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

June 12, 2018

Victor Lavy
University of Warwick, Hebrew University and NBER

Edith Sand
Bank of Israel

Abstract

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 in Tel-Aviv, Israel. We measure bias using class-gender differences in scores between school exams graded by teachers and national exams graded blindly by external examiners. For identification, we rely on the random assignment of teachers and students to classes in primary schools. Our results suggest that assignment to a teacher with a greater bias in favor of girls (boys) has positive effects on girls' (boys') achievements. Such gender biases have also positive impact on girls' (boys') enrollment in advanced level math courses in high school. These results suggest that teachers' biased behavior at early stages of schooling has 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.

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 and Dr. Haim Gat and Eliad Trefler 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 data at its protected research lab. We benefitted from comments and suggestions from Naomi Hausman, Shulamit Kahn, Larry Katz, Kevin Lang, Yoram Mayshar, Jonah Rockof, Analia Schlosser, Moses Shayo, Sarit Weisburd, Assaf Zussman, two referees of this journal, participants in seminars and conferences at Hebrew University, Tel-Aviv University, Ben Gurion University, Paris School of Economics, University of Warwick, NBER 2015 Summer Institute Education conference, CEPR 2015 Public Economics Annual Symposium, LAGV 2015 Conference in Public Economics, COSME 2016 Gender Economics Workshop, the Barcelona 2015 Summer Forum and the 2016 Applied Family Economics conference in Honk Kong. The first author acknowledges financial support from the European Research Council through ERC Advance Grant 323439, from the Falk Institute and from the Israeli Science Foundation.

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 has decreased consistently, to the extent that it has even reversed in many countries – in some countries there have been more female than male graduates in recent years (Goldin et al. (2006), Becker et al. (2010) and Goldin 2014). This trend is partly due to more women graduating in what used to be male-dominated fields such as math, science and engineering. The math and science test score gender gap is of special interest because it is a good predictor of future income (Murnane et al. (1995) and Paglin and Rufolo (1990), Brown and Corcoran 1997) and because there is still a considerable gender gap in employment in these fields. Although evidence based on recent PISA testing[1] shows that gender gap in math is closing in many countries, there is still a large disparity at the upper tail of the test scores distribution (Ellison and Swanson (2010), Hyde et al. (2008), Machin and Pekkarinen (2008), Fryer and Levitt (2010)). In addition, striking evidence from the UK shows that in 2012 about 80% of those who took A level physics 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% of medical degrees and 79% of veterinary science degrees went to women.[3] The related employment gaps are even larger, as females are only 6% of the engineering workforce, 5.5% of engineering professionals and 27% of engineering and science technicians.[4]

What explains these gender disparities in cognitive performance and in math and science scores 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 and stereotypes influence the gender gap by shaping the way parents raise their children[6], by affecting

[1] Programme for International Student Assessment (PISA), which surveyed 15-year old students from OECD countries in 2003, 2006, 2009 and 2012.

[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. See also Friedman (1989) and Wilder and Powell (1989) for reviews of the literature.

[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.

[6] Different parental treatment and expectations are manifested in several ways, such as a 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 a 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).

the environment at school and teachers' attitudes, and by determining social and cultural norms.[7] There is limited credible evidence for this debate because it is difficult to disentangle the impact of biological gender dissimilarities from environmental conditions, and because it is difficult to measure stereotypes and prejudices and test their causal implications.

In this paper we focus on the effect of gender bias in a schooling environment. Stereotypical attitudes of teachers towards boys and girls in class have been widely documented in the psychology and sociology literature, and have been argued to substantially influence students' self-image and educational outcomes. 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 feedback. They receive more praise for intellectual quality of their ideas whereas girls receive less instructional time, fewer challenges and are reinforced for conformism and passivity (Sadker et al. 2009). Teachers are also found to treat boys and girls differently, in particular with regard to math instruction: Hyde and Jaffe (1998) show that math teachers tend to encourage boys to exert independence by not using algorithms and that boys who pursue this rebellious approach are seen as having a 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 courses in math.

Using all of these mechanisms through which gender biases of teachers potentially affect their students' educational outcomes, we build a quantitative measure of primary school teachers' gender biases and estimate the impact on boys' and girls' academic achievements during middle and high school, and on the selection of advanced level courses in math and sciences during high school. We measure teachers' gender biased behavior by comparing their average marking of boys' and girls' papers in a "non-blind" exam to the gender means in an anonymously marked "blind" national exam. We assume this measure of teachers' biases captures her/his overall perception about gender cognitive

[7] 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 the cultural environment. They show that there is a positive correlation between gender equality and the 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 (which was based on a measure developed by Spolaore. and Wacziarg, 2009), 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.

differences. We then use it as a proxy for teachers' behavior towards the different groups in class. This measure might embody a wide range of behaviors, from conscious discriminating behavior to unintentional teaching style and personal traits that makes learning easier for one group over another.[8] Students might be affected by either one of these behaviors, all influencing their self-perceptions and functioning in a way of a sort of a self-fulfilling prophecy. We show that there is a 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 the difficulty of the math and science courses chosen in high school. These high stakes choices determine whether a student will meet requirements for admission to science and math studies at university.

The construction of the teacher bias measure is based on two tests that differ in timing (by one year) and because of that also differ somewhat in the material they cover. We note however that between the dates of these two tests there are no changes in teaching quality nor in the social environment in class because the class has the same teacher in 5th and 6th grades and because the class composition is unchanged as class enrollment is the same. In addition, since non-blind tests are graded by the class teacher who presumably knows their students more intimately, the blind score might capture more accurately students' unobserved characteristics than the national test. Being aware that these differences between blind and non-blind tests might lead to different interpretation of our findings, we address these confounding factors in several ways. First, we show that there is meaningful variation in gender based biases by teachers within schools and often it is even the case that within school and same subject teachers will have opposite gender biases. Therefore, since teachers and students are randomly assigned to classes within a given school, the within school by subject variation in our treatment variable permits us to use a school by subject fixed effect estimation strategy to estimate the effect of teachers' biases on their students' future education outcomes. Note that this identification strategy helps our interpretation of our findings against a variety of alternatives that rely on gender specific differences in behavior and characteristics, even if they are subject specific. Second, the within class variation in teachers' bias enable us to use also class fixed effect estimation strategy.[9] The similarity in the estimates we obtain from these two alternative model specifications reduces the possibility that our measure of teachers' gender bias simply pick up random (small sample) variation in the unobserved cross-subject stable "quality" of boys versus girls in a particular class. Furthermore, adding class level differences in average students' ability by subject to this class fixed effect estimation strategy accounts for any other subject-specific variation in achievements in class. Third, we define an alternative jackknife measure of teacher bias for each

[8] Few papers discuss the impact of unintended biases: see for example Bertrand et al (2005) for a discussion of the concept of 'implicit discrimination', ways of measuring it and possible ways of limiting its prevalence. See also Dee (2005) and Gershenson et al (2015) who highlight the role of teachers' perceptions and expectations and examine how they are affected by student-teacher demographic mismatch.

[9] We note that the term "class" describes the group of students who share the same lessons and teachers in all subjects.

student based on excluding his own exams' gap between "non-blind" and "blind" exams scores. This jackknife version addresses the problem that might arise if the student's own exam scores' gap is mechanically correlated with his future educational outcomes because of teacher's evaluation ("non-blind" test score) reflecting more accurately his unobserved traits than standardized tests ("blind" test score). Since this alternative measure might contain a measurement error due to the omission of one student from the class mean, we present both estimates: the benchmark measure represents a higher bound for the effect of teacher biases whereas the jackknife measure represents a lower bound. Finally, we provide additional and more direct evidence that our estimates reflect teachers' behavior and not students' characteristics or behavior. We show that teachers' biases are correlated with their own characteristics. For example, we find that having a bias in favor of girls is positively correlated with the proportion of daughters among teachers' children. In addition, we show that when a teacher is teaching two subjects, the correlation between teacher's biases measured in each of these two subjects is significantly higher than the correlation when the two subjects are taught by two different teachers.

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).[10] This approach was first applied to the economics of education 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), Burgess and Greaves (2013), and Botelho, Madeira and Rangel (2015), who implemented the same methodology using data from other countries and getting overall similar evidence about teachers' stereotypes/biases.[11] In the present paper, however, we go beyond measuring teachers' biased behavior and focus on the implications of this behavior for gender differences in human capital formation. We think this paper is the first to examine the consequences of teachers' biased behavior on their students' gender gap in human capital formation, in particular regarding gender differences in math and science studies.[12] Few papers pursued recently this idea which we introduced first in earlier

[10] 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. See Bertrand and Duflo (forthcoming) for a recent survey of the literature on discrimination, which reviews the existing field experimentation literature on the prevalence of discrimination, the consequences of such discrimination and possible approaches to undermine it.

[11] Lavy (2008) finds that in high schools, male students are being discriminated against in all subjects. Based on evidence from primary school in the U.S, Cornwell et al. (2013) found that boys who preform equally well as girls were graded less favorably by their teachers and that this gap can be largely explain by these students' non-cognitive skills. Other papers using a similar methodology examine the existence of racial discrimination: Burgess and Greaves (2013) find that in English public schools, black Caribbean and black African students are under-assessed relative to their white peers while other minority groups (such as Indian, Chinese and Asian) are over-assessed. Botelho et al (2015) find that black students are being discriminated against relative to their white classmates in Brazilian schools. Björn et al. (2011) report a similar attitude towards students from foreign backgrounds in Swedish high schools.

