On The Origins of Gender Human Capital Gaps: Short and Long Term Consequences of Teachers' Stereotypical Biases

Edith Sand

Bank of Israel

March 26-2015, 11:15-12:30

Bldg. 72, room 465

Abstract

In this paper, we estimate the effect of primary school teachers' stereotypes on boys' and girls' academic achievements during middle and high school and on the choice of higher level courses in math and sciences during high school. For identification, we rely on the conditional random assignments of teachers and students to classes in primary schools. Our results suggest that teachers' stereotypes favoring boys have asymmetric effect by gender, positive effect on boys' achievements and negative effect on girls. Such stereotypes also impact students' enrollment in advanced level math courses in high school, boys positively and girls negatively. These results suggest that stereotypical attitudes of teachers at early stage of schooling have long run implications for the gender gap in math achievements. This impact is heterogeneous, being larger for children from low socialeconomic background and in families with high parental gap in years of schooling.

On The Origins of Gender Human Capital Gaps: Short and Long Term Consequences of Teachers' Stereotypical Biases*

January 2015

Victor Lavy
University of Warwick, Hebrew University and NBER

Edith Sand
Bank of Israel

Abstract

In this paper, we estimate the effect of primary school teachers' gender biases on boys' and girls' academic achievements during middle and high school and on the choice of advanced level courses in math and sciences during high school. For identification, we rely on the random assignments of teachers and students to classes in primary schools. Our results suggest that teachers' biases favoring boys have an asymmetric effect by gender—positive effect on boys' achievements and negative effect on girls'. Such gender biases also impact students' enrollment in advanced level math courses in high school—boys positively and girls negatively. These results suggest that teachers' biased behavior at early stage of schooling have long run implications for occupational choices and earnings at adulthood, because enrollment in advanced courses in math and science in high school is a prerequisite for post-secondary schooling in engineering, computer science and so on. This impact is heterogeneous, being larger for children from families where the father is more educated than the mother and larger on girls from low socioeconomic background.

Victor Lavy, msvictor@huji.ac.il,
v.lavy@warwick.ac.uk
Edith Sand, edith.sand@boi.org.il

* We thank the Education Department of Tel-Aviv-Yafo Municipality and Yossef Shub, the CEO of Optimal Scheduling Systems, for making the data available for this study, Israel's Ministry of Education for allowing restricted access to secondary schooling data in the Ministry online protected research lab, and Israel's National Insurance Institute (NII) for allowing restricted access to post-secondary schooling and economic and social outcomes data at adulthood in the NII protected research lab. We benefitted from comments and suggestions from Naomi Hausman, Yoram Mayshar, Analia Schlosser, Moshe Shayo, Sarit Weisburd, Assaf Zussman, and participants in several seminars. The first author acknowledges financial support from the European Research Council through ERC Advance Grant 323439.

1. Introduction

Over the past decades there has been a large increase in female human capital investment and labor force participation. The ratio of male to female college graduates decreased continuously, and in recent years it was even reversed in many countries (Goldin et al. (2006), and Becker et al. (2010)). This trend is partly due to the rise in graduation rates of women in what used to be historically men-dominated fields such as math, science and engineering. The math test score gender gap is of special interest because it was shown to be a good predictor of future income (Murnane et al. (1995) and Paglin and Rufolo (1990)). However, there is still a considerable gender gap in academic outcomes and in employment in math- and science fields. For example, evidence based on recent PISA testing[1] show that in most countries, girls outscore boys in reading while being outscored in math (Machin and Pekkarinen (2008)). This gap is shown to grow during early years of schooling (Fryer and Levitt (2010)), and is larger at the upper tail of the test scores distribution (Ellison and Swanson (2010), Hyde et al. (2008)). Related striking evidence from the UK show that in 2012 about 80% of those who took A level physics in secondary schools were male[2], and that men were awarded 85% of engineering and technology degrees and 82% of computer science degrees while in the same year, 83% medical degrees and 79% of veterinary science degrees went to women.[3] The related employment gaps are even larger, as only 6% of the engineering workforce in the UK is female, only 5.5% of engineering professionals are female and only 27% of engineering and science technicians are female.[4]

What explains these gender disparities in cognitive performance and in math scores in particular is still an open question. Some emphasize the role of biological gender differences in determining gender cognitive differences[5] while others emphasize the social, psychological and environmental factors that might influence this gap. For example, some argue that gender role attitudes

[1] Programme for International Student Assessment (PISA), which surveyed 15-year old students from OECD countries in 2003, 2006 and 2009.
[2] Joint Qualification Council, quoted in The State of Engineering, Engineering UK 2013. HESA, 2010/11, quoted in WISE statistics 2012.
[3] HESA, 2010/11, quoted in WISE statistics 2012.
[4] These statistics on women in engineering compiled by Women's Engineering Society revised February 2014, Joint Qualification Council, quoted in The State of Engineering, Engineering UK 2013.
[5] This approach suggests that the difference in chromosomal determinants (Vandenberg (1968)), hormone levels (Benbow (1988) and Collaer and Hines (1995)) and brain structure (Witelson (1976), Lansdell (1962), Waber (1976)) can explain the evidence that men perform better in spatial tests, whereas women do better in verbal tests.

and stereotypes influence the gender gap by shaping the way parents raise their children[6]; by affecting educational environment at school and teachers' attitudes[7]; and by determining social and cultural norms.[8] This debate is based on limited credible scientific evidence because it is difficult to disentangle the impact of biological gender dissimilarities from environmental conditions and also because it is difficult to measure stereotypes and prejudices and test their causal implications.

In this paper we measure and test the effect of gender bias in a schooling environment. We estimate the impact of primary school teachers' biases about boys' and girls' math and language ability on cognitive outcomes in later grades. We measure teachers' gender biased behavior by comparing their average marking of boys' and girls' in a "non-blind" classroom exam to the respective means in a "blind" national exam marked anonymously. We view this measure of stereotypical grading bias as reflecting the perception of a teacher about gender cognitive differences and use it as a

---

[6] Different parental treatment and expectations are manifested in several ways, such as different attitude from birth (boy babies are handled more than girl babies, whereas girl babies are spoken to more than boy babies (Lewis and Freedle, 1973) to later stages of childhood (boys receive more encouragement for achievements and competition (Block 1976), and are trained to be more independent (Hoffman 1977); in addition, parents engage in more positive attitude when children engage in gender-appropriate behavior (Block 1976), and instruct their sons and daughters in the different behaviors expected of them by providing them with different toys: boys' are "moveable and active and complex and social;' whereas girls' are "the most simple, passive, and solitary" (Brooks-Gunn and Lewis 1979).

[7] Stereotypical attitudes of teachers towards girls and boys in class are said to affect students' self-image and self-confidence, and substantially influence their future educational outcomes. Such mechanisms have been widely documented in the psychology and sociology literature. For example, teachers are said to treat the successes and failures of boys and girls differently, by encouraging boys to try harder and allowing girls to give up (Dweck et al. (1978) and Rebhorn and Miles (1999)). Sadker and Sadker (1985) suggest that teachers give more attention to boys by addressing them more often in class, giving them more time to respond and providing them with more substantive feedbacks. Teachers are also found to treat boys and girls differently, in particular with regard to math instruction: Hyde and Jaffee (1998) show that math teachers tend to encourage boys to exert independence by not using algorithms and that this rebellious approach pursued by boys is perceived as a sign of promising future in mathematics; girls, on the other hand, are controlled more than boys, and are taught mathematics as a set of rules or computational methods. Leinhardt, Seewald and Engel (1979) find that teachers spent more time training girls in reading and less time in math, relative to boys. In addition, according to the National Center of Education Statistics (1997) girls are less likely than boys to be advised, counseled and encouraged to take mathematical courses.

[8] Social norms and beliefs are said to shape the perception of the appropriate division of roles in the home and family, paid employment and the political sphere (Inglehart and Norris 2003). Guiso et al. (2008) try to assess the relative importance of biological and cultural explanations, by exploring gender differences in test performances across countries. Their identification strategy relies on the fact that biological differences between sexes are much less likely to vary compared to cultural environment. They show that there is a positive correlation between gender equality and gender gap in mathematics achievements according to data from OECD's international tests (PISA 2003) and data that measure gender equality taken from the World Economic Forum's Gender Gap Index (GGI). Moreover, they show that these results are not driven by biological differences across countries, by using a genetic distance measurement between the populations. Pope and Sydnor (2010) and Fryer and Levitt (2010) replicate this methodology for different sets of countries. Also related is Alesina et al. (2013) who examine the historical origins of existing cross-cultural differences in beliefs and values regarding the appropriate role of women in society.

proxy for his/her level of prejudice and discriminatory behavior in class. We show that there is large variation within schools in this measure, and that it has a significant effect on the academic achievements of both genders during middle school and high school in math, science and language and also on the choice of the level of math and science study programs in high school. These high stake choices determine practically whether a student will be able to meet requirements of admission to science and math studies at universities later in life.

The systematic difference between non-blind and blind assessment across groups as a measure of discrimination or stereotypes was pioneered in economics by Blank (1991) and Goldin and Rouse (2000).[9] In economics of education, this approach was first implemented in Lavy (2008) to measure gender bias in grading by teachers and it was followed by others, for example, Björn, Höglin, and Johannesson (2011), Hanna and Linden (2012), Cornwell, Mustard, Van Parys (2013) and Burgess and Greaves (2013), who implemented the same methodology using data from other countries and getting overall similar evidence about teachers' stereotypes/biases. In the present paper, however, we go beyond measuring teachers' biased behavior and focus on their implications for gender differences in human capital formation. We think that this paper is perhaps the first one to link teachers' biased behavior as a source of gender gap in human capital, and in particular of gender differences in math and science studies.[10] In this regard, we focus on the choice made by boys and girls regarding the level of their math and science studies in high school. In the Israeli higher education context, as in many other countries, these choices have important longer term implications for occupational choices at adulthood, because successfully completion of advanced courses in math and science in high school is

[9] 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.

[10] A recent paper by Leslie et al. (2015) argue that women are underrepresented in disciplines whose practitioners believe that innate talent is the main requirement for success, controlling for disciplines' characteristics. This correlation is argued to be partly driven by the negative stereotype against women on this dimension, which are measured based on survey questionnaires. Also related is Reuben et al.(2014) who study the effect of stereotypes in an experimental market, where subjects were hired to perform an arithmetic task that, on average, both genders perform equally well. They find that when the employer had no information other than candidates' physical appearance, women were only half as likely to be hired as men, while revealing information on the candidate arithmetic ability reduced the degree of discrimination, but did not eliminate it completely. Terrier (2014) was recently brought to our attention as it addresses a similar question using a similar methodology as the one used in this paper though with a focus on short term effect in primary schools in France.

a prerequisite for post-secondary schooling in engineering, computer science and so on. We test whether teachers' biases towards one of the sexes, as reflected by a more positive evaluation on the "non-blind" tests relative to the "blind" tests of this group, reinforce this group's future achievements and affect their orientation toward enrollment in advanced math and science studies in high school.

