered}$ | $\begin{gathered} 0.086 \\ (0.068) \end{gathered}$ |
| Bias $\times$ Female Principal | $\begin{aligned} & -0.081 \\ & (0.098) \end{aligned}$ | $\begin{aligned} & -0.007 \\ & (0.097) \end{aligned}$ |
| Female Principal | $\underset{(0.027)}{0.004}$ | $\begin{gathered} -0.039 \\ (0.024) \end{gathered}$ |
| Sample Size | 9,099 | 8,823 |
| $\begin{aligned} & \text { Subject FE } \\ & \text { Year FE } \end{aligned}$ | $\checkmark$ | $\checkmark$ |
| Notes: Each panel presents estimates from a separate regression using an empirical Bayes estimation strategy. Standard errors are clustered using a two-step bootstrapping method. All specifications include the students' first-semester test score in grade 11 as a control. All scores are standardized z-scores. The first panel "All Subjects" includes all subjects from the core and the tracks. The second panel "Core Subjects" includes all core subjects. The third panel "Classics Subjects" includes relevant exams from the core (history and modern Greek) and all the classics track subjects. The fourth panel "Science Subjects" includes relevant exams from the core (algebra, geometry and physics) and all the science track subjects. The firth panel "Exact Science Subjects" includes relevant exams from the core (algebra, geometry and physics) and all the exact science track subjects. |  |  |

Table A10: Effect of Teacher Gender Bias on the Quality of the Program Students Enrolled In

|  | Boys |  | Girls |  |
| :---: | :---: | :---: | :---: | :---: |
|  | $\begin{gathered} (1) \\ \text { Rank } \end{gathered}$ | $\begin{gathered} (2) \\ \text { Stand Rank } \end{gathered}$ | $\begin{gathered} (3) \\ \text { Rank } \end{gathered}$ | $\begin{gathered} (4) \\ \text { Stand Rank } \end{gathered}$ |
| Bias in All Subjects | $\begin{gathered} 1.693 \\ (0.943) \end{gathered}$ | $\begin{gathered} 0.059 \\ (0.033) \end{gathered}$ | $\begin{aligned} & -1.841 \\ & (0.808) \end{aligned}$ | $\begin{aligned} & -0.064 \\ & (0.028) \end{aligned}$ |
| Sample Size | 8,452 | 8,452 | 10,125 | 10,125 |
| Bias in Core Subjects | $\begin{gathered} 1.258 \\ (1.034) \end{gathered}$ | $\begin{gathered} 0.043 \\ (0.036) \end{gathered}$ | $\begin{aligned} & -3.305 \\ & (0.939) \end{aligned}$ | $\begin{gathered} -0.114 \\ (0.032) \end{gathered}$ |
| Sample Size | 4,986 | 4,986 | 6,053 | 6,053 |
| Bias in Classics Subjects | $\begin{gathered} 2.583 \\ (1.740) \end{gathered}$ | $\begin{gathered} 0.090 \\ (0.061) \end{gathered}$ | $\begin{gathered} 0.201 \\ (1.527) \end{gathered}$ | $\begin{gathered} 0.007 \\ (0.053) \end{gathered}$ |
| Sample Size | 2,668 | 2,668 | 4,134 | 4,134 |
| Bias in Science Subjects | $\begin{gathered} 0.478 \\ (1.320) \end{gathered}$ | $\begin{gathered} 0.017 \\ (0.046) \end{gathered}$ | $\begin{gathered} -2.009 \\ (1.203) \end{gathered}$ | $\begin{gathered} -0.069 \\ (0.041) \end{gathered}$ |
| Sample Size | 3,637 | 3,637 | 4,342 | 4,342 |
| Bias in Exact Science Subjects | $\begin{gathered} 1.246 \\ (1.536) \end{gathered}$ | $\begin{gathered} 0.043 \\ (0.053) \end{gathered}$ | $\begin{aligned} & -1.221 \\ & (1.211) \end{aligned}$ | $\begin{gathered} -0.042 \\ (0.042) \end{gathered}$ |
| Sample Size | 4,264 | 4,264 | 4,199 | 4,199 |
| Year FE |  | $\checkmark$ | $\checkmark$ | $\checkmark$ |
| Subject FE | $\checkmark$ | $\checkmark$ | $\checkmark$ | $\checkmark$ |
| Grade FE | $\checkmark$ | $\checkmark$ | $\checkmark$ | $\checkmark$ |
| Class FE | $\checkmark$ | $\checkmark$ | $\checkmark$ | $\checkmark$ |