[12] In a recent paper, 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 the disciplines' characteristics. This correlation is argued to be partly driven by the negative stereotype against women on this dimension, which is

draft (Lavy and Sand 2014). Consistently with our finding Terrier (2015) shows that in primary schools in France there also exists a positive correlation between teachers' bias in favor of boys in a specific subject and the progress of boys relative to girls in class in that subject. Lavy and Megalokonomou (2016) document similar teachers' biases in secondary schools in Greece and their impact on students' outcomes. In a related paper, Carlana (2017) finds a negative correlation between gender biases of math teachers based on Implicit Association Test scores and their female students' math test scores.

To this end, we focus on boys' and girls' choices about the difficulty of the math and science courses they select in high school. In Israeli higher education, as in many other countries, these choices have important implications for occupational choices at adulthood, because advanced courses in math and science in high school are 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, influence this group's future achievements and affect their orientation toward enrollment in advanced math and science studies in high school.

Our data enables us to evaluate the impact of teachers' gender biases on students' test scores in later years by following three cohorts of 6^{th} 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 teachers' gender biases in primary school, and to estimate the effect on both 8^{th} 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 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.

Our results suggest that teachers' more positive assessment of boys (girls) in primary school in a specific subject has a positive and significant effect on boys' (girls') achievements in that subject in middle school and high school national tests. We find that the magnitudes of these effects are more pronounced for boys than for girls, especially when using the jackknife version of the teacher bias measure. In addition, we find that the favoring of boys (girls) by primary school math teachers also affects boys' (girls') successful completion of advanced courses in math and science in high school. Teachers' biases that favor boys (girls) encourage boys (girls) to choose advanced math courses and to successfully complete more units in math and science courses. Since these courses are prerequisites for admission to higher education in these subjects, teachers' biases contribute to the gender gap in

---

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's arithmetic ability reduced the degree of discrimination, but did not eliminate it completely.

qualifications in fields like engineering and computer science, and therefore 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 in adulthood. In addition, we show that these effects have interesting patterns of heterogeneity by parental years of schooling and parental education gap.

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

2. Empirical Strategy

Teacher gender bias measure

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). We define two different measures of teacher bias. The first measure is based on all the students' test scores' differences in class (basic measure) and the second is based all students' test scores' differences in class excluding the student own test score's difference (jackknife measure). We assume that these measures of teachers' biases capture the teachers' overall perception about gender cognitive differences. We then use it as a proxy for teachers' behavior towards the different groups in class. These measures might capture a wide range of teachers' behavior, from conscious discriminating behavior to unintentional teaching style and personal traits that makes learning easier for one group over another. Students might be affected either by teachers' evaluations or by the way they are taught or treated by their teachers, both influencing their self-perceptions and functioning like a type of self-fulfilling prophecy. Although we cannot identify the exact mechanism through which students are affected, all these channels are possible explanations consistent with the results.

More formally, we denote the difference between the school score and the national score of student *i*, from primary school class *c*, subject *j* and year *t* as:

(1) $Gap_{icjt} = S\_Score_{icjt} - N\_Score_{icjt-1}$

where S_Score denotes the non-blind (school) score and N_Score denotes the blind (national) score. These gaps are averaged for boys and for girls and the difference is then a measure of teacher bias. So, for a given teacher in class c, subject j and year t, the measure of bias is defined by:

(2) $Bias_{cjt} = 1/K \sum_k Gap_{kcjt}^{boys} - 1/L \sum_l Gap_{lcjt}^{girls}$

where *K* is the total number of boys in the class and *L* is the total number of girls. This bias measure is calculated in each subject (Hebrew, math and English) for each one of the 112 classes in our sample. The higher it is, meaning the higher is boys' average gap in scores between "non-blind" exams graded by their teachers and "blind" exams relative to that of the girls in class, the higher the bias in favor of boys and against girls.

This approach of measuring a teacher's bias is based on all of his classroom students. We also consider the jackknife version of teacher bias measure, where for each student we exclude his own exams' gap between "non-blind" and "blind" scores from the class average measure. This jackknife version addresses the problem that might arise if the student's own exam scores' gap is mechanically correlated with his future educational outcomes, for example, if the teacher's evaluation (the "non-blind" score) might capture more accurately a student's unobserved traits than the blind test score. We note that this measure of the teacher's bias will include a measurement error because of the omission of one student from the class mean but the use of empirical Bayes shrinkage mitigates to some extend this loss of information by scaling up the estimates. We therefore present both estimates, the basic measure represents the higher bound for the effect of teacher biases whereas the jackknife measure represents a lower bound.

The jackknife version of teacher bias measure is defined as follows:

$$(3)\quad {Bias_{cjt}}^{-i} = \begin{cases} 1/(K-1)\sum_{k\neq i} {Gap_{kcjt}}^{boys} - 1/L\sum_{l} {Gap_{lcjt}}^{girls} & if\ i = boy \\ 1/K\sum_{k} {Gap_{kcjt}}^{boys} - 1/(L-1)\sum_{l\neq i} {Gap_{lcjt}}^{girls} & if\ i = girl \end{cases}$$

where the notation is as in equation (2). This measure assigns different values of bias for each student in a particular classroom using outcomes from all other students in the classroom taught by the same teacher.

Identifying Teachers' Biases Impact on Test Scores

The main goal of this paper is to investigate how teachers' biases towards a specific gender influence this group's future achievements and affect educational choices. Our data allows us to track students from primary school, where students were exposed to different teachers' gender biases, through middle and then high school. Thus we can examine the implications of this exposure for their human capital formation, in particular test scores in middle school and high school national standardized tests, and choices about math and science studies in their final years of high school. Our main identification strategy relies on the random assignment of students and teachers in a specific subject to classes within a school. Using within-school by subject analysis (primary school by subject fixed effect framework), we compare students that study in the same primary school but were randomly exposed to different teachers in a specific subject, and potentially different gender biased behavior. In robustness exercises, we first replace the school by subject fixed effects with class fixed effects to further test the impact of variation in teachers' biases on students within the same class, and then add class level differences in average students' ability by subject to this class fixed effect estimation strategy in order to account for any other subject-specific variation in achievements in class.

The randomness of class composition results from the fact that students' assignments into class based on ability, family background or any other characteristics of the students are forbidden by law in

Israel and this law is strictly enforced.[13] In order to test explicitly for the randomness of class composition in our sample, we perform 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, four ethnicity groups, number of siblings, and level of parents' education) we find that out of 296 p-values, only 18 were equal to or lower than 5 percent. This implies that for only 6% of the classes we cannot reject that there is non-random assignment. In addition, of the 37 elementary 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.[14] 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 observed characteristics, teachers' by subject assignments to class are also unrelated to unobserved students' backgrounds.

In the empirical model we assume that the test scores and the choice of advanced level courses by pupils in middle/high school are determined by the following equation:

$$(4)\ y_{icjt} = \alpha + \beta_{js} + \gamma_t + \lambda_1 y_{ijt-1} + \lambda_2 X_i + \beta_1 Bias_{cjt} + \beta_2 Bias_{cjt} * Boy_i + \varepsilon_{icjt}$$

where y_{icjt} denotes the outcome of student *i*, from primary school s and class c, subject *j* and year *t*; y_{ijt-1} is student i's 5th grade national exam test score in subject j; X_i are the student characteristics including the gender of the student, and Boy_i is a dummy for boy; β_{js} is a primary school by subject fixed effect; γ_t is a year fixed effect; $Bias_{cjt}$ is the basic measure of teachers' biased behavior in subject *j*. The error term, ε_{icjt}, is clustered by both student and primary school class.[15]

The coefficients of interest are β_1 and β_2, where β_1 captures the effect of teacher's biases in subject j on girls' later scores in subject j and the sum of the coefficients β_1 and β_2 captures the effect of teacher's biases in subject j on boys' later scores in subject j.[16] We also estimate the effects of the

[13] The 1968 Integration Law in Israel clearly states that schools should be the focal point of integration of different socioeconomic and ethnic groups in Israeli society. Therefore, tracking students in primary or middle schools based on students' characteristics is prohibited. Numerous publications of the Director General's Circulars at the Ministry of Education note that a specific committee at the Ministry is responsible for the implementation of the integration policy. This committee monitors periodically the integration process between and within schools. (See for example the Director General's Circular publication regarding the integration policy of Ethiopian students:
http://cms.education.gov.il/EducationCMS/applications/mankal/arc/sd9ak3_7_47.htm). See also the Bank of Israel Report No. 2014.07 which examined whether the allocation of students to classes by socio-demographic characteristics was random during the years 2001-2010 and found very little segregation within schools in Israel.

[14] See also Lavy (2011) and Lavy and Sand (2017) for evidence that suggests no systematic nonrandom formation of classrooms in primary and middle schools in Israel.

[15] Changing the standard errors level of the clusters to class by subject clusters (in addition to students' clustering) has only marginal effects on the significance levels of the results. Furthermore, since homeroom teachers often teach their class more than one subject the standard errors in the baseline model are clustered by class.