The data we use in this paper enables us to evaluate the impact of gender biases of teachers on students' cognitive outcomes in later years, by following three cohorts of 6th grade students between the years 2002–2004 in Tel-Aviv, Israel. By tracking students from primary school to the end of high school we are able to measure students' exposure to gender biases of teachers in primary school, and to estimate its effect on both 8th grade (middle school) test scores in national tests as well as on the high stakes matriculation exam scores at the end of high school, more than six years after the timing of exposure to biased behavior. In addition, we are able to examine whether this measure of teachers' biases is correlated with certain teachers' characteristics, such as age, ethnicity, marital status and gender composition of own children.

For identification, we rely on the random assignments of teachers and students to classes within a given grade and a given primary school. The within school comparison of students who attended the same primary school but were randomly assigned to different teachers accounts for any observed and unobserved school characteristics. This empirical framework ensures that our measure of teachers' biases is not correlated with students' observables. In support of our identification strategy we also test directly for randomness of class assignment within a given school and we present robustness tests regarding the correlation between our measure of teachers' biases and students' and classes' predetermined characteristics.

Our results suggest that teachers' over-assessment of boys in primary school in a specific subject has a positive and significant effect on boys' achievements in that subject in the national tests administered during middle school and at end of high school, and it has an asymmetric significant negative effect on girls. In addition, we find that the favoring of boys over girls by primary school math teachers also affects the successfully completion of advanced courses in math and science in high school. Teachers' biases that favor boys encourage boys to enroll in advanced math courses while doing the opposite for girls; since these courses are prerequisites for admission to higher education in these subjects, such teachers' stereotypical biases contribute to the gender gap in academic degrees in fields like engineering and computer science, and by implication they also contribute to the gender gap in related occupations. These impacts on human capital outcomes by the end of high school have meaningful economic consequences for quantity and quality of post-secondary schooling and for earnings at adulthood. We also find large spillover effects of stereotypical biases of teachers across subjects, implying that a teacher bias against girls or boys in one subject can have a broader influence on students' achievements in other subjects. In addition, we show that these effects have interesting patterns of heterogeneity by parental years of schooling, parental education gap, ethnicity and birth order of children.

The rest of the paper is organized as follows. In Section 2, we present our data. Section 3 explains the identification and estimation methodologies. We detail our results in Section 4 and Section 5 offers conclusions and policy implications.

2. Data

In this study we use data from the administrative records of the Tel-Aviv municipality's school authority. The baseline sample is sixth-grade students in the city's schools in 2002–2004. Each record contains an individual identifier, a school and class identifier in the sixth grade and students' test scores from school exams in math, English and Hebrew held in 6th grade. These data are merged with Israel Ministry of Education students' registry files that include students' demographic information (gender, ethnicity, number of siblings, and parents' education). We combine this dataset with data from three additional sources:

1) The first is GEMS records (Growth and Effectiveness Measures for Schools - *Meizav* in Hebrew) for the three cohorts that we study. The GEMS records were administered by the Division of Evaluation and Measurement of the Ministry of Education.[11] This dataset includes test scores of fifth and eighth graders from a series of tests (in math, Hebrew and English), which were transformed into z-scores for each year and for each subject to facilitate interpretation of the results, as well as responses of fifth- through sixth-grade students to questionnaire. The GEMS questionnaire records

[11] For more information on the GEMS, see the Division of Evaluation and Measurement website (in Hebrew): http://cms.education.gov.il/educationcms/units/rama/odotrama/odot.htm.

include students' questionnaires addressing various aspects of the class environment. In these questionnaires, students are asked to rate in a 6-point scale ranging from 1 (strongly disagree) to 6 (strongly agree), the extent to which they agree with a series of statements. We select a section that focuses on student social behavior and teachers' behavior in the classroom. The GEMS tests and questionnaires were administered at the midterm of each school year to a representative 1-in-2 sample of all elementary and middle schools in Israel, so that each school participated in GEMS tests and questionnaires once every two years. The proportion of students tested is above 90 percent, and the rate of questionnaire completion is roughly 91 percent.

2) The second is matriculation exam test scores and credit units from the Israel Ministry of Education for the three cohorts that we study.[12] Matriculation exams are a series of national exams in core and elective subjects taken by the students between the tenth and twelfth grades. Students choose to be tested at various levels of proficiency, with each test awarding from one to five credit units per subject, depending on difficulty. Some subjects are mandatory, and, for many, the most basic level is three credit units. Advanced level subjects are those subjects taken at four or five credit units. A minimum of 20 credit units is required to qualify for a matriculation certificate which is a prerequisite for university admission. The average scores in the matriculation certificate, which are calculated by the Higher Education Council, are weighted based on the number of credits units taken (advanced level subjects are also given bonuses: four credit units are awarded a bonus of 12.5 points and five credit units 25 points). All schools in the sample offer an academic track leading to a matriculation diploma. We focus on the following matriculation exam outcomes: test score in math, English and Hebrew (transformed into z-scores by each year and each subject), the probability of receiving a matriculation diploma, the total number of successfully completed exams' units, and the number of successfully completed exams' units in English and in science related subjects (math, physics and computer science).[13]

[12] We access these data at the Ministry of Education protected research lab.

[13] The matriculation exams in math, English and Hebrew are mandatory: the number of credits units required in Hebrew is two credit units, and in math and English students are allowed to choose between the most basic level (three credit units) and the advanced level (four or five credit units). On the other hand, matriculation exams in computer science and physics are optional and the maximal number of credit units that students are able to take in these subjects is up to five credit units. Regarding the scores in the mandatory matriculation exams (in math,

3) The third is data from the Population Registry at the National Insurance Institute (NII) on demographic background of teachers and students.[14] We observe teachers' identifier from GEMS questionnaires for a sub-sample of teachers who are homeroom teachers and teach at least one of the relevant subjects: math, Hebrew or English. Merging teachers' identifier with data from the Population Registry enables us to observe teachers' demographic background such as gender, age, marital status, ethnicity and number and gender of children. Additional demographic information is also obtained for the students (the place of birth of the student's grandparents and the birth order of children).

To construct a measure of teachers' biased behavior we combine the scores from the GEMS 5th grade external exam with those of internal exams held in the middle of 6th grade. The GEMS test scores is a "blind" assessment since the GEMS exams are graded by an independent agency where at no stage are the identity and gender of the student revealed. In contrast, the internal exam is graded by the student's teacher and therefore it is a "non-blind" assessment. We assume that this measure of teacher's stereotypes captures her/his overall perception about gender cognitive differences and we use it as a proxy for her/his level of prejudice and discriminatory behavior in class.[15]

We estimate the impact of teachers' stereotypical biases on students' short term cognitive outcomes (8th grade GEMS national tests), on longer term cognitive outcomes (12th grade matriculation national exams) and on the choice of high level math and science courses.

In addition, we are able to examine whether teachers' biased behavior are correlated with teachers' characteristics and with teacher's behavior that differ across gender which we measure by analyzing 5th and 6th grade GEMS survey[16] replies of boys and girls separately.

The final merged dataset includes the national external test scores (blind) in the 5th grade, the school test scores (non-blind) in the 6th grade, GEMS surveys questions in 5th and 6th grade, national exam GEMS test scores in 8th grade, matriculation exam scores and units of study at the end of high

Hebrew and English), students got zero values in these matriculation exam scores if they did not take the exam, but did take exams in other subjects (in about 10 percent of cases).

[14] We access this data at the protected research lab of the National Insurance Institute.

[15] We note here that both exams are not high stakes tests because they are not used for matters important directly to students (GEMS tests and questionnaires are revealed only at the school level since they are mainly used for educational monitoring purposes).

[16] We note that teachers in elementary school are assigned to the same classes for two years consecutively, so that teachers in both 5th and 6th grades are the same teachers.

school for 2001–2008, 2002–2009 and 2003–2010, and student characteristics. In addition, we also observe teachers' characteristics for a sub-sample of the schools.

Table 1 presents descriptive statistics, and information about sample size, number of schools, and number of classes for the three sixth grade cohorts that we use: 2002, 2003 and 2004. The panel data includes 20 secular elementary schools and 5 secular middle schools each year.[17] The sample includes 867 students (in 33 classes) from the 2002 cohort, 1,127 students (in 41 classes) from the 2003 cohort, and 1,017 (38 classes) from the 2004 cohort. The table indicates that the three cohorts' samples are similar across all background variables: mean parental education, average family size, and ethnicity.

Table A1 in the online appendix presents descriptive statistics for the sub-sample of teachers for whom we have additional demographic information, sample size, and subject of instruction. The sample includes 13 math teachers, 29 Hebrew teachers and 36 teachers who teach both math and Hebrew. English teachers are not part of this sample because none of them also served as a homeroom teacher.

Table 2 presents the means of the internal (non-blind) and external (blind) test scores, the mean of the difference between them, separately for boys and girls. We also present the mean of teachers' biased behavior measured at the student level (defined as the difference between boys' internal and external exams scores (column 3) less the difference between girls' internal and external exams scores (column 6)).

The gender gap in test scores varies substantially by type of exam (internal versus external) and by subject. Girls in primary schools outscore boys in the Hebrew external and internal exams. This implies that there is no teachers' gender grading bias in Hebrew. In math we see a different pattern—girls outscore boys in the external exam and boys outscore girls in the internal exam, implying that teachers over-assess boys relative to girls. In English girls outscore boys in both type of exam, and they are over-assessed relative to boys.

[17] The number of middle schools presented in the table refers only to middle schools with GEMS test scores, which participate in GEMS once every two years. The overall number of middle schools in the sample is 12.

Next we examine whether the apparent gap between non-blind and blind test scores of boys relative to girls is statistically significant, using the same estimation framework used in Lavy (2008). We assume that the students' test scores depend on gender, type of test (non-blind test=1) and their interaction term. Appendix Table A2 presents estimates based on this basic specification. We first run a regression that includes individuals' characteristics and year, subject and class fixed effects, and then a second regression that includes year, subject and students fixed effects. The estimated coefficient of the interaction term, which measures the difference between the non-blind scores of boys and those of girls and that we use as a measure of teachers' biases (measured at the student level and presented in the last column of Table 2), is positive in math, negative in English, and practically zero in Hebrew. While the estimates in Hebrew and English were not statistically different from zero in both regressions, the positive estimate in math is statistically different from zero in the first regression (OLS), and marginally significant in the second (student fixed effect specification). These results suggest that there is systematic bias against girls in marking of math exams.