Notes: Year 2003 is excluded from the analysis, as it is used to calculate the quality measures for the post-secondary program students enroll in. "Rank" represents the ranking of each post-secondary program based on the 2003 threshold score for each university department. This is calculated as the admission score of the last admitted/marginal student. "Stand Rank" is the standardized measure of "Rank" which is normalised to have a zero mean and a standard deviation of 1 . We assign this measure of program quality to the relevant post-secondary programs and drop the year 2003 from the regressions. We then look at the effects of teacher biases on student quality of enrolled postsecondary program. All specifications include the students' first semester $11^{\text {th }}$ grade performance and the teacher's gender as controls. All estimates are adjusted for the empirical Bayes technique. Standard errors are adjusted using a two-step bootstrapping technique.

Table A11: Summary Statistics for Sample Used to Estimate Value-Added Models

|  | Mean | Std.Dev | Min | Max |
| :---: | :---: | :---: | :---: | :---: |
|  | (1) | (2) | (3) | (4) |
| Student Characteristics |  |  |  |  |
| Born in the First Quarter of Birth Year (Yes=1) | 0.238 | 0.426 | 0 | 1 |
| Gender (Female=1) | 0.576 | 0.494 | 0 | 1 |
| Test Scores (s.d.) | -0.045 | 0.995 | -2.937 | 3.608 |
| Proportion of Students in Classics | 0.391 | 0.488 | 0 | 1 |
| Proportion of Students in Science | 0.174 | 0.379 | 0 | 1 |
| Proportion of Students in Exact Science | 0.435 | 0.496 | 0 | 1 |
| Lagged Test Scores (s.d) | 0.025 | 0.963 | -4.365 | 3.807 |
| No.of Subject-School Combinations per Student | 10.044 | 5.233 | 1 | 22 |
| Missing Lagged Test Scores | 0.041 | 0.198 | 0 | 1 |
| Class/School/Neighborhood Characteristics |  |  |  |  |
| Class Size | 18.818 | 3.761 | 4 | 25 |
| School Grade Enrollment Size | 62.170 | 28.504 | 4 | 104 |
| Neighorhood Income (in Euro) | 19,718.846 | 3,794.528 | 12,265.880 | 26,586.330 |
| Urban Locality (yes=1) | 0.781 | 0.414 | 0 | 1 |
| Teacher Characteristics |  |  |  |  |
| Teacher Gender (Female=1) | 0.337 | 0.473 | 0 | 1 |
| No.of other Classes a Teacher Taught (Experience) | 6.529 | 5.582 | 1 | 22 |