[16] We note that the specific model that we estimate, whether by splitting the sample or pooling both gender and interacting the bias measure with the gender of the student, does not influence substantially the estimated effect of the teacher bias.

jackknife version of teachers' biased behavior on boys and girls test scores which are determined by the following equation:

$$(5)\ y_{icjt} = \alpha + \beta_{js} + \gamma_t + \lambda_1 y_{ijt-1} + \lambda_2 X_i + \beta_1 Bias_{cjt}{}^{-i} + \beta_2 Bias_{cjt}{}^{-i} * Boy_i + \varepsilon_{icjt}$$

We present both effects of teacher bias measures on boys and girls test scores according to the basic teacher bias measure and according to the jackknife teacher bias measure. In addition, we implement a Bayes shrinkage estimation strategy for both definitions of teacher bias measures and construct unbiased measures of teacher bias that accounts for noise in the measurement. Using this approach the noisy measures of a teacher bias is multiplied by an estimate of their reliability, where the reliability of a noisy measure is the ratio of signal variance to signal variance plus noise variance. Thus, less reliable measures are shrunk back toward the mean of the distribution of teacher bias measure.[17] In addition, all measures are normalized to be mean zero and have a standard deviation of one.

2. Data

Data Description

In order to construct the teachers' bias measures (basic and jackknife) we combine two datasets and compare the students' scores from a "non-blind" exam and a "blind" exam.

The first dataset is from the school authority for the municipality of Tel-Aviv. The data contains information on 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 exams in three subjects (math, English and Hebrew) held in the midterm of 6th grade. These tests were graded by the students' teachers[18] ("non-blind" assessments) and were created and administered by Tel-Aviv municipality for monitoring purposes. 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).

The second dataset is GEMS records (Growth and Effectiveness Measures for Schools - *Meizav* in Hebrew) for the three cohorts that we study. The GEMS records were created and

[17] Following Morris (1983) and the teacher value added literature (for example, Kane and Staiger 2008) we construct the EB shrinkage factor for teacher i by the ratio of signal variance to signal variance plus noise variance of teacher i. Similarly to the teacher value added literature, we assume that the measure of teacher bias includes an error component. Thus, estimating teachers' effects on students' weighted difference between "non-blind" and "blind" scores (where the weights are the inverse proportion of each gender in class, defined positively for boys and negatively for girls) enables to separate between the signal variance (variance of teachers' effects) and noise variance of teacher i (variance of the residuals for teacher i). We note that since we observe only part of the teachers teaching multiple classes we do not distinguish between the student-level noise and the teacher-class shocks. The EB estimate for each teacher is a weighted average of the teacher estimated effect and the mean of teacher estimates, where the weight is the EB shrinkage factor. While implementing this methodology, the less reliable estimates of teacher bias (those with a large variation in estimated residuals) are shrunk towards the mean of teacher estimates.

[18] Students were tested in these three subjects only, and the tests in each subject (Hebrew, math and English) were graded by the class teacher for the subject. We note that all tests in each subject were the same and teachers did not have the ability to construct their own tests.

administered by the Division of Evaluation and Measurement of the Ministry of Education.[19] The students' GEMS records include test scores of fifth and eighth graders for a series of tests (in math, Hebrew and English) as well as an individual identifier, a school and class identifier at the 5 and 8 grades. The GEMS tests were administered during 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 once every two years. GEMS tests were graded blind by an independent agency: the identity of the student is never revealed. The GEMS tests in 5th grade serve to construct teacher bias measure, while 8th grade GEMS tests is being used to examine the short term impact of teachers' biases on their students' scores.

The 5th grade GEMS test is a "blind" assessment since the GEMS exams are graded by an independent agency and the identity and gender of the student are never revealed. In contrast, the other exam, which is graded by the students' teacher, contained the name of the student and therefore is a "non-blind" assessment. Thus, using these two tests enables us to define the measures of teacher bias (basic and jackknife) at the class level by the difference between boys' and girls' average gap between the school score (non-blind) and the national score (blind).

We note that in addition to the different ways these two tests are administered, they differ in some other aspects: First, the structure of these tests is different. While some of the questions in the GEMS are multiple choice, most of the "non-blind" tests questions are open responses. Although the way students answer these types of questions may differ across gender, the fact that the "non-blind" tests consist mostly of open questions give teachers more freedom in grading those tests and enables scores to reflect other possible factors besides students' knowledge. Second, since non-blind tests are graded by the class teacher who presumably knows their students more intimately, the blind score might capture more accurately students' unobserved characteristics than the national test. Third, the timing of these tests differs. Since GEMS test are administered at the mid-term of the 5th grade, only three or four months after the teachers have begun instructing the class, we posit that their biases in class only marginally affects students' GEMS test scores, while 6th grade test scores are influenced much more by the behavior of the teachers.[20] Furthermore, the fact that internal scores are revealed to students only after the GEMS test eliminates the possibility that GEMS scores are affected directly by grading of their teachers. Fourth, the material being evaluated in both tests might not completely overlap, although most of the topics covered should be comparable. This results from the fact that the time gap between these two tests is about a year though students' educational environment remains almost unchanged throughout these two consecutive years (teachers in both 5th and 6th grades are

[19] 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.

[20] If we were to assume that teachers' biases affected to some extent also 5th grade external scores, it would have biased our teacher measure towards zero (i.e., underestimating the magnitude of teachers' biases effects).

usually the same teachers[21], and students stay in the same classes and have the same curriculum). Finally, both tests are low stakes tests because they are not used for matters important directly to students and are mainly used for monitoring purposes.[22] In addition, since they are both created and administered by external agencies (the Division of Evaluation and Measurement of the Ministry of Education and Tel-Aviv municipality) all tests in each subject were the same and teachers did not have the ability to construct their own tests.

Although the timing of these tests differ as well as the material they cover, since the class has the same teacher and there is no change in class composition, students are experiencing no change in teaching quality or in their social environment. Thus, it increases the likelihood that nothing else occurs between these two tests that can affect student achievements differentially by gender, other than the tests being graded differently (blind vs. non-blind). Though we cannot completely rule out that the differences in what is being evaluated by these two exams are not gender-neutral, our results are not sensitive to these cross gender differences, since the teachers' bias measure, which relies on the average scoring of boys and girls in these exams at the class level, would have been affected in a similar way in all the classes of the sample.

In order to test teachers' biases effect on students' test scores, we examine the short term impact of teachers' biases on 8th grade GEMS tests (in math, Hebrew and English), and their long term impact on matriculation tests taken at the end of high school. Thus, in addition to 8th grade GEMS tests scores, we merge the data also with matriculation exam scores and credits units from the Israel Ministry of Education. Matriculation exams are national exams in core and elective subjects, taken between the tenth and twelfth grades. Students choose to be tested at various levels of proficiency, with each test awarding from one to five credits per subject, depending on difficulty. Some subjects are mandatory, and for many the most basic level is three credits. Advanced level subjects carry four or five credits. The matriculation exams in math, English and Hebrew are mandatory: the number of credits required in Hebrew is two, and in math and English students are allowed to choose between the most basic level (three credits) and the advanced level (four or five credits). On the other hand, matriculation exams in computer science and physics are optional and students can take a maximum of 5 credits in these subjects. A minimum of 20 credits is required for a matriculation certificate, which is a prerequisite for university admission. We focus on the following matriculation exam outcomes: test score in math, English and Hebrew, the probability of matriculating, the number of successfully completed exams, and the number of successfully completed units in English and in math related subjects (math, physics and computer science). We note that we weight the test scores based on

[21] Homeroom teachers in Israel elementary schools are generally assigned to the same classes for two years consecutively.

[22] The school tests were used by the Tel-Aviv school authority to monitor the composition of schools in terms of students' primary school achievements along similar comparison in terms of socio-economic background of students. The national tests are used mainly to give schools feedback on their average performance relative to other schools in the district and in the country. We note that only the school means of GEMS tests results are sent to schools.

number of credit units assigned to each test and the respective weights determine by the Higher Education Council (four credits are awarded a bonus of 12.5 points and five credits are awarded 25 points).

We use two additional datasets. The first dataset is teachers' GEMS questionnaire which contains data on homeroom teacher identifier and their class identifiers and main subjects of instruction. This information enables us to relate teacher bias measures to the relevant teachers, and check the consistency of the biases' measures of the same teacher who teaches the same class different subjects. The teachers' GEMS questionnaires were addressed to almost all teachers in schools for which we have students' GEMS scores in the relevant years (except from the first year of the sample for which we have only partial data because it was also the first year that the GEMS was administered).[23] Although all teachers were asked to fill in these questionnaires, we can only merge the information of homeroom teachers with our teachers' bias measure, since other teachers were not asked which classes they teach. Thus, the information we gather from these files are teachers' identifier, if they are homeroom teachers, and if so which class do they teach and what are their subjects of instruction. This information allows us to classify which homeroom teachers instruct both Hebrew and math courses and which ones instruct only one of these subjects. Since elementary school homeroom teachers teach math, Hebrew or both (Appendix Table A1 presents the subjects of instruction of identified teachers) it enables us to make distinguish between classes in which both subjects are taught by the same teacher and classes where these subjects were taught by different teachers.

The second dataset is the Population Registry at the National Insurance Institute (NII) that contains data on the demographic background of teachers.[24] This data enables us to observe homeroom teachers' demographic background (such as gender, age, marital status, ethnicity and number and gender of children) and tests whether high or low level of bias is correlated with certain attributes of teachers.