Table 3 presents the means of both middle schools and high schools test scores in external exams, separately for boys and girls. The gender gap in favor of girls in Hebrew and English external exams persists to a large extent also in middle school and high school: the gap in Hebrew is 0.3 in middle school and 0.2 in high school, while the gap in English is about 0.035 in middle school and in high school. Yet at the same time, the gender gap in favor of girls in math external exams in primary school is reversed in middle school and in high school: from a gap of 0.024 in favor of girls, to a gap of 0.022 in favor of boys.

Table A3 in the online appendix presents the distribution of students across matriculation exams' units of study, for boys and girls separately. Although girls have a higher probability of receiving a matriculation diploma and outnumber boys in overall number of successfully completed matriculation exams' units, boys outnumber girls in English and in science oriented advanced courses. The proportion of boys and girls who successfully completed advanced English studies is 60.5% and 58.2%. The proportion of boys who successfully completed the 5 credit-units course in math is 21.1%, while the respective proportion of girls is 14.1%. In science courses this gender gap is even larger: 21.6% of boys successfully completed advanced physics and 13% in an advanced computer science while the rates for girls are only 8.1% and 4.5%, respectively. In the remaining part of the paper we will test whether these differences in achievements, especially in math scores, and in successfully completion of advanced math and science courses, can partly be explained by exposure to teachers' gender biases during earlier stages of schooling.

The teachers' biased behavior measure is defined at the class level by the difference between boys' and girls' average gap between the school score (non-blind) and the national score (blind). It is estimated uniquely for each subject and the higher it is, the higher the stereotypical bias in favor of boys and against girls. The distributions of this measure by subject are presented in Figure 1. English teachers in primary school over-assess girls (mean is -0.74) and the same pattern is seen for Hebrew teachers (mean is -0.41). Math teachers' assessment in primary school, on the other hand, is on average gender neutral (0.01). However, these means hide quite a large heterogeneity among teachers. The range and the standard deviation of the stereotypical biases measures, which are similar across subjects (SD= 0.45, min= -1.5, max= 1.1), reveal that there is a considerable variation in gender biased behavior among teachers. We will exploit this significant variation and test whether teachers' stereotypes have short and long term effects on students' cognitive outcomes.

3. Identification and Estimation

The main goal of this paper is to investigate how teachers' biases toward one of the genders reinforce this group's future achievements and affect educational choices. As noted, the data allows us to track students from primary school, where students were exposed to different teachers' stereotypical biases, through middle and then high school. We are thus able to examine the implications of this exposure for their human capital formation, in particular test scores in middle school and high school national standardized tests as well as choices of math and science studies in the high school matriculation program. Our main identification strategy relies on the random assignments of students and teachers to classes within a school.[18] Using within-school analysis (primary school fixed effect

[18] The randomness of class composition results from the fact that students' assignment in primary and middle school classes cannot be based on ability, family background or any other students' characteristics. The 1968 Integration Law in Israel clearly states that schools should be the focal point of integration of different socio-economic and ethnic groups in Israeli society. Therefore practices of tracking students in primary or middle schools based on students' characteristics are prohibited.

framework), we compare students that study in the same primary school but were randomly exposed to teachers with potentially different gender biased behavior.

We first test the randomness of class composition in our sample by a series of Pearson Chi-Square (χ^2) tests that check whether the student's characteristics and the class assignment are statistically independent. Based on 37 elementary schools (with two or more classes) and eight characteristics (gender, ethnicity, number of siblings, and level of parents' education) we find that out of 296 p values, only 14 were equal to or lower than 5 percent. This implies that for only 5% of the classes we cannot reject that there is non-random assignment. In addition, of the 37 middle schools in our sample, the p value was equal or lower than 5% in only two schools. We therefore conclude that in our sample of schools and classes there is no evidence of systematic non-random formation of classrooms with respect to students' characteristics.[19] The implication of this evidence is that since there is no difference in all classes within a school in terms of average students' ability or any unobserved characteristic, teachers assignment to classes are also unrelated to observed and unobserved students' background. As a result it is safe to assume that teachers' biased behavior is also not correlated with students' and classes' observable and unobservable characteristics.

In the empirical model we assume that the cognitive achievements of pupils in middle/high school are determined by the following equation:

$$(1)\quad y_{icjt} = \alpha + \beta_s + \delta_j + \gamma_t + \lambda X_{icjt} + \beta_1 DS_{icjt} + \beta_2 CS_{icjt} + u_c + \varepsilon_{icjt}$$

where y_{icjt} denotes the outcome of student *i*, from class c, subject *j* and year *t*; X_{icjt} are the student characteristics; β_s is a school fixed effect; δ_j is a subject fixed effect; γ_t is a year fixed effect; DS_{icjt} is the measure of teachers' biased behavior in subject *j* (direct-subject effect); CS_{icjt} is a measure of the average teachers' biased behavior in the other subjects (other than *j*) and we denote its effect as a cross-subject effect. The error term in the equation includes a class-specific random element u_c that allows for any type of correlation within observations of the same school across classes and an individual random element ε_{icjt}.

[19] See also Lavy (2011) and Lavy and Sand (2014) for evidence that suggest no systematic nonrandom formation of classrooms in primary and middles schools in Israel.

The coefficients of interest are β_1 and β_2. The first captures the direct-subject effect of teacher's biases and the second captures the cross-subject effect of teachers' biases. We will also consider a specification where we include a measure of teachers' average biases in all three subjects, and we denote its effect as the average-subject effect instead of the two separate measures of the direct-subject effect and cross-subject effect. In that case we assume that middle/high school test scores are determined by the following equation:

$$(2)\quad y_{icjt} = \alpha + \beta_s + \delta_j + \gamma_t + \lambda X_{icjt} + \beta_1 AS_{icjt} + u_c + \varepsilon_{icjt}$$

where AS_{icjt} is the average of the teachers' biases in all three subjects. The coefficient of interest in this case is β_1 which captures the average-subject effects of teacher's biased behavior in all three subjects on the outcome in subject *j*.[20] The average-subject effect captures the overall stereotypical biased environment that students are exposed to in primary school.

For the purpose of comparison, we will present first estimates based on a regression specification that includes only year and subject dummies as controls, a second specification that also includes primary school fixed effect and in a third specification where we will include also pupil's characteristic (including the mother's and father's years of schooling, number of siblings, immigration status, and ethnic origin) as controls. These various specifications will provide indirect evidence about whether our measure of teachers' stereotypical biases is correlated with students' predetermined characteristics.

4. Results: Effect of Teachers' Biases

A. Main Results

Table 4 reports the estimated effect of teachers' gender biased behavior on students' academic achievements based on estimating equations 1 and 2. We present the estimates of the direct-subject effect of teachers' stereotypical biases and of their cross-subject effect where both are included jointly in a regression.[21] The estimates based on the sample of boys are presented in columns 1-2 and based

[20] We note that this coefficient is by construction exactly the sum of the direct-subject coefficient and the cross-subject coefficient in a simple OLS regression without controls.

[21] Appendix Table A4 reports the estimates of the direct-subject effect of teachers' stereotypical biases and of their cross-subject effect from two separate regressions. The estimates are very similar to that reported in Table 4, suggesting that the direct-subject and cross-subject effects are not very correlated, though, as will be discussed

on the sample of girls in columns 3-4. In columns 5-6 we present the estimated coefficient of the average over the three subjects of the teachers' biases. Each regression includes subject and year fixed effects. Panel A and B shows results of the estimated effect of teachers' biases on 8th grade GEMS test scores and on matriculation test scores respectively. In both panels test scores in all three subjects (math, English, and Hebrew) are pooled together. These test scores are standardized scores, by year and subject. Panel C shows results of the estimated effect of teachers' stereotypical biases on the probability of receiving a matriculation diploma and Panel D shows results of the estimated effect of teachers' biases on the total number of successfully completed matriculation exams' units.[22]

Short term effects

In Panel A of Table 4, we report results from three different specifications. The simple OLS estimates (first row) are positive and marginally significant for boys for the direct-subject effect (column 1), the cross-subject effect (column 2) and the average-subject effect (column 5); for girls on the other hand, these estimates are considerably lower and are not significantly different from zero. Adding primary school fixed effects to the regressions (second row) do not change the estimates for boys, but it lowers the estimated standard errors and as a result the estimated effects are now significantly different from zero. The estimates for girls in this second specification are now all negative, but only the cross-subject effect and the average-subject effect estimates are significantly different from zero. Remarkably, adding students' characteristics leaves the estimates for boys and for girls unchanged, implying that pupil's characteristics are not correlated with the teacher's stereotypical bias measure once we control for primary school.

The estimated effect of teachers' biases on boys' outcomes are positive—this indicates that teachers' over-assessment of boys' test scores increases their achievements at a later age. The estimate of the direct-subject effect for boys is 0.116 (SE=0.058), the respective estimate of the cross-subject effect is 0.139 (SE=0.083) and the average-subject effect is 0.254 (SE=0.113). Calibrating the effect

---

in Appendix Table A6, these correlations depend on whether it is the same teacher or different teachers instructing multiple subjects.

[22] Panel C reports the estimated effect on the probability of receiving a matriculation diploma based on a linear probability regression. We also estimated logit regressions and we present the estimated marginal effects from this model in Appendix Table A5. Since these marginal effects are similar to the estimates obtained from the linear probability regressions, we focus our discussion here on the later estimates presented in Table 4.

size, increasing a teacher's stereotypical bias in a specific subject from zero (no gender bias) to one (the maximal value observed in the sample), will increase boys' test score in that subject by 0.116 of a standard deviation. Increasing the average biased behavior measures in the other two subjects, from zero to one, will similarly increase boys' test scores in that subject by 0.139 of a standard deviation. If in this scenario, we change the exposure of a male student in all three subjects from no gender bias to the highest stereotypical bias observed in the data, his test score in that subject will improve by 0.254 of a standard deviation.