Notes: All statistics reported are for the sample used in estimating the baseline TVA model. This sample includes only students who have non-missing baseline controls. Student data are from the administrative records of 21 schools in Greece. The sample period of 2003-2005 is used to obtain the TVA estimates. All test scores are standardized. Summary statistics (number of observations, mean, s.d., min., and max.) for the baseline variables are reported in the table. Outcome test scores are measured in $11^{t h}$ and $12^{t h}$ grade and the prior test scores are measured in $10^{t h}$ and $11^{t h}$ grade. The variables are weighted by the number of students in the school-year-class-subject-year cell. Only the "Number of other Classes a Teacher Taught" is weighted by the number of teachers in the school-year-class-subject-year cell. The age is measured in years for students the day they start the $11^{\text {th }}$ or $12^{\text {th }}$ grade class. Born in the First Quarter of Birth Year is a dummy that takes the value of one if a student is born in the first quarter of the calendar year, and zero otherwise. Students who are born in the first quarter of the calendar year are eligible to enroll a year earlier in the $1^{\text {st }}$ grade. There are three tracks available: classics, science or exact science. The school grade enrollment size denotes the number of $11^{t h}$ or $12^{t h}$ graders in a given school and year. The number of subject-school combinations per student is the number of subjects that students take. Each student takes on average 10 subjects. When prior test scores are missing, we set the prior score equal to 0 and include an indicator for missing data. On average, 6 percent of lagged scores are missing. The total number of observations used here is 50,970 . In the last panel that reports teachers' characteristics, a class corresponds to a subject-class-year-grade cell.

Table A12: Comparisons of Mean Teacher Value Added for Pro-Boys, Neutral, and Pro-Girls Teachers
$\left.\begin{array}{lccc}\hline \hline & & \text { (1) } & (2) \\ \text { Neutral } & \text { Pro-Boy } & \text { (3) } \\ \text { Difference }\end{array}\right)$

Notes: We pool data on test scores for $11^{\text {th }}$ and $12^{\text {th }}$ grades for the period 2003-2005. The TVA measures are derived following the procedure described in the text. Pro-boy teachers exhibit bias larger than or equal to 0.10. Pro-girl teachers exhibit bias smaller than or equal to -0.10 . We define as neutral teachers who exhibit bias that is larger than -0.10 and smaller than or equal to 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 TVA model. Our baseline TVA 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 TVA 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. In Panel A, we compare the mean TVA of teachers who are neutral (column 1) to the mean TVA of teachers who are pro-boy (column 2). The related standard deviations are reported below the means. Column 3 reports the difference of the means and the respective standard error. In Panel B, we compare the mean TVA of teachers who are neutral (column 1) to the mean TVA of teachers who are pro-girl (column 2). The related standard deviations are reported below the means. Column 3 reports the difference of the means and the standard error.

Table A13: 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-Girls Teachers | $\begin{gathered} 0.040 \\ (0.028) \end{gathered}$ | $\begin{gathered} 0.039 \\ (0.029) \end{gathered}$ | $\begin{gathered} 0.041 \\ (0.029) \end{gathered}$ | $\begin{gathered} 0.042 \\ (0.030) \end{gathered}$ |
| Spline for Pro-Boys Teachers | $\begin{aligned} & -0.021 \\ & (0.022) \end{aligned}$ | $\begin{aligned} & -0.020 \\ & (0.022) \end{aligned}$ | $\begin{gathered} -0.019 \\ (0.022) \end{gathered}$ | $\begin{aligned} & -0.021 \\ & (0.022) \end{aligned}$ |
| Female Teacher |  | $\begin{gathered} 0.011 \\ (0.011) \end{gathered}$ | $\begin{gathered} 0.011 \\ (0.011) \end{gathered}$ | $\begin{gathered} 0.011 \\ (0.011) \end{gathered}$ |
| Class Size |  |  | $\begin{gathered} -0.002 \\ (0.002) \end{gathered}$ | $\begin{gathered} -0.002 \\ (0.002) \end{gathered}$ |
| Experience |  |  |  | $\begin{gathered} -0.001 \\ (0.002) \\ \hline \end{gathered}$ |
| Obs. | 418 | 418 | 418 | 418 |
| Year FE | $\checkmark$ | $\checkmark$ | $\checkmark$ | $\checkmark$ |
| School FE | $\checkmark$ | $\checkmark$ | $\checkmark$ | $\checkmark$ |
| Grade FE | $\checkmark$ | $\checkmark$ | $\checkmark$ | $\checkmark$ |

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 ProBoy 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 20032005 sample and is described in detail in the text. "Teacher experience" measures the different combination of classes and subjects a teacher has taught in $11^{\text {th }}$ and $12^{\text {th }}$ grades in the sample period 2003-2011. Standard errors are clustered by school and year and are reported in parentheses. The empirical Bayes estimates of teacher gender biases are used.