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, national exam GEMS test scores in 8th grade, matriculation exam scores and units of study at the end of high school for 2001–2008, 2002–2009 and 2003–2010, school and class assignments and student characteristics. In addition, we also observe teachers' characteristics for a sub-sample of teachers (homeroom teachers).

Summary statistics

The schools in the sample consist of elementary schools and middle/high schools in Tel-Aviv. Elementary schools are all K-6, whereas middle schools are part of secondary schools that include

[23] For 2002 only 13 homeroom teachers from 33 classes were identified: in 2003 and 2004 more than 30 are identified.
[24] We accessed this data at the protected research lab of the National Insurance Institute.

both the three middle schools grades and the three high school grades. Students attend their local elementary school and are placed in the same class from grade 1 to 6. There is very low mobility between schools and classes between the primary school and middle school years, so that class composition remains almost unchanged throughout these years. In contrast, the transition between elementary school to middle school was based on a school choice program where students' assignment to middle school depended on the preference of students and schools.[25] Moreover, this school choice program allowed students who completed primary school a choice of a middle school which was met by school capacity limits and balancing requirements.[26] Most of the student continues in the same school during all secondary school years (the proportion of students that changes school is about 10 percent). Since all the schools in our sample offer studies in an academic track leading to a matriculation diploma, most of the students in our sample were enrolled in the matriculation study program. The dropout rate during secondary schooling in our sample is about 10%.

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 40 secular elementary schools and 12 secular middle/high schools. The number of middle school students in the sample in each year is slightly smaller than half of the total number of elementary schools students in the sample because each school participated in GEMS tests once every two years (implying that only about a fourth of each cohort has test scores from both testing rounds)[27]. The fact that all secondary schools in our sample offer study programs in an academic track towards a matriculation diploma enable us to track most of these students until the end of high school. Therefore the student's panel data includes 5^{th} and 8^{th} grades GEMS test scores as well as a matriculation test scores. There are on average two classes in each elementary school (3 schools have only one classroom). The sample in elementary school 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.

Students in primary schools in Israel are randomly assigned into classes as any form of tracking is forbidden by law and this strictly enforced. Each student in primary school usually studies several different subjects, with the same group of peers. Homeroom teachers are the instructors of a

[25] The Tel Aviv school choice program allowed choice of secondary school at end of primary school. Each student could choose from a set of five schools, three of which were outside his/her school district. The school choice program opened the possibility for a better match between students and schools, and the system had the potential to increase school productivity by introducing competition among schools (see Lavy 2010 and Lavy 2016 for more details of this program and its medium and long term consequences).

[26] Tel-Aviv school authority maintained a balanced enrollment of students across schools based on socioeconomic level, educational achievement, gender, and disciplinary record.

[27] The proportion of students that take the GEMS exams is above 90 percent. We note that the bias measure is not correlated with the gender difference in this rate: the estimated correlations between the probability of missing test scores for girls (boys) and the bias measure are not statistically significant for all tests (5^{th} grade GEMS test, 8^{th} grade GEMS test and matriculation test).

quarter to a third of these subjects, while other subjects are taught by subject-specific teachers, such as English. Homeroom teachers in primary school generally teach the same class during two consecutive grades, while subject-specific teachers might teach several different classes.[28]

Appendix Table A1 presents descriptive statistics for the sub-sample of teachers of homeroom teachers for whom we have additional demographic information. 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 are homeroom teachers. We note that all identified teachers in our sample are female and this is expected given that the majority of teachers in primary school in Israel are female.[29] We also note that only a few schools appear in the sample twice and therefore we could not track homeroom teachers over time. [30]

Table 2 presents the means of the "non-blind" and "blind" test scores, and the mean of the difference between them, separately for boys and girls.[31] We also present in column 7 the difference between column 3 (the difference between boys' "non-blind" and "blind" exam scores) and column 6 (the difference between girls' "non-blind" and "blind" exam scores). The gender gap in test scores varies substantially by type of exam ("non-blind" versus "blind") and by subject. Girls in primary schools outscore boys in the Hebrew "non-blind" and "blind" exams. In math we see a different pattern—girls outscore boys in the "blind" exam and boys outscore girls in the "non-blind" exam. In English girls outscore boys in both types of exam.

Next we examine whether the apparent gap between "non-blind" and "blind" test scores of boys relative to girls (column 7) is statistically significant, using the estimation framework suggested 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 two specifications. 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" and "blind" scores of boys relative to that of girls (similar to the measure presented in the last column of Table 2), is positive in math, it is negative in English, and it is practically zero in Hebrew. While the estimates in Hebrew and English are not statistically different from zero in both specifications, the positive estimate in math is statistically different from zero in the first regression (OLS), and positive but not significantly different from zero in the second (student fixed effect specification).

[28] See the Director General's Circular regarding the syllabus regulation of primary schools in Israel: (http://cms.education.gov.il/EducationCMS/Applications/Mankal/EtsMedorim/3/3-1/HoraotKeva/K-2006-3a-3-1-25.htm).

[29] According to a recent publication of the Ministry of Education, 92 percent of primary school teachers in Israel (in secular Jewish schools) in 2007/8 were female. This statistic is from: http://meyda.education.gov.il/files/MinhalCalcala/facts-and-figures_v2_2014.pdf.

[30] Although our sample contains three consecutive years, we note that the first year of our sample contains only a small sample of teachers that participated in the GEMS (13 identified homeroom teachers).

[31] All test scores are standardized scores, by year and subject.

Although "non-blind" and "blind" tests differ in some aspects, we think these results imply that math teachers in our sample are to some extent more in favor of boys than Hebrew and English teachers, and that since boys tend to perform better in math and science subjects courses (see the discussion in the introduction) it suggests that math teachers might have stereotypical biases against girls.[32]

Table 3 presents the means of both middle school and high school test scores in the external exams, separately for boys and girls.[33] In all three subjects the gaps between girls' and boys' scores decline from middle to high school: the gender gap in Hebrew is 0.3 standard deviation in middle school and it declines to 0.15 standard deviation in high school. The gender gap in English is 0.16 in middle school and 0.02 in high school. The gender gap in math in middle school is -0.024 and in high school it is -0.09.

Appendix Table A3 presents the distribution of students by matriculation exam units of study, for boys and girls separately. Although girls have a higher probability of receiving a matriculation diploma and outnumber boys in the number of completed matriculation exam units, boys outnumber girls in math and in science oriented advanced courses. The proportion of boys and girls who successfully completed the advanced 5 credits course in English is almost the same and is around 56% for both genders. The proportion of boys who successfully completed the 5 credits course in math is 23.9%, while the respective proportion of girls is 12.8%. In science courses this gender gap is even larger: 16.5% of boys successfully completed advanced physics and 11% advanced computer science, while the rates for girls are only 4.5% and 3.5%, respectively. In chemistry the completion rates are even. In the remaining part of the paper we will test whether these differences in achievements, especially in math scores, and in successful completion of advanced math and science courses, can partly be explained by exposure to teachers' gender biases during earlier stages of schooling.

The distributions of the Bayesian measure for each subject are presented in Figure 1. These measures are normalized to be mean zero and a unit standard deviation. There is a large variation in the bias measures in all subjects (the ranges of the bias measures are: min=-2.85 and max=2.58 in English; min=-3.65 and max=2.65 in Hebrew; and min=-4.02 and max=2.65 in math). Actually, 66% of this variation is within school by subject. In the next section we will exploit this significant variation to test whether teachers' biases have short and long term effects on students' test scores.

Does the Bias Capture Teachers' Behavior?

In this sub-section we provide direct evidence that our gender bias measure captures teachers' and not students' behavior. We first examine the within classroom correlation between the bias measure in math and Hebrew when these two subjects are taught by the same teacher and compare it

[32] We note that these results differ from Lavy (2008) who finds that in a sample that includes all high school students in Israel, male students face discrimination when "blind" and "non-blind" scores of matriculation examinations are compared.

[33] All test scores are standardized scores, by year and subject.

to the within classroom correlation when the two subjects are taught by two different teachers. We exclude English teachers from this analysis because they do not teach math or Hebrew. The majority of teachers in this sample are homeroom teachers who teach the class several subjects, including math and Hebrew. Since we identify the classes and subjects of instruction of those homeroom teachers, we are able to divide our sample into homeroom teachers who teach their classroom both math and Hebrew, and those who teach only one of these subjects. If our gender bias measure indeed captures teachers' and not students' behavior or classroom characteristics, we expect the correlation between the math and Hebrew bias measures of the same teacher to be higher relative to the case where there are two different teachers for these subjects.[34] We also examine the correlations between teachers in different subjects from the overall sample of teachers (without restricting it to homeroom teachers). In this analysis we also expect the correlation between math or Hebrew teachers' bias and the English teachers' bias to be lower than the correlation between math and Hebrew teachers' biases, because English teachers do not teach math or Hebrew, while math and Hebrew are often taught by the same teacher.

Table 4 presents the correlations between Bayesian biases of teachers by subjects of instruction: The (OLS) estimates in column 1 are based on a sub-sample of classes where the same teacher teaches both math and Hebrew while the (OLS) estimates in column 2 are based on the sample of classes where math and English are taught by two different teachers. The estimates in each row in columns 3-4 are the correlation coefficients between bias measures using the sample of all teachers (same or different teachers for the two subjects), estimated in separate OLS regressions. The estimated coefficients in each row in columns 5-6 are from regressions that include primary school fixed effects.