The estimated effects of all biased behavior measures on girls' test score in 8th grade are negative but only two of the three are precisely measured. The estimated direct-subject effect is -0.071 (SE=0.075), the estimated cross-subject effect is -0.243 (SE=0.098) and the average-subject effect is-0.317 (SE=0.135). These estimates suggest that the overall classroom stereotypical biased environment have a much stronger influence on girls' achievements than just the specific subject teacher's stereotypical bias. In terms of effect size, these estimates indicate that increasing the average stereotypical bias against girls from zero to its maximal value of one will reduce girls' outcomes by 0.317 of a standard deviation.[23]

Long term effects

In the other panels of Table 4, we present evidence of teachers' gender biases effects on several long term educational outcomes: Panel B reports the effects on test score in the high school matriculation exams in the three subjects; Panel C reports the effects on the probability of receiving a matriculation diploma; and Panel D reports the effects on the total number of successfully completed matriculation exams' units. These exams are taken at the end of 12th grade, more than 6 years after 'exposure' to teachers' gender stereotypical biases in primary school. Similar to the pattern we found in Panel A, only a few OLS estimates are significant, though in most cases the estimates for boys are positive, while for girls they are negative. The within school estimation lowers again the estimated standard errors which makes most of the estimates statistically significant. Adding student

[23] Terrier (2014) uses a similar idea to the one we pursue in this paper and in an earlier draft (Lavy and Sand 2014). Her paper focus on whether teachers implement a gender discriminatory behavior in a sample of primary schools in France. In addition, the paper shows that there exists a positive correlation between teachers grading bias in favor of boys' in a specific subject and the gap between boys' and girls' average progress in class in that subjects (within the same school year).

characteristics as additional controls in the regressions leaves again the estimated effect almost unchanged.

Comparing the estimates based on the third specification in Panel B to those in Panel A reveals that the effects of the stereotypical biases measures persist through high school, since most of the point estimates are very much the same. For example, the effect of the overall stereotypical biased environment in the classroom on boys' matriculation scores (0.236) and on boys' GEMS test scores (0.254) are very similar. The effects on girls' matriculation scores have a slightly different pattern: the direct-subject effect of teachers' biases on girls' matriculation scores is -0.086 and significantly different from zero (SE=0.034), whereas the respective cross-subject effect on girls is smaller, -0.080, and less precisely measured (SE=0.064) on girls. The overall effect is -0.166 (SE=0.084).

In both Panel C and Panel D, the estimates for boys are positive and the estimates for girls are negative, and as expected in most cases only the average-subject effects are precisely measured. The overall teachers' biases effect on boys' probability of receiving a matriculation diploma is 0.071 (SE=0.047), and the overall effect on their number of successfully completed matriculation exams' units is 1.945 (SE=0.792). Girls' outcomes are affected in the opposite way (-0.081, SE=0.042 and -1.683, SE=0.851 respectively). These results imply that the overall stereotypical biased environment in the classroom that students are exposed to in primary school increases boys' probability of receiving a matriculation diploma and their total number of successfully completed matriculation exams' units while lowering that of girls.

B. Estimated Effects by Subject

Short term effects

In this section we present and discuss results of estimating the effect for each subject separately. In Table 5, we present evidence based on estimating a separate regression for each subject, using the specification of a regression that includes students' characteristics, year and school fixed effects. As before, we present the estimates of the direct-subject effect of teachers' biases and of their cross-subject effect from one joint regression, for boys and for girls separately. The last two columns show the estimated coefficient of the average effect of teachers' biases in all three subjects from separate regressions for boys and girls. In Panel A the dependent variable is 8th grade GEMS test scores whereas in Panel B the depended variable is matriculation test scores.

In Panel A of Table 5, the estimates of the direct subject effect of teachers' biases are relatively small and not significantly different from zero for both genders, except in mathematics. The estimated effect of math teachers' biases on boys’ 8th grade math test scores is the largest direct-subject effect and is positive and significant, 0.374 (SE=0.142). The estimated effect of math teachers' stereotypical biases on girls’ math test scores is also the largest direct-subject effect, though it is not precisely measured, -0.135 (SE=0.143). This result suggests that the 8th grade students' test scores in math are mainly affected by their math teachers' biases. In Hebrew and English[24], the cross-subject effects are larger than the direct effect on students' 8th grade test scores for both genders. The average-subject effect is significant in three of the six estimates and marginally so in two others, indicating that the overall stereotypical biased environment is positive and significant for boys in math and Hebrew; while the opposite is true for girls in English.

Long term effects

In Panel B of Table 5, we present the estimated effect on the matriculation score by subject. Focusing on the overall effect presented in columns 5-6, we note that the estimated effects of teachers’ biases on boys’ test scores in all subjects are positive and significantly different from zero at the 5 percent level of significance. The respective effect on girls is negative in all subjects and it is significantly different from zero at the 5 percent level of significance for the Hebrew and math test scores. These estimates indicate that increasing the average stereotypical bias against girls from zero to its maximal value of one generates a 0.46 standard deviation gap in math test scores in the matriculation exam in favor of boys (0.263+0.195). The respective magnitudes for Hebrew and English are equally large.

C. Effects on Choice of Advanced Courses in Math and Science

The evidence on the effects of teachers' stereotypical biases on students’ successfully completion of advanced courses in science, math and English in high school (equivalent to honors

[24] We note that the estimated effect of teachers’ stereotypical biases on English outcomes is based on a smaller sample (311 observations) because of missing GEMS test scores in English in eighth grade in 2006.

classes in the US) are presented in Table 6 and Table 7. Table 6 presents the estimated effect of teachers' biases on the probability of successfully completing such courses and Table 7 presents the estimated effect of teachers' biases on the total number of matriculation credits a student gained in each of these advanced courses. We note that an advanced class yields 5 matriculation credits and a basic class yields only 3 matriculation credits. In science we included advanced computer science and physics courses. Both tables present evidence based on estimating a separate regression for each subject, using the specification that includes students' characteristics, year and school fixed effects. As in earlier tables, we present the estimates of the direct-subject effect of teachers' stereotypical biases[25] and of their cross-subject effect, for boys and for girls separately. In columns 5-6 we present the estimates for the overall exposure in all three subjects.

Table 6 presents estimates from linear probability regressions. We also estimated logit regressions and we present the marginal effects estimated from this model in Appendix Table A5. Since these marginal effects are very similar to the estimates obtained from the linear probability regressions, we focus our discussion here on the later estimates presented in Table 6. The estimated effect of math teachers' biased behavior on the probability of successfully completing advanced studies in math (4 or 5 credit units) is positive and significant for boys (0.093, SE=0.049) and negative and significant for girls (-0.073, SE=0.044). The respective estimates in English and science are not precisely measured, though they are in most cases positive for boys and negative for girls. The estimated average-subject effects are positive and significant for boys in English and in math and they are negative but not different from zero for girls. In order to assess the magnitude of the effect for boys, we simulate a situation where a boy is moved from a neutral teachers' stereotypical biased environment to one with a boys' bias of one. This will increase the completion rate of boys in advanced math studies by 11 percent and in advanced English program by 5.8 percentage point.

Table 7 presents the estimated effect of teachers' stereotypical biases on students' total number of matriculation credits gained in these study programs. The average-subject estimated effects on math credits are significant for both boys and girls. For boys the estimated effect is also significant on

[25] We note that the direct-subject effect of teachers' biased behavior on students' science test scores (in both computer science and physics courses) refers to the effect of math teachers' stereotypical behavior on science test scores.

physics and English matriculation units. As before, we can simulate the impact of moving from a neutral teachers' stereotypical biased environment to one with a boys' bias of one. Such change will increase boys' number of matriculation units in math by 0.338, and decrease girls' number of math units by 0.291. Similarly, it will increase boys' number of matriculation units in physics by 0.365 and in English by 0.257.

The estimates of the direct-subject effect in math are of special interest because of the considerable gender gap in math achievements and its impact on future labor market outcomes.[26] Based on our evidence in Appendix Table A2, simulating a 0.07 decrease in the math teachers' biased behavior (see Table A2 in the appendix) will decrease boys' math achievements in middle school by 0.026 standard deviations, and as a result will also eliminate the positive gender gap in favor of boys in math achievements in middle school (0.024). Such change will also decrease advanced math studies' completion rates of boys in in high school by 0.7 percentage point and will increase girls' completion rates by 0.5. As a result, the gender gap in studying math at the highest level in high school will decline from 3 to 1.8. A more drastic decline in math teachers' biases, say a decrease of one standard deviation of the math teachers' stereotypical biases measure (0.4), will reverse the gender gap in math achievements in middle school from a positive gap of 0.024 SD in favor of boys to a negative gap of 0.126 SD in favor of girls. The same change will also change the gender gap in advance math studies' completion rates of from 3 in favor of the boys, to 3.6 in favor of girls.

The long term effect on high school matriculation study program and test scores have meaningful economic consequences for quantity and quality of post-secondary schooling and on earnings at adulthood because. In Table A6 in the online appendix we present results of regressions of three key matriculation exams' outcomes on post-secondary enrollment and attainment and on earnings at age 30, based on a sample of older cohorts of Tel Aviv high school graduates. Each of the three outcomes is a good predictor of the various outcomes at adulthood. All three matriculation exams' outcomes are positively and significantly correlated with enrollment and attainment of post-secondary schooling in general and by quality (university schooling, academic colleges and other).

[26] In addition, these results are consistent with the findings regarding gender biased behaviors of teachers especially with regard to math instruction, as have been documented in the psychology and sociology literature (reviewed in the introduction section).

They are also positively correlated with annual earnings at age 30; for example, each credit unit is associated with a gain of ILS 1,270 ($343) per annum and having a matriculation certificate is associated with a ILS 15,648 ($4,230).

D. Robustness and Falsification Tests

We turn next to discuss results that support our identification strategy and our interpretation of the evidence presented above. As implied from the evidence presented in Table 4, our measure of teachers' stereotypical biases is not correlated with students' predetermined characteristics; we further examine directly the correlation of this measure with average classroom predetermined characteristics. We first test how sensitive are the treatment estimates of Table 4 to adding some classroom level controls, such as, the proportion of boys in class, the difference between the boys' and girls' mean blind test scores in class and the difference between boys' and girls' violent behaviors in class.[27] We also examine to what extent the measure of teacher's stereotypical bias is a characteristic of the teacher and not of the classroom by comparing the correlations between the biases measures of the same class in different subjects where these subjects are taught by the same teacher versus when they are taught by different teachers. Finally, we examine whether our results are derived by the fact that girls outscore boys in 5th grade national exams by replicating the same analysis for higher/lower achievers and test whether teachers' encouragement of low achievers leads to similar results.