Table A14: Comparisons of Mean Teacher Value Added for Pro-Boys, Neutral, and Pro-Girls Teachers (Measured by TVA on Enrolled Program Quality)

|  | (1) <br> Neutral /sd | $\begin{gathered} (2) \\ \text { Pro-Boy } \\ / \text { sd } \end{gathered}$ | (3) Difference /se |
| :---: | :---: | :---: | :---: |
|  |  | Panel A |  |
| Teacher Value Added | $\begin{aligned} & 1.685 \\ & 9.525 \end{aligned}$ | $\begin{gathered} -0.979 \\ 9.503 \end{gathered}$ | 2.665 1.542 |
| N | 67 |  | 155 |
|  | Neutral /sd | $\begin{gathered} \text { Pro-Girl } \\ \text { /sd } \end{gathered}$ | $\begin{gathered} \hline \text { Difference } \\ / \mathrm{se} \end{gathered}$ |
|  | Panel B |  |  |
| Teacher Value Added | $\begin{aligned} & 1.685 \\ & 9.525 \end{aligned}$ | $\begin{gathered} -1.480 \\ 9.771 \end{gathered}$ | $\begin{aligned} & 3.165 \\ & 1.467 \end{aligned}$ |
| N | 67 | 125 | 192 |

Notes: Year 2003 is excluded from the analysis, since it is used to define the quality of the enrolled university degree program. We pool data on test scores for $11^{\text {th }}$ and $12^{\text {th }}$ grades for the period 2004-2005. The TVA measures are derived following the procedure described in the text. Pro-boy teachers exhibit bias larger than or equal to 0.10. Pro-girl teachers exhibit bias smaller than or equal to -0.10 . We define as neutral teachers who exhibit bias that is larger than -0.10 and smaller than or equal to 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 TVA model. Our baseline TVA 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 TVA 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. In Panel A, we compare the mean TVA of teachers who are neutral (column 1) to the mean TVA of teachers who are pro-boy (column 2). The related standard deviations are reported below the means. Column 3 reports the difference of the means and the respective standard error. In Panel B, we compare the mean TVA of teachers who are neutral (column 1) to the mean TVA of teachers who are pro-girl (column 2). The related standard deviations are reported below the means. Column 3 reports the difference of the means and the standard error.

Table A15: Correlations Between Teacher Gender Bias And Teacher Quality (Measured by TVA
on Enrolled Program Quality)
Dependent Variable: Teacher Quality (Measured by TVA)

|  | $(1)$ | $(2)$ | $(3)$ | $(4)$ |
| :--- | :---: | :---: | :---: | :---: |
| Pro-Girl Teacher Indicator | -0.586 | -0.601 | -0.607 | -0.553 |
|  | $(0.816)$ | $(0.807)$ | $(0.809)$ | $(0.786)$ |
| Pro-Boy Teacher Indicator | 0.171 | 0.170 | 0.212 | 0.260 |
|  | $(0.677)$ | $(0.680)$ | $(0.688)$ | $(0.688)$ |
| Female Teacher |  | -0.151 | -0.198 | -0.160 |
|  |  | $(0.707)$ | $(0.693)$ | $(0.687)$ |
| Class Size |  |  | -0.174 | -0.184 |
|  |  |  | $(0.156)$ | $(0.158)$ |
| Experience |  |  |  | -0.068 |
| Obs. | 279 | 279 | 279 | 279 |
| Year FE | $\checkmark$ | $\checkmark$ | $\checkmark$ | $\checkmark$ |
| School FE | $\checkmark$ | $\checkmark$ | $\checkmark$ | $\checkmark$ |
| Grade FE | $\checkmark$ | $\checkmark$ | $\checkmark$ | $\checkmark$ |

Notes: Year 2003 is excluded from the analysis, since it is used to define the quality of the enrolled university degree program. The "Pro-Girl Teacher Indicator" takes the value of one if the teacher exhibits a bias that is smaller than or equal to -0.10. The "Pro-Boy Teacher Indicator" 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 -0.10 and smaller than or equal to 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 $11^{t h}$ and $12^{\text {th }}$ grades in the sample period 20032011. The empirical Bayes estimates of teacher gender biases are used. Standard errors are clustered by school and year and are reported in parentheses.