Comparing the correlation coefficient estimate that is based on a sample of classes where the same teacher teaches math and Hebrew reveals that the estimate is positive, large and statistically significant (0.486, SE=0.135). In contrast, the respective estimates based on a sample of classes with different teachers of math and Hebrew is much smaller and not significantly different from zero (0.214, SE=0.182). In addition, the estimates in columns 1-4 reveal that the correlation between the math and the Hebrew teachers' bias measures is higher than the correlation between the bias shown by English teachers and the teachers of the other two subjects. Furthermore, once we add as controls primary school fixed effects, the correlation between the math and the Hebrew biases is positive and statistically significant, whereas the correlations between math/Hebrew bias and the English bias are both not significantly different from zero. Since most math teachers instruct Hebrew as well, and no English teachers instruct the other two subjects (Hebrew/math), these findings reinforce our interpretation that the teachers' bias measures do not capture students' or classes' behavior.

We find further evidence linking the bias measure to teachers' behavior by relating it to teachers' characteristics. We propose that if the bias measures captured students' and not teachers'

[34] We note that an even better strategy would have been to examine the correlation between bias measures of the same teacher teaching different classes. However, our data does not have such panel structure.

behavior, they should not be correlated with any of the teacher's personal characteristics. We find the opposite, however. Using administrative data from NII we are able to examine the characteristics of a sub-sample of homeroom teachers. In Table 5 we present the estimated correlations between several teachers' characteristics and the Bayesian teachers' bias measure. The estimates are from separate regressions for each of the teachers' characteristics that we have, using a school fixed effects regression with year and subject fixed effects. Teachers' characteristics include age, ethnicity, marital status and number of children and their gender. We note that all the identified teachers in the sub-sample are female, thus we could not test this aspect in our analysis.[35]

The proportion of daughters among teachers' offspring has a statistically significant pro-girls bias: the estimated effect is larger the lower is the proportion of daughters (-1.013, SE=0.449). The estimated effect of having at least one daughter is also negative, but not significantly different from zero (-0.556, SE=0.517). Psychologists and recently also economists have shown that parenting daughters increases feminist sympathies. For example, Washington (2008) has demonstrated that the propensity to vote liberally among legislator fathers, especially on reproductive rights, increases significantly with their proportion of daughters. The intuition here is that personal experience within the family can influence parental behavior. Consistently with the literature, we also find a correlation between offspring gender and parental beliefs. Moreover, while most of this literature focus on fathers, our sample includes mainly female teachers. Therefore our results suggest that among mothers as well, beliefs and gender bias are similarly influenced by their offspring's' gender.

In addition, we find that pro-girls bias is also larger among single teachers (-0,465, SE= 0.262). Teachers with a Asian/African origin (relative to other teachers of an European/North American origin or Israeli ethnic origin) has a bias in favor of boys (0.556, SE=0.329). The effects of the other three teacher's characteristics that we examine are not precisely measured: being married, the number of children and older teachers are all positively correlated with the bias measure but all of the effects are not significantly different from zero. Although these findings suggest that teachers' bias is correlated with characteristics that are not randomly assigned to teachers, they support our claim that the bias measure captures teachers' behavior and that teacher-specific component is a dominant factor.[36] It is difficult to provide reasonable explanations that link students' behavior to this pattern of correlations between teachers' bias measures and demographic characteristics.

[35] Although the issue of the correlation of teachers' gender with the measure of teachers' stereotypical biases is irrelevant in our context since all identified teachers in our sample are women (as it is also in many developed countries), the literature has documented different patterns of discriminatory behavior across gender. Carrell et al (2010) focus on the role of professor gender in affecting female students' math and science performances. Dee (2005) presents evidence that gender and race matches between students and teachers influence the teacher's subjective evaluations of student. Fershtman and Gneezy (2001) find a lower level of discriminatory behavior among females towards minority groups, while Reuben et al. (2014) report that both males and females tend to discriminate among job candidates based on their gender in a similar way.

[36] The R-squared in these regressions ranges between 0.4 and 0.48.

4. Results: Effect of Teachers' Biases

The estimated effect of Bayesian teachers' gender bias on students' academic achievements, based on estimating equations 1 and 2, is shown in Table 6. We present the estimates of the effect of the Bayesian measure of teacher bias using the basic teacher bias measure (column 1-2) and secondly using the jackknife teacher bias measure (column 3-4), from two separate regression specifications. Each regression includes a dummy for boy, the teacher bias measure and their interaction, subject and year fixed effects and student's 5th grade test score. The estimated effects for boys are presented in columns 1 and 3, and the estimated effects for girls are presented in columns 2 and 4. Panels A and B show 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 stacked (i.e. each student appear three times for each of these subjects). Panel C reports the estimated effect of teachers' biases on both 8th grade test scores and matriculation test scores, where the scores in all three subjects and in all tests (8th grade test scores and matriculation test scores) are stacked and a dummy variable for type of test (GEMS or matriculation tests) is added to the regression. All test scores are standardized scores, by year and subject.

Short term effects

In Panel A, Table 6, we report results from three different specifications. The simple OLS estimates (first row) are positive and significantly different from zero for boys (β_1) for the two bias measures, the basic and the jackknife measures (columns 1 and 3 respectively); for girls ($\beta_{1+}\ \beta_2$), the estimates are not statistically significant in both cases. Adding primary school by subject fixed effects to the regressions (second row) reduces boys and girls estimated standard errors. Adding students' characteristics leaves the estimates for boys and for girls almost unchanged, implying that pupil's characteristics are not correlated with the teacher's bias measures once we control for primary school by subject and the student 5th grade test scores.

The estimated effect of Bayesian teacher bias measures on boys' (girls') outcomes is positive (negative) —this indicates that teachers with more positive assessment of boys' (girls') test scores improve their achievements at a later age. The estimate of the basic teacher bias measure is positive and statistically significant for boys, 0.094 (SE=0.021) and is negative and marginally significant for girls -0.038 (SE=0.025, p value=0.128). In the jackknife specification both estimated effects are smaller, the estimate for boys is also positive and statistically significant, 0.054 (SE=0.022), while the estimate for girls is practically zero, 0.005 (SE=0.026).

Long term effects

In Panel B of Table 6, we present evidence of the effects of the Bayesian teacher bias measures on test scores in the high school matriculation exams in the three subjects. These exams are taken at the end of 12th grade, more than 6 years after 'exposure' to teachers' gender biases in primary

school. Similar to the pattern we found in Panel A, the estimates for boys are positive and significant, while for girls they are not statistically significant. The within school by subject estimation reduces again the estimated standard errors, which makes most of the estimates statistically significant. Both estimates for boys are smaller but still positive and significant, while the estimates for girls are both negative but only the estimated effect of the basic teacher bias measure is statistically different from zero. Adding student characteristics as additional controls in the regressions again leaves the estimated effect almost unchanged.

Comparing the estimates based on the third specification in Panel B to those in Panel A of Table 6 reveals that the effects of teacher bias measures persist to a large extent through high school. The estimated effects obtained from the basic and the jackknife measure of teachers' biases measures are very similar in both panels. The estimated effect of the basic measure on matriculation scores for boys (0.091, SE=0.025) is very similar to its estimated effect on 8th grade test scores (0.094, SE=0.021), and so are the estimated effect for girls, though the effect is significant on matriculation scores (-0.064, SE=0.023) and only marginally significant on 8th grade test scores (-0.038, SE=0.025). This pattern is reproduced when using the jackknife measure, where the estimated effects on both matriculation tests scores and 8th grade tests scores are positive, significant and of the same magnitude for boys (0.052, SE=0.025 compared to 0.054, SE=0.022) and not statistically different from zero for girls (-0.026, SE=0.024 compared to 0.005, SE=0.026).

In Panel C we take advantage of the estimates in Panels A and B being similar and report estimates based on pooling the middle school GEMS test scores and the high school matriculation scores data. In this pooled short- and longer-term outcome analysis, we use the third regression specification, which includes school by subject fixed effects, students 5th grade test scores and students characteristics as controls. The estimates from this regression are approximately the average estimated short-term effects (Panel A) and longer-term effects (Panel B).[37] We note that in all these panels there are large differences between the basic and jackknife measures' estimated effects, which are especially pronounced for girls. These differences result from the fact that the basic measure includes the student's own exams' gap between "non-blind" and "blind" exams scores (higher bound for teacher biases' effects) while the jackknife measure doesn't (lower bound). Thus, the first might be biased because of the correlation between student's own exam scores' gap and his future educational outcomes, while the latter might contain a measurement error due to the omission of one student. Considering the range between them as capturing the possible estimated teacher biases effects, we calibrate the effect size of increasing these teacher bias measures in a specific subject from zero (no

[37] Appendix Figure A1 presents the partial regression plots of teacher bias measure in each subject on pooled 8th grade GEMS and 12th grade matriculation test scores for boys and girls in each subject. Each regression includes a dummy for boy, the teacher bias measure and their interaction, and school, subject and year fixed effects. Appendix Figure A2 presents the regression plots of teacher bias measures on predicted test scores for boys and girls, where predicted test scores are based on 5th grade tests scores and all other observables. The correlations between predicted scores and both bias measures (the basic and the jackknife) are not statistically significant.