Appendix Table A7 presents the estimated effect of teachers' stereotypical biases on test scores when classroom level controls are added to the regressions. We report only the estimated average-subject effect of teachers' stereotypical biases on 8^{th} grade GEMS test scores (Panel A) and matriculation exams test scores (Panel B), separately for boys and girls. The test scores in all three subjects (math, English, and Hebrew) are stacked together, and each regression includes students' characteristics and year and school fixed effects as controls. In the first row of both panels, the proportion of boys in class is added as a control, and in the second row, the difference between boys' and girls' mean blind test scores in class is added as a control. In the last row of Panel A, the difference

[27] The level of violence was based on students' report on classroom environment, available from GEMS questionnaire survey. Students were asked the extent to which they agree with the following statement: "I was involved in violence (physical fights) in school many times this year". Note that since GEMS questionnaire data were not included in the data available in the ministry of education lab, this control was added only in Panel A.

between boys' and girls' violent behaviors in class is added as control. The estimated effects of the average stereotypical bias measure on boys’ and on girls' outcomes are very similar to the estimates of our preferred specification (last row of Table 4) in both panels. Adding classrooms level controls to the regression leads to only minor changes in the estimates. We think that these results provide further evidence that the measure of teachers’ biased behavior is not correlated with classrooms' characteristics.

Appendix Table A8 presents additional evidence regarding the measure of teachers’ stereotypical attitude being a characteristic of the teacher and not of the classroom. The Table presents the correlations between stereotypical biases of teachers by subjects of instruction: The correlations in each row in columns 1-2 are the correlation between teachers' biases measures across subjects from the overall sample; The correlations in each row in columns 3-4 are similar to those in columns 1-2, but the school means of teachers biases measures in each subject are netted out. The correlation in column 5 is between biases measures of the same teachers who instruct students from the same class in both math and Hebrew; and the correlation in column 6 is between biases measures of different teachers who instruct students from the same class in both math and Hebrew. Comparing the correlation in columns 1-4 suggests that the correlation between math teachers' biased behavior and Hebrew teachers' biased behavior is higher than the correlation between teachers' biased behavior in other subjects. Since most math teachers instruct Hebrew as well, and no English teachers instruct the other two subjects (Hebrew/math) (see Table A1), this finding reinforces our assumption regarding teachers' biased behavior not being correlated with students’ and classes’ observable or unobservable characteristics. This assumption is further strengthened by comparing explicitly the correlation between the biases measures of the same teacher (0.654) to the biases measures of different teachers (0.140).

Appendix Table A9 presents the estimated effect of teachers' attitude towards low achievers on 8th grade GEMS test scores in math, English, and Hebrew, separately for low and high achievers. Higher/lower achievers are defined as students with higher/lower scores in 5th grade than the average (i.e., their mean scores in all three subjects are higher/lower than zero). The measure of teachers' attitude towards low achievers is defined at the class level by the difference between low performing students' and high performing students' average gap between the school score (non-blind) and the national score (blind). Test scores in all three subjects (math, English, and Hebrew) are stacked together, and similarly to Table 4 we present result from three specifications. The estimates of the direct-subject effect of teachers' attitude and of their cross-subject effect based on the sample of low performing students are presented in columns 1-2 and based on the sample of high performing students in columns 3-4. In columns 5-6 we present the estimated coefficient of the average over the three subjects of teacher's attitude, for low and high performing students. According to Table A9, the estimates of the effect of teachers' attitude, on both high and low achievers are not significantly different from zero when primary school controls are added to the regressions. This finding suggests that our results are not the consequence of girls outscoring boys in 5th grade national exam, which might have been the reason for teachers' over-encouragement of boys because of their lower grades, leading them to better performances.

E. Heterogeneous Treatment Effects of Teachers' Biases

In order to gain further insight into the effects of teachers' gender biased behavior on students' academic success we explore these heterogeneous effects across different dimensions. In Table 8 we present the estimated effect of the average over the three subjects of the teacher's stereotypical bias on GEMS test scores[28] for boys and for girls separately, based on different stratifications of the full sample. We use the specification that includes students' characteristics, year and school fixed effects. The first part reports the heterogeneous treatment effects of teachers' stereotypical biases by mother's education level (whether mother's years of schooling is above the median – 12 years); the second reports the heterogeneous treatment effects by the gap in parental education (high parental education gap is defined as a non-negative gap between the fathers' and the mothers' years of schooling); the third reports the heterogeneous treatment effects by ethnicity (whether grandparents' place of birth is

[28] Appendix Table A10 presents the estimates of the average-subject effect on matriculation test scores based on only two stratifications of the full sample (by parental education and by the gap in parental education), since the matriculation test scores that are available at the Ministry of Education lab could not be merged with the additional demographic information from the population registry (available only at National Insurance Institute lab). Comparing the estimates of Table 8 to the estimates of Table A10, reveals similar effects of teachers' stereotypical biases on both GEMS and matriculation tests scores by parental education; though the effect of teachers' stereotypical biases on matriculation test scores is more pronounced for girls with low parental education gap, rather than with high parental education gap as is the case according to GEMS tests scores.

Asia/Africa) and the last reports the heterogeneous treatment effects by the child birth order (whether the student is a firstborn child).

The estimates presented in Panel A of Table 8 indicate that the overall stereotypical biased environment is positive and significant only for boys with high parental education; while the opposite is true for girls with low parental education. The estimated average-subject effect for boys with high parental education is 0.492 (SE=0.193), while the estimated average-subject effect for girls with low parental education is -0.407 (SE=0.212). In contrast, the average-subject effects for boys with low parental education and for girls with high parental education are of similar signs but are not significantly different from zero. These results seem partly counterintuitive since one would expect that the overall stereotypical environment to which students are exposed to would be more influential among students with low parental education for both genders and not only for girls. Moreover, as the relevant sociology and psychology literature suggest mother's employment status is correlated with a more egalitarian gender role attitude[29], students of educated mothers should be less influenced by teachers' stereotypical biases. Therefore, we also consider the heterogeneous treatment effects of teachers' stereotypical biases based on a slightly different stratification of the sample where we divide students based on the within family parental education gap. We postulate that a child from a family with lower parental education gap could be less susceptible to the effect of external gender stereotypes in school.

The treatment effects of teachers' stereotypical attitudes by parental education gap are presented in panel B of Table 8. Indeed, the table indicates that the overall stereotypical biased environment effect is more important for students with high parental education gap, in particular when the student's father has more education than the mother. This is true for boys and girls. The estimated average-subject effect is positive and significant for boys with a high parental education gap (0.298, SE=0.158); while it is negative and significant for girls with a high parental education gap (-0.431, SE=0.168). On the other hand, the average-subject effect for students with low parental education gap

[29] See, for example, Hoffman (1977) and Herzog et al (1983).

is not significantly different from zero. These results suggest that students from a more gender-equal home environment are indeed less influence by their teachers' stereotypical biases.

Panel C in Table 8 presents the heterogeneous treatment effects by ethnicity—ethnic origin Asia-Africa versus all others, which includes mainly Europe-North America origin. This division proxies a division by income and wealth as well as other socioeconomic background characteristics. For example, the Asia-Africa ethnic group has a much lower level of parental education in comparison to the North America/Europe ethnic group. We also note that Jewish families from Africa/Asia ethnicity tend to be more patriarchal with an enhanced role for the male in family matters and decision making. Such a growing up environment might be more conducive for children being affected by gender biases at school. Indeed the estimated effects by ethnicity have a similar pattern to those by parental education. The estimated average-subject effect is negative and significant for girls of Asia-Africa ethnicity (-0.556, SE=0.215); while the opposite is true for boys from other ethnic groups (0.289, SE=0.135).

Panel D of Table 8 reports estimated treatment effects by child birth order. This stratification is not common, but much has been argued about the impact of birth order on children's' personality and behavior, especially with regard to firstborn children who are said to be a more sociable, dependent and conforming.[30] We therefore posit that a stereotypical biased environment could affect children differently by their birth order. Interestingly, the results suggest that firstborn children of both sexes tend to be less influenced by teacher biases: the average-subject effect is significant only among non-firstborn children. The estimated effect on non-firstborn boys is positive and significant (0.313, SE=0.185)); whereas the estimated effect on non-firstborn girls is negative and significant (-0.417, SE=0.189)).

F. Teacher's Characteristics and Teachers' Classroom Behavior

Using administrative data from NII we are able to further explore the measure of teachers' gender biases and its correlation with characteristics of teachers. In Table 9 we examine the correlations between teachers' biased behavior and teachers' characteristics based on a sub-sample of

[30] See Adams (1972) for a review of the literature.

teachers that we can match to NII data. The estimates are from a separate regression for each of the teachers' characteristics that we have, using a simple OLS regression with year and subject fixed effects. We postulate that the findings signify that such correlations will be an indication that the measure of stereotypical bias that we use indeed reflect teachers attitude, and not students' attributes.

Older and single teachers seem to favor boys over girls: the coefficient of a dummy indicator of being older than 50 years old is positive and significant (0.206, SE=0.104), and so is the estimate of the indicator for single teachers (0.315, SE=0.202). The estimated coefficient for teachers from Europe-North America origin is negatively and significantly correlated with teachers' biases (-0.204, SE=0.113). The other individual characteristics that we examined are being married (positive but insignificant) and the number of children and the proportion of daughters, both of which have negative coefficient but not significantly different from zero.[31]

In Table 10, we test whether teachers' over-assessment of boys according to our bias measure is correlated with teaching behaviors that differ across gender. The information about teachers' behavior towards boys and girls in the classroom is based on students' report of classroom environment and teachers' attitudes (of all teachers) available from five GEMS questionnaire items. In these questions, students are asked to rank from 1 (strongly disagree) to 6 (strongly agree) the extent to which they agree with a series of statements. These are the items we use in this analysis: (1) "Teachers help students to learn their subjects of interest"; (2) "Sometimes teachers insult me"; (3) "There are good relations between teachers and students"; (4) "The relations between students and teachers are of mutual respect"; and (5) "I am satisfied in school".

In Table 10 we report estimates of the effect of the overall classroom stereotypical biased environment on each of these five behavioral outcomes, using regressions that include students' characteristics, and school and year fixed effects. We estimate these regressions for boys and girls separately. We note that the measure of the overall teachers' gender biased behavior captures only the behavior of math, English and Hebrew teachers while the survey questions guide the students to refer to all of their teachers. The estimates in Table 10 suggest teachers' biased attitude against girls leads

[31] We note that all identified teachers in our sample are female, thus we could not test the correlation of teachers' gender with the measure of teachers' stereotypical biases.

girls to feel that they are getting less support from their teachers (-0.333, SE= 0.190), and a bias against boys make them sense an insulting attitude from their teachers (-0.381, SE=0.262), though this effect is only marginally significant. The estimated effects on more general questions regarding good student-teacher relations and overall satisfaction in class are not significantly different from zero for both genders.