Figure A1: Distribution of Teacher Gender Bias in $11^{\text {th }}$ and $12^{\text {th }}$ Grade, Sample of 21 Schools

Panel A: $11^{\text {th }}$ grade


Panel B: $12^{\text {th }}$ grade


Notes: Panels A and B show the teacher-class level distribution of bias based on all classes (including the own) in grades $11^{\text {th }}$ and $12^{\text {th }}$, respectively. We use data for the sample period 2003-2011. The teacher bias in all classes that a teacher taught is measured as the average bias that a teacher exhibited in all classes she/he ever taught in the sample period 2003-2011 in each grade.

Figure A2: Histograms of the Bias Measured in All Classes in Core and Track Subjects in $11^{\text {th }}$ Grade


Notes: This figure presents a teacher-class level distribution of bias based on all classes (including the own) for core and track subjects. In $11^{\text {th }}$ grade the core subjects taught are: modern Greek, history, physics, algebra and geometry. There are the following tracks in $11^{\text {th }}$ grade: Classics, Science and Exact Science. In the classics track the $11^{\text {th }}$ grade subjects are: ancient Greek, philosophy and Latin; in the science track: mathematics, physics, chemistry, and in the exact science track: mathematics, physics and technology and computers. The mean (s.d.) of the teacher-level measure of the gender bias based on all classes is $-0.123(0.418),-0.166(0.514),-0.029(0.502)$ and $-0.065(0.516)$ in core, classics track, science track, and exact classics track subjects, respectively.

Figure A3: Histograms of the Bias Measured in All Classes in Core and Track Subjects in $12^{\text {th }}$ Grade


Notes: This figure presents a teacher-class level distribution of bias based on all classes (including the own) for core and track subjects. In $12^{\text {th }}$ grade the core subjects taught are: modern Greek, history, physics, biology and mathematics. There are the following tracks in $12^{\text {th }}$ grade: Classics, Science and Exact Science. In the classics track the $12^{\text {th }}$ grade subjects are: ancient Greek, Latin, literature and history; in the science track: biology, mathematics, physics and chemistry, and in the exact science track: mathematics, physics, business administration and computer science. The mean (s.d.) of the teacher-level measure of the gender bias based on all classes is -0.139 ( 0.491 ), -0.197 (0.478), $-0.197(0.472)$ and $-0.171(0.430)$ in core, classics track, science track, and exact classics track subjects, respectively.

Figure A4: Evolution of Teacher Bias, All, Male and Female Teachers who Teach 8 or 9 Years in the Sample


Notes: We calculate the annual bias across subjects and classes for all teachers in the sample. We then plot the evolution of teacher bias for male and female teachers who teach 8 or 9 years in the sample. We consider years from 2003 up to and including 2011.

Figure A5: Proportion of University Students Enrolled By University Field of Studies, Sample of 21 schools


Notes: This figure presents the proportion of enrolled students in each field of university study and year. We consider four broad categories: humanities, science, social science and exact science. The remaining students enroll in vocational studies. Humanities include the departments of liberal Arts, literature, psychology, 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 science and European studies. Exact Science includes the departments of mathematics, engineering, physics and computer science. Science includes the departments of biology, chemistry, medicine, pharmacy, veteひnary studies and dentistry.