gender bias) to one (biases measures are standardized). In such a simulation, boys' test scores rise by 0.092 (/0.052) of a standard deviation based on the basic (/jackknife) bias measure, and girls' test scores decline by 0.051 of a standard deviation based on the basic bias measure and are not effected by teacher biases based on the jackknife version of the bias measure.[38]

We preformed several sensitivity tests of our main specification results (Table 6 Panel C) that we present in the online appendix. These sensitivity tests yield results that are very similar to those of our preferred specification presented in Table 6 Panel C. 1) In Appendix Table A5 we replicated the analysis while imposing two sample restrictions: the first is excluding from the sample schools with only one class per grade because they do not use random assignment of students to classes, Secondly, we test the sensitivity of the results by excluding outliers (extreme cases) in the teacher bias measures by restricting these measures to be in the interval [2,-2]; 2) In Appendix Table A6 we preform two additional robustness tests. We first test the sensitivity of our results to the inclusion of 5th grade students test scores. The fact that we control for student's characteristics as well as 5th grade test scores in our baseline specification, might be a concern if it creates a mechanical correlation between the bias measure and the student's 5th grade test scores. Therefore in the first row of Appendix Table A6 we test the results of a specification similar to our baseline specification but without controlling for student's 5th grade test scores. We note that the estimates of the basic version of the bias measure are lower that our preferred specification estimates, but the jackknife version estimates are less sensitive to the exclusion of 5th grade students test scores. We next test the possible influence of some type of tracking in middle/high school, by adding to the basic specification high school by subject fixed effects; 3) In Appendix Table A7 we estimated the main specification using a two-step bootstrapping algorithm in order to account for the estimation of teachers' biases as a first step and adjust their estimated standard errors[39]; 4) The last robustness checks presented in Appendix Table A8 examine the sensitivity of our results to the Bayesian estimation strategy, by comparing the result of our main specification to a comparable analysis without implementing a Bayes shrinkage estimation strategy and test for possible non-linearity in the effect of our bias measures by adding the square of the

[38] In Appendix Table A4 we replaced the test z-score with student's percentile rank in each subject. We compute percentile ranking by subject, type of test (GEMS test or matriculation test) and year. These results are presented in Appendix Table A4. The estimates in this table are consistent with those presented in Table 6 Panel C.

[39] We employ a two-step bootstrapping algorithm similar to the one computed in Ashraf and Galor (2013). The bootstrap estimates of the standard errors are constructed as follows. In a first step, a random sample with replacement is drawn from each class by gender group of students. A new measurement of teacher bias for each teacher is created, based on the new sample of students in class. In a second step, the effect of these new bias measures on student test scores in 8th and 12th grades based on the bootstrapped sample are estimated (based on the preferred specification presented in Table 6 Panel C) and the coefficients are stored. This process of two-step cluster bootstrap sampling and estimation, which also deals with school-student correlations, is repeated 1,000 times. The standard deviations in the sample of 1,000 observations of coefficients estimates from the second step are the cluster bootstrap standard errors of the estimated effects of teacher biases. The bootstrap estimates of the standard errors which are presented in Appendix Table A7 are very similar to the ones presented in our main specification, Table 6 Panel C.

measure of the grading bias to the baseline specification. We note that the fact that the results of our preferred specification are very similar to the modified ones reinforce the stability of our results.

Effects on Other Long Term Outcomes

In Table 7, we present evidence of the effects of Bayesian teachers' gender biases on two additional long term outcomes: Panel A of Table 7 reports the Bayesian estimated effects on the probability of receiving a matriculation diploma based on logit regressions; and Panel B of Table 7 reports the Bayesian estimated effects on the total number of successfully completed matriculation unit exams. The three specifications presented for each of these two dependent variables are similar to the specifications in Table 6 Panel B.

In the last row of panel A the estimated effects based on each of the two measures of teachers' biases (the logit estimates are transformed to marginal effects at the means) on the probability of receiving a matriculation diploma are positive and significant for boys. In the case of the basic bias measure the estimated effect is 0.035 (SE=0.010) and based on the jackknife bias measure the estimated effect is 0.022 (SE=0.010). The estimated effect on girls is negative and significant based on the two different bias measures, -0.037 (SE=0.011) and -0.020 (SE=0.010), respectively. The estimated effects on the number of successfully completed matriculation exam units, presented in the last row of Panel B, have a very similar pattern: the two respective estimates are positive and significant for boys (1.095, SE=0.338 and 0.654, SE=0.341) and they are negative and significant or marginally significant for girls (-0.821, SE=0.264, -0.338, SE=0.258). These results suggest that biases in favor of boys (girls) that students are exposed to in primary school increases boys' (girls') probability of receiving a matriculation diploma and their total number of successfully completed matriculation exam units, the latter being viewed as a good proxy for quality of the study program. We also note that these two outcomes feature prominently in the admission criteria of students to universities, in particular to highly demanded fields of study, and therefore they can have far reaching implications for students' careers. We discuss these implications in more detail at the end of this Section.

Does the Bias Measure Capture Variation in Student Characteristics Such as Ability or Non-Cognitive Skills?

It can be argued that teachers may take into account in determining the "non-blind" scores factors other than student's actual performance in the tests, in particular factors which are not necessarily related to teachers' gender biases. For example, teachers may know the true ability of students better than their external assessments might reveal, or teachers may 'award' bonus marks to students with good behavior, to those who are popular among peers or to those who make more effort in school. If so, the question that remains to be answered is whether the reason for these grading

patterns is gender based or not (for example, these grading patterns might result from systematic gender differences in popularity level or good behavior).

We argue that such 'threats' are unlikely to bias the estimates that we obtain based on the specification that includes school by subject fixed effects. For such a concern to be valid, it must be that gender differences of this nature vary across classes and within subjects in school. Otherwise they will be controlled for in the regression that includes school by subject fixed effect where we rely solely on within school by subject variation across teachers. For example, for this argument to be valid, it must be that in one class in a school the girls have higher ability in a specific subject, are better behaved or they make more effort and that their teacher rewards these attributes in terms of higher non-blind scores, while in the other class in the same school and in the same subject the boys have higher ability or are better behaved or they make more effort and that their teacher reward them for these attributes by higher non-blind scores. So the argument against the school by subject fixed effects estimates cannot rely on gender specific behavioral differences. On the contrary, they should vary in a convoluted way across gender, classes and subjects within a school in order to be consistent with our results.

The same kind of rationale holds against the validity of other such alternative interpretations. For example, suppose that the math curriculum in 6^{th} grade includes new material, for example geometry, and that girls do not do as well as boys in geometry questions in the internal assessments in 6^{th} grade. If that is the source of the gender marking bias in math, then it must lead to the same marking bias for all math teachers in Tel Aviv because the teaching curriculum is identical in all primary schools in the city. But we find that some of the teachers are in favor of boys and some are in favor of girls and that this variation holds within school. So this alternative interpretation seems irrelevant with regard to the estimates based on the school by subject fixed effect specification. Another example is the claim that girls do worse under pressure of external exams, or that girls do worse under the stereotypical threat environment of internal assessments, girls are more prone to peer pressure that leads to underperformance when test scores are not anonymous[40], and so on with other explanations that are based on gender specific characteristics or behavior.

Since alternative interpretations cannot rely on gender behavioral differences in general, it might be argued that teacher bias measures reflect random variation in boys' versus girls' cognitive or behavioral outcomes in class. We present several pieces of empirical evidence in an attempt to rule out such alternative interpretations.

We first test the sensitivity of our result to adding class mean difference between the means of boys' and girls' 5^{th} grade GEMS external exam test scores in each subject. Furthermore, we also add to this specification class fixed effects instead of school by subject fixed effects. Controlling for the average difference between boys and girls in 5^{th} grade GEMS scores by subject accounts for any

[40] Burnsztyn and Jensen (2015) find that when academic effort is observed by peers, students may conform to the prevailing norms in the peer group.

subject specific variation in achievements, whereas the within class variation in teachers' bias permits additionally to dismiss the possibility that our measure of gender bias of teachers just pick up random (small sample) variation in the unobserved cross-subject stable "quality" of boys vs. girls in a particular single class.

Table 8 presents the estimated effect of Bayesian teachers' biases on test scores when classroom level controls are added to the regressions. The table reports the estimated effect of teachers' biases based on a specification similar to our preferred specification (Table 6 Panel C). In the first row we report results from a regression that include as an additional control the lass level difference in 5th grade GEMS test scores between boys' and girls'. In the second row we present results when we also include class fixed effects and drop the school by subject fixed effects. Adding the classrooms' level control to the regression leads only to minor changes in the estimates. The estimated effect on boys of teachers biases using the basic or the jackknife bias measures are both positive and significant, 0.094, SE=0.021 and 0.055, SE=0.021, respectively, and they are negative and statistically significant for girls only when using the basic teachers' biases measure (-0.048, SE=0.02). Applying a similar robustness check while including class fixed effect instead of school by subject fixed effects, allows also to compare students in the same class that were randomly exposed to teachers of different subjects who have different biases.[41] Adding these two controls eliminates within classroom gender differences in average ability, by subject, as well as all stable (across-subject) class characteristics. The point estimates of the effect stay relatively stable in comparison to basic specification: using the basic bias measures, the estimates are 0.076, SE=0.023 for boys and -0.059, SE=0.026 for girls. Using the jackknife bias measure the estimates are 0.036, SE=0.025, p-value=1.43 for boys and -0.019 SE=0.027 for girls.