5. Conclusions

In this paper we investigate how primary school teachers' biases toward one of the genders reinforce this group's future academic achievements and orientation toward enrollment in advanced math and science studies in high school. We base the measure of teachers' gender-biased behavior on a comparison of primary school classroom boys' and girls' average test scores in a "non-blind" exam that the teacher marks versus a "blind" exam marked externally. We then estimate the impact of this measure of teachers' stereotypical biases on the academic achievements of students in standardized national exams during middle school and high school, and on completion of higher level courses in math and sciences during high school.

For identification, we rely on the random assignments of teachers and students to classes within a given primary school. Comparing students in the same primary school who are exposed to different teachers who might have different patterns of gender stereotypical biases, eliminates any selection bias due to observed and unobserved teacher, class and school characteristics. We support this identification approach with evidence that clearly indicates that teachers' biased behavior are indeed not correlated with students' and classes' predetermined characteristics.

The results we present suggest that teachers' over-assessment of boys in a specific subject has a positive and significant effect on boys' overall future achievements in that subject while having a significant negative effect on girls. We also provide evidence that suggests spillover effects across biased behavior of teachers of different subjects can also impact students' achievements in other subjects. These effects persist through middle school and high school and actually have dramatic implications for end of high school matriculation exam scores and on the probability of receiving a matriculation diploma. Interestingly, we find that teachers' biases have a greater influence on students from families where the student's father has more education than the mother, as well as on students who are youngest among their siblings; and girls with low parental schooling or from an Asia/Africa ethnicity.

We also find that favoritism of boys among math and science teachers has an especially large and positive effect on boys math test score and on their successfully completion of advance math and science studies in high school; the respective effect on girls is negative and statistically significant. The estimates of the direct-subject effect in math are of special interest because of the considerable gender gap in math achievements and its impact on future labor market outcomes. Moreover, since this gap in math achievement partly results from teachers' stereotypical biases against girls in mathematics, eliminating these biases will go a long way toward reducing the math achievements gender gap, and it will also decrease the gender gap in enrollment in advanced math studies. The impact on the various end of high school matriculation outcomes carries meaningful economic consequences because these high stakes outcomes affect sharply the quantity and quality of post-secondary schooling and impact earnings at adulthood as well.

6. References

Alesina, Alberto, Paola Giuliano, and Nathan Nunn. 2013. "On the Origin of Gender Roles: Women and the Plough", *Quarterly Journal of Economics* 128, no. 2: 469-530.

Adams, B.N., 1972, "Birth Order: A Critical Review", Sociometry, 35(3), 411-439.

Bae, Y. and T.M. Smith, 1997, "Women in Mathematics and Science". Findings from "the Condition of Education", *National Center for Education Statistics 1997*, no. 11.

Becker, G.S., W.H. Hubbard and K.M. Murphy, 2010, "Explaining the Worldwide Boom in Higher Education of Women", *Journal of Human Capital* 4, 203-241.

Benbow, C.P., 1988, "Sex-Related Differences in Precocious Mathematical Reasoning Ability: Not Illusory, not Easily Explained", *Behavioral and Brain Sciences* 11, 217-232.

Björn, T.H., Höglin, E. and M. Johannesson, 2011, "Are boys discriminated in Swedish high schools?", *Economics of Education Review* 30*(*4), 682-690.

Blank, R.M., 1991, "The Effects of Double-Blind versus Single-Blind Reviewing: Experimental Evidence from the American Economic Review", *American Economic Review* 81, 1041-1067.

Block, J.H., 1976, "Issues, Problems, and Pitfalls in Assessing Sex Differences: A Critical Review of The Psychology of Sex Differences", Merrill-Palmer *Quarterly of Behavior and Development* , 283-308.

Brown, C. and M. Corcoran, 1997, "Sex-Based Differences in School Content and the Male-Female Wage Gap", *Journal of Labor Economics* 15, 431-465.

Burgess, S. and E. Greaves, 2013, "Test Scores, Subjective Assessment, and Stereotyping of Ethnic Minorities", *Journal of Labor Economics* 31, 535-576.

Collaer, M.L. and M. Hines, 1995, "Human Behavioral Sex Differences: a Role for Gonadal Hormones during Early Development?", *Psychological Bulletin* 118, 55.

Cornwell, C., D. Mustard and J. Van Parys, 2013, "Non-cognitive Skills and Gender Disparities in Test Scores and Teacher Assessments: Evidence from Primary School", *Journal of Human Resources*, 48(1), 236-264.

Dweck, C.S., W. Davidson, S. Nelson and B. Enna, 1978, "Sex Differences in Learned Helplessness: The Contingencies of Evaluative Feedback in the Classroom and An Experimental Analysis", *Developmental Psychology* 14, 268.

Ellison, G. and A. Swanson, 2010, "The Gender Gap in Secondary School Mathematics at High Achievement Levels: Evidence from the American Mathematics Competitions", *Journal of Economic Perspectives* 24, 109-128.

Friedman, L., 1989, "Mathematics and the Gender Gap: A Meta-Analysis of Recent Studies on Sex Differences in Mathematical Tasks", *Review of Educational Research* 59, 185-213.

Fryer, R.G. and S.D. Levitt, 2010, "An Empirical Analysis of the Gender Gap in Mathematics", *American Economic Journal: Applied Economics* 2, 210-240.

Gneezy, U., M. Niederle and A. Rustichini, 2003, "Performance in Competitive Environments: Gender Differences", *The Quarterly Journal of Economics* 118, 1049-1074.

Goldin, C., "A Grand Gender Convergence: Its Last Chapter", *American Economic Review*. Forthcoming 104.

Goldin, C., L.F. Katz and I. Kuziemko, 2006, "The Homecoming of American College Women: The Reversal of the College Gender Gap", *Journal of Economic Perspectives* 20, 133-156.

Goldin, C. and C. Rouse, 2000, "Orchestrating Impartiality: The Impact of "Blind "Auditions on Female Musicians", *The American Economic Review* 90, 715-741.

Guiso, L., F. Monte, P. Sapienza and L. Zingales, 2008, "Culture, Gender, and Math", *Science* 320, 1164- 1165.

Hanna, R.N., and L.L., Linden, 2012, "Discrimination in Grading", *American Economic Journal: Economic Policy*, 4(4), 146-68.

Herzog, A.R., Bachman, J.G., and L.D., Johnston, 1983, "Paid Work, Child Care, and Housework: A National Survey of High School Seniors' Preferences for Sharing Responsibilities Between Husband and Wife", *Sex Role*, 9(1), 109-135.

Hoffman, L. W., 1977, "Changes in Family Roles, Socialization, and Sex Differences", *American Psychologist*, 32(8), 644-657.

Hyde, J.S., Lindberg, S.M., Linn, M.C., Ellis, A.B., and Williams, C.C, 2008, "Diversity. Gender similarities characterize math performance", *Science* 321(5888), 494–495.

Hyde, J.S., and S. Jaffee, 1998, "Perspective from Social and Feminist Psychology", *Educational Research* 27 (5), 14-16.

Inglehart, R. and P. Norris, 2003, "Explaining the Rising Tide of Gender Equality", In Inglehart, R. and P. Norris, Rising Tide: Gender Equality and Cultural Change Around the World. (Cambridge University Press)

Lansdell, H., 1962, "A Sex Difference in Effect of Temporallobe Neurosurgery on Design Preference", *Nature* 194, 852-854.

Lavy, V., 2008, "Do Gender Stereotypes Reduce Girls' or Boys' Human Capital Outcomes? Evidence from a Natural Experiment", *Journal of Public Economics* 92, 2083-2105.

Lavy, V., 2011, "What Makes an Effective Teacher? Quasi-Experimental Evidence", Forthcoming *CESifo Economic Studies.*

Lavy, V. and E. Sand “On the Origins of the Gender Human Capital Gap: Short and Long Term Effect of Teachers’ Stereotypes”, Draft, Applied Micro Seminar, Department of Economics, Hebrew University of Jerusalem, July 2014.

Leinhardt, G., A.M. Seewald and M. Engel, 1979, "Learning what's Taught: Sex Differences in Instruction", *Journal of Educational Psychology* 71, 432-439.

Leslie, S.J., A. Cimpian, M. Meyer and E. Freeland, 2015, “Expectations of brilliance underlie gender distributions across academic disciplines”, *Science* 347, 262-26.

Lewis, M. and J. Brooks-Gunn, 1979, "Towards a Theory of Social Cognition: The Development of Self", *New Directions for Child and Adolescent Development*, 4, 1-20.

Lewis, M. and R. Freedle, 1972, "Mother-Infant Dyad: The Cradle of Meaning", In P.K., Pliner and T. Lester Alloway (Eds), "Communication and Affect: Language and Thought", Oxford England: Academic Press.

Machin, S. and T. Pekkarinen, 2008, "Global Sex Differences in Test Score Variability", *Science* 322, 1331-1332.

Murnane, R.J., J.B. Willett and F. Levy, 1995, "The Growing Importance of Cognitive Skills in Wage Determination", *Review of Economics and Statistics* 77, 251-266.

Niederle, M. and L. Vesterlund, 2007, "Do Women Shy Away from Competition? Do Men Compete Too Much?" *The Quarterly Journal of Economics* 122, 1067-1101.

Paglin, M. and A.M. Rufolo, 1990, "Heterogeneous Human Capital, Occupational Choice, and Male-Female Earnings Differences", *Journal of Labor Economics* 8, 123-144.

Pope, D.G. and J.R. Sydnor, 2010, "Geographic Variation in the Gender Differences in Test Scores", *The Journal of Economic Perspectives* 24, 95-108.

Reuben, E., Sapienza P. and L. Zingales, 2014, 'How Stereotypes Impair Women's Careers in Science," *Proceeding of the National Academy of Science,* Forthcoming.

Sadker, M. and D. Sadker, 1986, "Sexism in the Classroom: From Grade School to Graduate School", *Phi Delta Kappan* 67, 512-515.

Spolaore, E. and R. Wacziarg, 2009, "The Diffusion of Development", *The Quarterly Journal of Economics* 124, 469-529.

Terrier, C., 2014, "Giving a Little Help to Girls? Evidence on Grade Discrimination and its Effect on Students Achievement", PSE Working Papers n. 2014-36

Vandenberg, S. G. 1968, "Primary Mental Abilities or General Intelligence? Evidence from Twin Studies", In J.M. Thoday and A.S. Parkers (Eds), "Genetics and Environmental Influences on Behaviour", New York: Plenum.