Figure A6: Proportion of University Students Enrolled By Field of Studies and By Gender, Sample of 21 schools


Notes: The top figure presents the proportion of enrolled boys (out of enrolled students) in each field of university study and year. The bottom figure presents the proportion of enrolled girls (out of enrolled students) in each field of university study and year. We consider four broad categories: humanities, science, social science and exact science.

Figure A7: Proportion of University Students Enrolled in STEM by Subject, Sample of 21 schools


Notes: This figure presents the proportion of enrolled students in each stem field and year. We use the following five broad fields: mathematics, engineering, physics, computer science and architecture.

Figure A8: Proportion of University Students in STEM fields by Gender, Sample of 21 schools



Notes: The top figure presents the proportion of enrolled boys in each STEM field and year. The bottom figure presents the proportion of enrolled girls in each STEM field and year. Five broad fields are used: mathematics, engineering, physics, computer science and architecture.

Figure A9: Evolution of Teacher Gender Bias by School Track, Sample of 21 Schools


Notes: The $11^{\text {th }}$ grade bias in Classics is the average bias in ancient Greek, philosophy and Latin. The $11^{\text {th }}$ grade bias in Science ${ }^{2} T_{s}$ the average bias in mathematics, physics and chemistry. The $11^{\text {th }}$ grade bias in Exact Science is the average bias in mathematics, physics and computer science. The $12^{\text {th }}$ grade bias in Classics is the average bias in ancient Greek, Latin, literature and history. The $12^{\text {th }}$ grade bias in Science is the average bias in mathematics, physics, biology and chemistry. The $12^{\text {th }}$ grade bias in Exact Science is the average bias in mathematics, physics, business administration and computer science. All teacher biases that are used here are calculated based on all other classes that a teacher taught in the sample.

Figure A10: Evolution of Teacher Gender Bias by Field of Study, Sample of 21 schools


Notes: These figures show the evolution (over ${ }^{2}{ }^{2}$ ime) of the teacher bias in terms of the fields of study at the university level. These fields of study include departments in the following fields: humanities, social science, science and exact science. The $11^{\text {th }}$ grade bias (measured in all other classes) for humanities and social science departments is the average bias in modern Greek and history in $11^{\text {th }}$ grade. The $11^{\text {th }}$ grade bias (measured in all other classes) for Science and Exact Science departments is the average bias in algebra, geometry and physics in $11^{\text {th }}$ grade. The $12^{\text {th }}$ grade bias (measured in all other classes) for exact science, science, humanities and social science departments is the bias in mathematics and physics

Figure A11: Histogram of Teacher Value-Added Measure on the Quality of the Program Students Enrolled in


Notes: This figure presents the distribution of the TVA measure, which is weighted by the number of students in the school-year-grade-subject-class year cell. Teacher value-added is measured on the enrolled university degree quality. The enrolled university degree quality ranges from 0 to 100 , which the value of 100 being assigned to the program of the highest quality. Year 2003 is excluded, as it is used to calculate the quality measures for the post-secondary program students enroll in. The program quality is measured as the ranking of each post-secondary program based on the 2003 mean performance of enrolled students in each post-secondary program. To derive these value-added measures we pool the $11^{t h}$ and $12^{t h}$ grade data for the years $2004-2005$. We use $10^{t h}$ and $11^{t h}$ 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 TVA 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 cell.

Figure B1: Example of a School Exam Script in a Random School in Academic Year 2005-2006


## E®ETAZOMENO MAOHMA ГENIKHะ ПAIAEIA乏： MAӨHMATIKA KAI ЕTOIXEIA ЕTATI亡TIKH亡

## OEMA A：



$$
\mathrm{P}(\mathrm{AUB})=\mathrm{P}(\mathrm{~A})+\mathrm{P}(\mathrm{~B})
$$

（Mováósç 10）

（Mováósç 6）

a）$(x)^{\prime}=$ $\qquad$ ß）$(\eta \mu x)^{\prime}=$ $\qquad$ y）$\left(x^{3}\right)^{\prime}=$ $\qquad$ （Mováósç 9）