In addition, we present two falsification tests. As a complementary analysis to Table 4 where we compare the correlations between biases measures of the same teacher and those of different teachers, we report the results of the first falsification test which compares between the estimated effects of teacher biases on students test scores when teacher biases measures are switched for the same teacher versus for different teachers. The second falsification test assesses the possibility that our results are driven by the variation in boys' versus girls' cognitive ability in class. Assuming that teachers' behavior is driven by students' ability which is unrelated to their gender, we replicate our analysis based on the alternative measure of teacher bias that reflects teacher's attitude toward low/high achievers in class.

Appendix Table A9 reports the results of the first falsification test. This falsification test is based on switching the bias measures once when the same teacher is teaching both subjects and once

[41] We note that the variation in the bias measures using the class fixed effects estimation strategy is lower (46% of the variation is within classes relative to 66% in the case of the school by subject estimation strategy) partly due to the fact that the correlation between the bias measures of the same teacher who teach the same classroom both math and Hebrew is relatively high.

when two different teachers are teaching a subject. We divided the sample to two sub-samples of schools. The first sub-sample includes schools in which at least in one class the same teachers instruct both math and Hebrew. The results based on this sample are presented in columns 1-4. The second sub sample includes the schools where different teachers instruct math and Hebrew.[42] The results based on this sample are presented in columns 5-8. We use here a regression specification that includes school fixed effects, year fixed effects, a dummy for type of exam (GEMS exam or matriculation exam), students 5th grade test scores and students characteristics as controls (as in Table 6 Panel C). Panel A presents the estimated effect of the basic measure of teacher bias, whereas Panel B presents the jackknife approach.

In the first sub-sample, we switched between the math and Hebrew biases of the same teacher (i.e. teacher that instruct both math and Hebrew). The estimated Bayesian effects of teachers' bias for the "switched biases" case (columns 1-2) are almost identical to the estimated effect obtained from the "regular" sample where teachers biases are not switched (columns 3-4). The estimated effects of the "switched biases" are 0.087 (SE=0.027) and 0.048 (SE=0.025) for boys based on the basic bias measure and -0.050 (SE=0.025) and -0.025 (SE=0.028), when based on jackknife bias measure. These estimates should be compared to the respective estimates when we do not 'switch' across subjects: 0.093 (SE=0.027) and 0.051 (SE=0.024) for boys and -0.067(SE=0.026) and -0.034 (SE=0.029) for girls. Repeating this exercise for the second sub-sample yields different results (columns 5-8)). The Bayesian estimates obtained here are not statistically significant and they even change sign for girls (0.071 SE=0.055 and 0.048 SE=0.051 for boys and 0.022 SE=0.035 and 0.057 SE=0.034 for girls). We view these results as additional evidence that the bias measure reflects teachers' behavior and not classroom unobserved skills or characteristics.

In the second falsification test, we examine if teacher biases might be driven by boys versus girls differences in cognitive skills in class. We define an alternative measure of teachers' attitude toward low/high achievers in class instead of our previous bias measure based on students' gender. This measure is defined at the class level by the difference between high performing students' and low performing students' average gap between the school score (non-blind) and the national score (blind). Higher/lower achievers are defined as students with higher/lower scores in GEMS 5th grade than the class average score (i.e., their mean scores in all three subjects are higher/lower than the average scores in all subjects in class). Thus, this alternative measure captures teachers' attitude toward low versus high achievers in class. We therefore test if having a teacher with a more pro-lower achievers in class have a differential effect on their students later outcomes based on their 5th grade achievements (Table 6 panel C), and whether same teacher biases measures according to this alternative definition are correlated with other teacher biases (Table 4).

[42] We note we exclude altogether from this analysis and estimation the schools for whom we do not have full information about teachers of different subjects in different classes (about 20% of the classes in the sample).

Appendix Table A10 replicates the results reported in Table 6 panel C and Appendix Table A11 replicates the results reported in Table 4 for the Bayesian alternative measure of teacher attitude toward low versus high achievers on test scores.[43] The estimated effects of the alternative measure on low versus high achievers in class are not significantly different from zero when using the basic or jackknife versions. Moreover, the correlation between same teacher's alternative bias measures is not higher than the correlation between the alternative bias measures of two different teachers. This falsification test provides further reinforcement that our results are not driven by the variation in boys' versus girls' cognitive ability in class.[44]

Effects on Choice of Advanced Courses in Math and Science

In this sub-section we present and discuss results of estimating the effect for each subject separately. In Table 9 we report the estimated effect of Bayesian teachers' biases in a specific subject on students' later scores in that subject based on pooling the middle school GEMS test scores and the high school matriculation scores data. We use a regression specification that includes students' characteristics and 5th grade test scores, year and primary school by subject fixed effects. As in Table 6 Panel C, we present the estimates based on the Bayesian teachers' biases, using the basic or the jackknife bias measures.

The estimated effect of teachers' biases are positive and significant for boys in all three subjects and based on both definitions of teachers' biases measures. Furthermore, the estimated effect of math teachers' biases for boys is the largest effect though it is not statistically different from the estimated effect on each of the other two subjects. For girls, the estimated effects are all negative but are statistically different from zero only in math. The implication of these estimates is that increasing the math teachers' biases against girls, for example from zero to one (biases measures are standardized), will increase boys' test scores by 0.108 (/0.067) standard deviation based on the basic (/jackknife) bias measure and decrease girls' test scores by 0.064 standard deviations (based on the estimate obtained when using the basic measure of teachers' biases.

The evidence for the effects of teachers' biases on students' choice of advanced courses[45] in English, math and science in high school (equivalent to honors classes in the US) is presented in Table 10. In panel A, we present the effect of math Bayesian teachers' biases on the probability of

[43] We note that we implemented a Bayes shrinkage estimation strategy and normalized both basic and jackknife versions of this alternative measure. The range of this measure is [6.663, -3.31] and 78.6% of this measure's variation is within schools.

[44] In order to further test if our results are derived by the natural variation in the gap between teachers' perception of student ability and the exam's measurement of student ability which is not related to students' gender, we simulate one thousand random assignments of gender to students. These simulations yield similar results for the basic measure (same signs and significance level of estimates) in only 11 percent of the cases and for the jackknife measure in only 3 percent of the cases (replicating both the effect of the bias measure on students' later outcome (Table 6 Panel C) and the correlations of same teacher bias measures vis-à-vis that of different teachers (Table 4)).

[45] An advanced class yields 5 matriculation credits, and intermediate level class yields 4 credit and a basic class yields 3 matriculation credits.

successfully completing a math advance level course, and in panel B we present the estimated effect of Bayesian teachers' biases measure on the total number of matriculation credits a student gains in each of these advanced courses. Both panels present evidence based on estimating a separate regression for each subject, using the specification that includes students' 5th grade test scores and characteristics, year and primary school fixed effects. We also note that in order to test the effect of teachers' biases on the number of science matriculation credits (which includes computer science, chemistry and physics courses) [46] we use the math teachers' biases since science matriculation scores are more correlated with math scores than with English or Hebrew scores.

Table 10 panel A presents the marginal effects at the means of math Bayesian teachers' biases on the probability of successfully completing an advanced course in math (4 or 5 credits). The estimated effect of math teachers' biased behavior on the probability of successfully completing advanced studies in math based on using the basic measure of teacher bias is positive and significant for boys (0.037, SE=0.018) and negative and significant for girls (-0.040, SE=0.019). The estimates based on the jackknife definition have the same signs but are not statistically different from zero. In order to assess the magnitude of the effects, we simulate a scenario where a group of boys/girls is moved from a neutral teacher to one with a boys' bias of one. Referring to the basic definition of teacher biases measure, this will increase the completion rate of boys in an advanced math program by 3.7 percentage points and decrease that of girls by 4 percentage points.

Table 10 panel B presents the estimated effect of Bayesian teachers' biases on students' total number of matriculation credits gained in English, math and science. The estimated effects of math teachers' biases on the number of math and science credit units are positive and significant for boys according to both basic and jackknife definitions of teachers' biases, and negative and significant for girls in math credit units according to the first definition. Similarly, the estimated effects of English teachers' biases on the number of English credits is positive and significant for boys according to both basic and jackknife definitions of teachers' biases. As before, we can simulate the impact of moving from a neutral teacher to one with a boys' bias of one. Based on the basic definition of teacher biases measure, such change will decrease girls, number of matriculation units in math by 0.167 units and increase boys' number of units in several subjects: in math it will increase the number of units by 0.192, in English by 0.136 units, in overall science subjects by 0.341 units.

The estimated effects of teachers' biases on math test scores are of special interest because of the considerable gender gap in math achievements at the end of high school and the impact on future labor market outcomes.[47] Our results suggest that students' math test scores and advanced math

[46] These subjects are chosen since they constitute the basic requirement for university admission to STEM studies in most universities in Israel.