Voyer, D., S. Voyer and M.P. Bryden, 1995, "Magnitude of Sex Differences in Spatial Abilities: a Meta-Analysis and Consideration of Critical Variables", *Psychological Bulletin* 117, 250.

Waber, D.P., 1976, "Sex Differences in Cognition: a Function of Maturation Rate?", *Science* 192, 572-574.

Wilder, Gita Z., and K. Powell, 1989, "Sex Differences in Test Performance: A Survey of Literature". No. 89. New York: College Entrance Examination Board.

Witelson, D.F., 1976, "Sex and the Single Hemisphere: Specialization of the Right Hemisphere for Spatial Processing", *Science* 193, 425-427.

Table 1: Summary Statistics of Students' Characteristics by Cohort

	2002	**2003**	**2004**
	(1)	(2)	(3)
Mean Father's Education	13.477	13.339	12.992
	(3.391)	(3.468)	(3.482)
Mean Mother's Education	13.614	13.610	13.287
	(3.073)	(3.115)	(3.116)
Mean Number of Siblings	2.190	2.336	2.259
	(0.996)	(1.039)	(1.130)
Proportion of Asia/Africa Ethnicity	0.114	0.110	0.103
	(0.318)	(0.313)	(0.304)
Proportion of Europe/America Ethnicity	0.171	0.182	0.189
	(0.376)	(0.386)	(0.392)
Proportion of Israel Ethnicity	0.611	0.615	0.601
	(0.488)	(0.487)	(0.490)
Proportion of Former Soviet Union	0.081	0.063	(.083
	(0.273)	(0.244)	(0.276)
Number of Students	867	1127	1017
Number of Elementary Schools	17	20	20
Number of Elementary School Classes	33	41	38
Number of Middle Schools	5	7	5
Number of Middle School Classes	26	32	31

Notes: Each column is based on one different cohort of sixth grade students. Number of middle schools and middle school classes refers only to middle schools with GEMS test scores. Standard deviations are reported in parentheses.

Table 2: Means and Standard Deviations of National and School Primary School Exams Scores and Teachers' Biases Measure at the Student Level, by Gender

	Boy			Girl			Teachers' Biases Measure at the Student Level
	School Score Exams	National Score Exams	Difference Between School and National Exams Scores	School Score Exams	National Score Exams	Difference Between School and National Exams Scores	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Hebrew	-0.099	-0.136	0.037	0.168	0.139	0.029	0.008
	(1.001)	(1.022)	(1.047)	(0.951)	(0.952)	(0.925)	
Math	0.052	-0.014	0.066	0.003	0.014	-0.011	0.077
	(0.985)	(1.034)	(0.960)	(0.971)	(0.963)	(0.903)	
English	-0.036	-0.047	0.010	0.113	0.049	0.064	-0.054
	(1.002)	(1.036)	(0.999)	(0.940)	(0.962)	(0.933)	
Number of Students	4246	4246	4246	4122	4122	4122	8368

Notes: The national exams scores and the school scores are standardized scores. The number of students refers to the number of students in all three subjects. The teachers' biases measured at the student level are equal to the difference between boys' school and national exams scores (column 3) less the difference between girls' school and national exams scores (column 6). Standard errors are reported in parentheses.

Table 3: Means and Standard Deviations of National Exams Scores in Middle School and High School at the Student Level, by Gender

	Boy	Girl	Boy	Girl
	National Score Exams	National Score Exams	National Score Exams	National Score Exams
	(1)	(2)	(3)	(4)
	Middle School		High School	
Hebrew	-0.147	0.151	-0.101	0.097
	(1.061)	(0.908)	(1.055)	(0.932)
Math	0.012	-0.012	0.011	-0.011
	(1.043)	(0.952)	(1.060)	(0.937)
				0.022
English	-0.017	0.019	-0.020	0.019
	(1.004)	(0.995)	(1.039)	(0.959)
Number of Students	1490	1406	3883	4033

Notes: The national exams scores are standardized scores. The number of students referees to the number of students in all three subjects. Matriculation test scores are weighted based on the number of credits units taken, as computed by the Ministry of Education. Standard errors are reported in parentheses.

Table 4: Estimated Effect of Teachers' Biases on Educational Outcomes

	Boy		Girl		Boy	Girl
	Direct-Subject Effect	Cross-Subject Effect	Direct-Subject Effect	Cross-Subject Effect	Average-Subject Effect	Average-Subject Effect
	(1)	(2)	(3)	(4)	(5)	(6)
A. 8th Grade GEMS Test Scores						
OLS	0.110	0.103	0.031	-0.019	0.214	0.012
	(0.084)	(0.132)	(0.097)	(0.153)	(0.182)	(0.220)
6th Grade School Fixed Effects	0.118	0.112	-0.064	-0.249	0.229	-0.316
	(0.056)	(0.085)	(0.076)	(0.105)	(0.115)	(0.145)
6th Grade School Fixed Effects and Student Characteristics	0.116	0.139	-0.071	-0.243	0.254	-0.317
	(0.058)	(0.083)	(0.075)	(0.098)	(0.113)	(0.135)
Number of Students	1420	1420	1317	1317	1420	1317
B. Matriculation Test Scores						
OLS	0.085	0.214	-0.050	-0.036	0.299	-0.086
	(0.062)	(0.122)	(0.063)	(0.107)	(0.176)	(0.159)
6th Grade School Fixed Effects	0.073	0.190	-0.118	-0.162	0.262	-0.280
	(0.036)	(0.062)	(0.038)	(0.067)	(0.083)	(0.088)
6th Grade School Fixed Effects and Student Characteristics	0.071	0.165	-0.086	-0.080	0.236	-0.166
	(0.034)	(0.056)	(0.034)	(0.064)	(0.074)	(0.084)
Number of Students	3883	3883	4033	4033	3883	4033

Table 4: continued

	Boy		Girl		Boy	Girl
	Direct-Subject Effect	Cross-Subject Effect	Direct-Subject Effect	Cross-Subject Effect	Average-Subject Effect	Average-Subject Effect
	(1)	(2)	(3)	(4)	(5)	(6)
C. Probability of Receiving a Matriculation Diploma						
OLS	0.005	0.110	-0.049	-0.012	0.115	-0.061
	(0.052)	(0.065)	(0.046)	(0.055)	(0.077)	(0.059)
6th Grade School Fixed Effects	0.049	0.039	-0.051	-0.066	0.088	-0.116
	(0.034)	(0.045)	(0.032)	(0.035)	(0.050)	(0.040)
6th Grade School Fixed Effects and Student Characteristics	0.040	0.031	-0.047	-0.036	0.071	-0.081
	(0.031)	(0.043)	(0.033)	(0.037)	(0.049)	(0.042)
Number of Students	1243	1243	1269	1269	1243	1269
D. Total Number of Successfully Completed Matriculation Exams' Units						
OLS	0.050	2.032	0.105	-0.860	2.081	-0.742
	(0.108)	(1.135)	(1.024)	(1.166)	(1.244)	(1.118)
6th Grade School Fixed Effects	0.565	1.539	0.255	-2.602	2.109	-2.455
	(0.717)	(0.721)	(0.647)	(0.708)	(0.871)	(0.873)
6th Grade School Fixed Effects and Student Characteristics	0.564	1.378	0.333	-1.954	1.945	-1.683
	(0.595)	(0.649)	-(0.673)	(0.715)	(0.792)	(0.851)
Number of Students	1243	1243	1269	1269	1243	1269

Notes: The table reports the estimates of teachers' biases on several educational outcomes: Panel A and B shows results of the estimated effect of teachers' biases on 8th grade GEMS test scores and on matriculation test scores respectively. In both panels test scores in all three subjects (math, English, and Hebrew) are pooled together. These test scores are standardized scores, by year and subject. Panel C shows results of the estimated effect of teachers' biases on the probability of receiving a matriculation diploma and Panel D shows results of the estimated effect of teachers' biases on the total number of successfully completed matriculation exams' units. The first specification is a simple OLS regression with subject and year fixed effects; The second specification includes also primary school fixed effects; The third specification includes additionally students' characteristics (gender, parental education, number of siblings, and dummies for four ethnicity groups). The direct-subject effect is the effect of teacher bias in a specific subject on the test scores in the same subject; the cross-subject effect is the impact of the average teacher bias in the other subjects on the test scores in the referred subject. Average-subject effect is the impact of the average teacher bias in all subjects. The estimates in each row in columns 1-2 are each from a joint regression and so are the estimates in columns 3-4. The estimates in each row in columns 5-6 are each from separate regressions. Standard errors are clustered by class and are reported in parentheses.

Table 5: Estimated Effect of Teachers' Biases on Test Scores, by Subject

	Boy		Girl		Boy	Girl
	Direct-Subject Effect	**Cross-Subject Effect**	**Direct-Subject Effect**	**Cross-Subject Effect**	**Average-Subject Effect**	**Average-Subject Effect**
	(1)	(2)	(3)	(4)	(5)	(6)
A. 8th Grade GEMS Test Scores						
Hebrew	0.038	0.282	-0.020	-0.207	0.325	-0.229
	0.125	0.182	0.180	0.160	0.177	0.179
Math	0.374	0.029	-0.135	-0.137	0.368	-0.266
	(0.142)	(0.145)	(0.143)	(0.170)	(0.193)	(0.178)
English	0.018	-0.089	-0.111	-0.468	-0.066	-0.580
	(0.112)	(0.162)	(0.137)	(0.215)	(0.159)	(0.244)
B. Matriculation Test Scores						
Hebrew	0.018	0.221	-0.113	-0.079	0.233	-0.194
	0.096	0.117	0.079	0.104	0.094	0.095
Math	0.148	0.119	-0.050	-0.145	0.263	-0.195
	(0.120)	(0.115)	(0.083)	(0.119)	(0.103)	(0.099)
English	0.109	0.102	-0.002	-0.101	0.209	-0.105
	(0.052)	(0.059)	(0.071)	(0.085)	(0.066)	(0.088)

Notes: See Table 4. Each row present estimates from separate regression for each subject. Each regression includes students' characteristics, elementary school, year and subject fixed effect. The estimates in each row in columns 1-2 are each from a joint regression and so are the estimates in columns 3-4. The estimates in each row in columns 5-6 are each from separate regressions. Standard errors are clustered by class and are reported in parentheses.