## OEMA B：

 опияіо $\mathrm{A}(1,-4)$ ．

（Mováóec 8）
B2． $\mathrm{Na} \beta$ prite to $\lim _{x \rightarrow 3} \frac{x^{2}-9}{f(x)}$ ． （Mováóeç 8）

（Mováóȩ 9）

## OEMA Г：








（Movábec 8）
（Mováoeç 8）
（Mováoec 9）

Figure B2: Example of a National Exam Script in Academic Year 2005-2006


## APXH $2 H \Sigma \Sigma E \Lambda I \Delta A \Sigma$

## NEO \& ПAAAIO EYETHMA - Г' HMEPHEIQN

ӨEMA B

B1. $N \alpha \beta \rho \varepsilon і т \varepsilon$ та акро́тата тпऽ бuvápтпопऽf.
Mováбะऽ 9



Mováбะ؟ 8
B3. Na ито入оүібєєє то о́рıо $\lim _{x \rightarrow-1} \frac{f^{\prime}(x)-12}{x+1}$.
Mováठェร 8
ӨEMA Г


 бгєббробıа́үра $\mu \alpha$.

Mováסes 4



A: «то прш́то паıб́i вivaı корітбı»


Mováбะऽ 6




$$
\Delta=\mathrm{A} \cap \mathrm{~B}, \quad \mathrm{E}=\mathrm{A} \cup \mathrm{~B}, \mathrm{Z}=\Gamma-\mathrm{E} .
$$

( $\mu$ ováठ́єऽ 9 )




Mováరॄऽ 15

TE $\Lambda \mathrm{O} \Sigma 2 \mathrm{H} \Sigma \mathrm{A} \Pi \mathrm{O} 4 \Sigma \mathrm{E} \Lambda \mathrm{I} \Delta \mathrm{E} \Sigma$

## APXH $3 H \Sigma \Sigma E \Lambda I \Delta A \Sigma$

## NEO \& ПAAAIO EYETHMA - $\Gamma^{\prime}$ HMEPHEIQN

## $\Theta E M A \Delta$


 тірака:

| Xpóvos ( $\sigma \varepsilon \lambda \varepsilon \pi \tau_{1} \dot{\text { ) }}$ |  | $\begin{gathered} \overline{\text { Euxvótnta }} \\ v_{i} \end{gathered}$ |
| :---: | :---: | :---: |
| [8, ) |  | 20 |
| [, ) | 14 | 15 |
| [, ) |  | 10 |
| [ , ) |  | $\mathrm{v}_{4}$ |
| £YNO^O |  | $\mathrm{v}=\ldots \ldots \ldots$. |


Mováס̌ऽ 4

 ката́ $\lambda \lambda \eta \lambda \alpha \sigma u \mu \pi \lambda \eta \rho \omega \mu \varepsilon ́ v o ~(\mu \circ v \alpha ́ \delta \varepsilon \varsigma ~ 2) . ~$

Mováбะऽ 6

 про́үрациа.

Mováठes 5



Mováסes 6





Mováбะऽ 4

TE $\triangle \mathrm{O} \Sigma 3 \mathrm{H} \Sigma$ AПO $4 \Sigma \mathrm{E} \Lambda \mathrm{I} \Delta \mathrm{E} \Sigma$

## APXH 4H $\Sigma \Sigma E \Lambda I \Delta A \Sigma$

NEO \＆ПAAAIO EYITHMA－［＇HMEPHIIQN

## 
















## £A乏 EYXOMA乏TE KAへH EПITYXIA

TEAOE MHNYMATOE

TE $\Lambda \mathrm{O} \Sigma 4 \mathrm{H} \Sigma$ AПO $4 \Sigma \mathrm{E} \Lambda \mathrm{I} \Delta \mathrm{E} \Sigma$

Figure B3: Map of Schools in the Sample


This figure shows the counties with schools that we use in the main analysis.