Table 6: Estimated Effect of Teachers' Biases on Students' Probability of Successfully Completing Advanced Courses in High School

	Boy		Girl		Boy	Girl
	Direct-Subject Effect	**Cross-Subject Effect**	**Direct-Subject Effect**	**Cross-Subject Effect**	**Average-Subject Effect**	**Average-Subject Effect**
	(1)	(2)	(3)	(4)	(5)	(6)
English (dummy=1 if # units=5|4)	0.025	0.033	-0.011	-0.031	0.058	-0.042
	(0.021)	(0.027)	(0.029)	(0.034)	(0.035)	(0.034)
Math (dummy=1 if # units=5|4)	0.093	0.020	-0.073	-0.009	0.110	-0.083
	(0.049)	(0.050)	(0.044)	(0.057)	(0.060)	(0.055)
Physics/Computer Science (dummy=1 if units=5)	0.020	0.001	0.020	-0.002	0.020	0.019
	(0.053)	(0.057)	(0.029)	(0.034)	(0.047)	(0.028)

Notes: See Table 5. Each row present estimates from separate linear probability regression for each subject (English /Math/Science oriented subjects). The dependent variables are discrete and equals one if the number of matriculation credit units exceed a certain level. Each regression includes students' characteristics, elementary school and year fixed effect. The estimates in each row in columns 1-2 are each from a joint regression and so are the estimates in columns 3-4. The estimates in each row in columns 5-6 are each from separate regressions. Standard errors are clustered by class and are reported in parentheses.

Table 7: Estimated Effect of Teachers' Biases on Students' Total Number of Successfully Completed Units in Science, Math and English Courses in High School

	Boy		Girl		Boy	Girl
	Direct-Subject Effect	Cross-Subject Effect	Direct-Subject Effect	Cross-Subject Effect	Average-Subject Effect	Average-Subject Effect
	(3)	(4)	(5)	(6)	(7)	(8)
English	0.156	0.103	0.045	-0.188	0.257	-0.149
	(0.087)	(0.096)	(0.108)	(0.132)	(0.116)	(0.137)
Math	0.214	0.130	-0.056	-0.236	0.338	-0.291
	(0.160)	(0.171)	(0.119)	(0.173)	(0.168)	(0.154)
Computer Science	-0.035	0.011	0.025	-0.088	-0.022	-0.062
	(0.164)	(0.171)	(0.109)	(0.141)	(0.190)	(0.108)
Physics	0.255	0.117	-0.036	0.088	0.365	0.050
	(0.241)	(0.255)	(0.108)	(0.108)	(0.219)	(0.115)
Sum of Number of Units in Math, Physics and Computer Science	0.435	0.258	-0.064	-0.242	0.681	-0.305
	(0.497)	(0.527)	(0.226)	(0.283)	(0.427)	(0.236)

Notes: See Table 5. Each row present estimates from separate OLS regression for each subject (English /Math/Science oriented subjects). The dependent variables in each row are continuous and equals to the total number of matriculation units students' gained in each of these study programs. Each regression includes students' characteristics, elementary school and year fixed effect. The estimates in each row in columns 3-4- are each from a joint regression and so are the estimates in columns 5-6. The estimates in each row in columns 7-8 are each from separate regressions. Standard errors are clustered by class and are reported in parentheses.

Table 8: Estimated Average-Subject Effect of Teachers' Biases on 8th Grade GEMS Test Scores, by Sub-Groups

	Boy	Girl	Boy	Girl
	(1)	(2)	(3)	(4)
	Low Parental Education		**High Parental Education**	
Mother's Education Level	0.200	-0.407	0.492	-0.186
	(0.190)	(0.212)	(0.193)	(0.158)
Number of Students	805	723	615	594
	Low Parental Education Gap		**High Parental Education Gap**	
Parental Education Gap (Father's Education Less Mothers' Education)	0.009	-0.051	0.298	-0.431
	(0.261)	(0.223)	(0.158)	(0.168)
Number of Students	424	403	996	914
	Ethnicity Asia/Africa		**Other Ethnic Groups**	
Ethnicity Asia/Africa (defined by grandparents' place of birth)	0.242	-0.556	0.289	-0.223
	(0.237)	(0.215)	(0.135)	(0.186)
Number of Students	502	495	918	822
	Firstborn Children		**Non-Firstborn Children**	
Birth Order Of Children	0.179	-0.189	0.313	-0.417
	(0.229)	(0.237)	(0.185)	(0.189)
Number of Students	527	496	893	819

Notes: The table presents the estimated average-subject effect of teachers' biases on 8th grade GEMS test scores**.** Each regression includes pupil's characteristics, 6th grade, year and subject fixed effect. High parental education is defined as more than 12 years of mothers' schooling. High parental education gap is defined as a non-negative gap between the fathers' and the mothers' years of schooling. The estimates in each row in columns 1-4 are each from separate regressions. Standard errors are clustered by class and are reported in parentheses.

Table 9: Correlation of Teachers' Biases Measure with Characteristics of Teachers

	Age Dummy (dummy=1 if Older than 50 Years Old)	Ethnicity Europe/America	Married	Single	Number of Teachers' Offspring	Proportion of Daughters among Teachers' Offspring	At Least one Daughter among Teachers' Offspring
	(1)	(2)	(3)	(4)	(7)	(5)	(6)
OLS	0.206	-0.204	0.032	0.315	-0.034	-0.047	-0.090
	(0.104)	(0.113)	(0.141)	(0.202)	(0.046)	(0.186)	(0.173)
Number of Teachers	114						

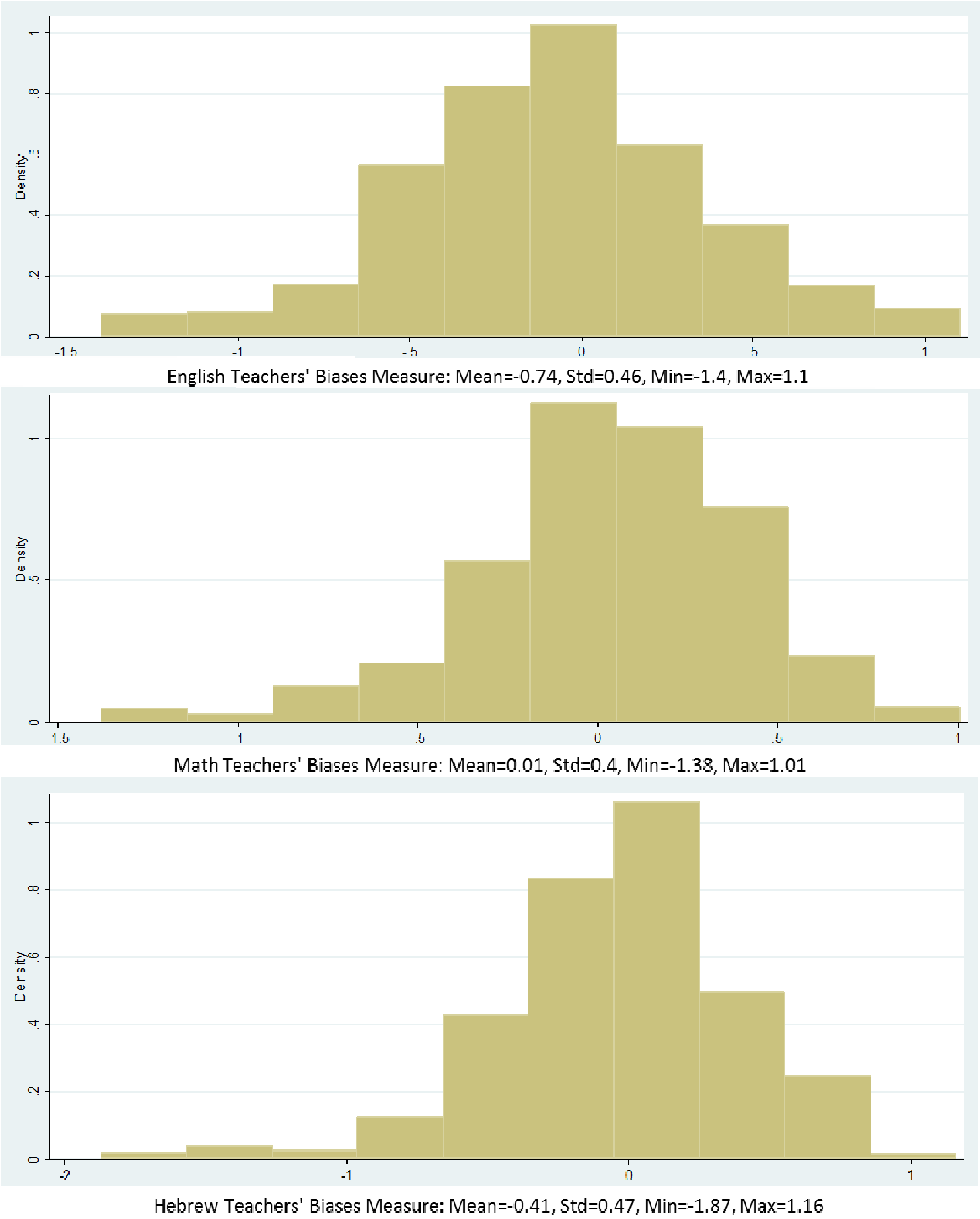
Notes: The table presents the estimated coefficient of teachers' characteristics on teachers' stereotypical bias measure. Each regression includes subject and year fixed effects. The estimates in each column in columns 1-6 are from a separated regression. Standard errors are clustered by class and are reported in parentheses.

Table 10: Estimated Effect of Average Teachers' Biases on 5th Grade Behavioral Outcomes

	Boy	Girl
	(1)	(2)
A. Teachers Encourage Students to Learn their Subjects of Interest	0.162	-0.333
	(0.233)	(0.190)
B. Insulting Attitude of Teachers Toward Students	-0.381	0.019
	(0.262)	(0.315)
C. Good Relations Between Teachers and Students	-0.062	-0.001
	(0.222)	(0.239)
D. Good Mutual Relations between Teachers and Students	-0.114	0.039
	(0.265)	(0.227)
E. Overall Satisfaction from School	-0.054	0.048
	(0.227)	(0.167)
Number of Students	1009	967

Notes: The table presents the estimated effect of the overall classroom stereotypical biased environment on each of these five behavioral outcomes. Each regression includes students' characteristics, school and year fixed effects. The estimates in each row in columns 1-2 are each from a separated regression, for boys and girls separately. Standard errors are clustered by class and are reported in parentheses.

Figure 1: The Distributions of Teachers' Biases Measure, by Subject



Notes: The teachers' biases measure is defined at the class level by the difference between boys' and girls' average gap between the school exams scores (non-blind) and the national exams scores (blind), by subject.