[47] Several papers which have documented the correlation between students' math test scores and their future labor market income, suggest that the gender gap in math test scores in later stage of high school leads to the underrepresentation of women in STEM careers and that this sorting might be one of the reasons for gender differences in adult wages (Paglin and Rufolo (1990), Brown and Corcoran (1997)).

studies' completion rates are affected mainly by their math teachers' biases. These results are in line with the different teaching practices towards boys and girls in class.[48] To shed light on the effect size of these estimates, we examine how eliminating teacher gender bias against girls in math affects the gender gap in math achievements. Based on our evidence in Appendix Table A2, a simulated 0.07 decrease in a math teacher biased behavior will decrease boys' math achievements in middle/high school based on the basic(/jackknife) definition of teacher bias by 0.008(/0.005) standard deviations and increase girls' achievements by 0.005 standard deviations. Such effect size will decrease the gap in favor of boys in middle/high school from 0.057 to 0.044(/0.052) standard deviations. It will also decrease boys' advanced math studies' completion rate in high school by 0.3 percentage point and will increase girls' completion rate by 0.3 percentage point according to the basic definition of teacher bias. As a result, the gender gap in studying math at the highest level in high school would decline from 3.2 to 2.6 percentage points. A more drastic decline in math teachers' biases, say a decrease of one standard deviation in the math bias. Such decline will reverse the gender gap in math achievements in middle/high school according to the basic (/jackknife) teacher bias definition from a gap of 0.06 SD in favor of boys to a gap of 0.115 SD (/0.01) in favor of girls. A similar change will also impact the gender gap in completion rates of advanced math studies from 3.2 percentage points in favor of the boys, to 4.6 percentage points in favor of girls according to the basic definition of teacher bias. In addition, it will also affect both number of math and science credit units: it will reverse the gender gap in total number of math credit units from a gap of 0.1 credit units in favor of boys to a gap of 0.1 (/0.06) credit units in favor of girls according to the basic (/jackknife) teacher bias definition and reduce the gender gap in science units in favor of boys from a gap of 1.12 credit units to 0.8 (/0.9) credit units according to the basic (/jackknife) teacher bias definition.

The long term effects on high school matriculation programs and test scores have meaningful economic consequences for quantity and quality of post-secondary schooling and on earnings at adulthood. In Appendix Table A12 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 with quality (university schooling, academic colleges and other). They are also positively correlated with annual earnings at age 30; for example, each credit unit is associated with a gain of NIS 1,270 ($343) per annum and having a matriculation certificate is associated with a gain of NIS 15,648 ($4,230).

Pursuing a similar question to the one we address in the current paper and in an earlier draft (Lavy and Sand 2014) and using a comparable methodology, Terrier (2015) presents similar evidence

[48] See the discussion in the introduction on the different teaching method implemented by math teacher toward boys and girls in class.

on the effects of teachers' gender biases using French data. She relies on a similar definition of teachers' gender biases and tests for the effect of teachers' gender biases on the gender gap in achievements in class. She finds that the classes in which teachers present a high degree of discriminatory in favor of girls are also classes in which girls tend to progress significantly more than boys and choose a high level of general training at a higher probability compared to boys. For example, she reports that having a math teacher who is one SD more biased in favor of girls increases girls' probability to select a scientific track by 2.7 percentage points compared to boys. Although the outcomes of the two papers are not completely equivalent, we find that increasing math teacher biases by one SD in favor of girls increase their completion rates in advanced math studies relative to boys by 7.8 percentage points, while it does not affect the relative completion rates in advanced science courses.

Heterogeneous Treatment Effects of Teachers' Biases

To gain further insight into the effects of teachers' gender biased behavior on students' academic success we explore heterogeneous effects across two dimensions. In Table 11 we present the estimated effect of the teacher's bias on test scores for boys and for girls separately, based on different stratifications of the full sample. We present the estimates of the Bayesian teacher bias effect on both GEMS and matriculation test scores. We use the specification similar to the one presented in Table 6 Panel C, which includes a dummy variable for type of exam, students' 5^{th} grade test scores and characteristics, year and primary school fixed effects. We examine first the heterogeneous treatment effects of teachers' biases by parental education level (whether the average parental years of schooling is above the median of 12 years)[49], and then the heterogeneous treatment effects by the gap in parental education (referring to cases were mothers are more educated than fathers and vice versa).

The first part reports the estimated effects based on stratifying the sample by parental level of schooling. According to the relevant sociology and psychology literature, the mother's level of education and employment status is correlated with a more egalitarian attitude towards gender roles.[50] We thus posit that students of educated mothers should be less influenced by teachers' biases. The table indicates that the estimated effect is stronger for students with low parental education for both gender according to both definitions of teacher bias measure, though the differences between the groups are not statistically different from zero: the estimated effect of teachers' biases on students with low parental education is 0.107 (SE=0.033) for boys and -0.065 (SE=0.031) for girls based on the basic definition of teacher bias, while the estimated effect on students with high parental education is only 0.073 (SE=0.024) for boys and -0.033 (SE=0.021) for girls. The jackknife measure exhibits a similar pattern, but is statistically significant only for boys with low parental education.

[49] We note that stratifying the sample by mothers' or fathers' education levels yields similar results.
[50] See, for example, Hoffman (1977) and Herzog et al. (1983).

Following a similar line of reasoning, we also consider the heterogeneous treatment effects of teachers' biases based on a slightly different stratification of the sample, where we group students based on the within-family parental education gap. We postulate that children from families where the mothers are more educated than the fathers might also be less prone to the influence of gender biases at school. The treatment effects of teachers' biases by parental education gap are presented in the second part of Table 11. The table indicates that the teachers' biases have similar effects on boys and girls according to the parental education gap in their family. However, the estimates are statistically significant for boys in both cases of parental education gap and for girls whose mothers are less/equally educated than their fathers according to the basic measure of teachers' biases.

5. Conclusions

In this paper we investigate how primary school teachers' positive 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 also define another measure which exclude the gap between the two exams' scores of the student himself from the measure (jackknife version) .We then estimate the impact of this measure of teachers' 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 in a specific subject. We compare students in the same primary school and in the same subject who are exposed to teachers, who might have different patterns of gender biases. We address several threats to the interpretation of our findings and demonstrate that our estimates reflect teachers' behavior and not students' characteristics or behavior.

Based on a sample of Tel-Aviv schools, the results we present suggest that teachers' more positive relative assessment of boys in a specific subject has a positive (negative) and significant effect on boys' (girls') overall future achievements in that subject. Moreover, the magnitudes of these effects are more pronounced for boys than for girls, especially when estimated using the jackknife version of the teacher bias measure. These effects persist through middle school and high school and actually have dramatic implications for matriculation exam scores and on the probability of receiving a matriculation diploma.

We also find that gender bias among math teachers has an especially large effect on students' math test scores and on total number of matriculation credit units in advanced math and science studies in high school. The estimates of the 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 might partly results from teachers' biases against girls in

math, eliminating these biases will go a long way toward reducing the math achievement gender gap, and it will also decrease the gender gap in enrollment in advanced math and science studies. The impact on the various end-of-high-school matriculation outcomes carries meaningful economic consequences, because these high stakes outcomes sharply affect the quantity and quality of post-secondary schooling as well as impacting earnings in adulthood.

6. References

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

Alesina, A., P. Giuliano, and N. Nunn. 2013. "On the Origin of Gender Roles: Women and the Plough", *Quarterly Journal of Economics* 128(2), 469-530.

Ashraf, Q. and O. Galor, 2013. "The Out of Africa" Hypothesis, Human Genetic Diversity, and Comparative Economic Development", *American Economic Review*, 103(1), 1-46.

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.

Bertrand, M., D. Chugh and S. Mullainathan, 2005, "Implicit Discrimination", *American Economic Review*, 95(2), 94-98.

Bertrand, M. and E. Duflo, "Field Experiments on Discrimination", Forthcoming in: A. Banerjee and E. Duflo (Eds.), Handbook of Field Experiments.

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.

Blass, N., Tsur, S. and N. Zussman, 2014. "Segregation of Students in Primary and Middle Schools", Bank of Israel Discussion Paper No. 2014.07.

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.

Botelho, F., Mdeira, R.A. and M.A., Rangel 2015. "Racial Discrimination in Grading: Evidence from Brazil", *American Economic Journal: Applied Economics*, 7(4), 37-52.

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.

Burnsztyn, L. and R. Jensen, 2015, "How Does Peer Pressure Affect Educational Investments?", *The Quarterly Journal of Economics*, 130(3), 1329-1367.

Carlana, M., 2017, "Stereotypes and Self-Stereotypes: Evidence from Teachers' Gender Bias"_ Draft, October 2017.

Carrell S.E., M.E., Page and J.E., West, 2010. "Sex and Science: How Professor Gender Perpetuates the Gender Gap" *The Quarterly Journal of Economics*, 125 (3), 1101-1144.

Chetty, R, Friedman J.N., Hilger N,. and E. Saez, 2011. "How Does Your Kindergarten Classroom Affect Your Earnings? Evidence from Project Star." *Quarterly Journal of Economics* 126(4), 1593-1660.

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.

Dee, T. S., 2005. "A Teacher Like Me: Does Race, Ethnicity, or Gender Matter?" *American Economic Review*, 95(2), 158-165.

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.

Fershtman, C. and U. Gneezy, 2001. "Discrimination in a Segmented Society: An Experimental Approach", *The Quarterly Journal of Economic* 116(1), 351-377.

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.

Gershenson, S., Holt, S., and N. Papageorge, 2016. "Who believes me? The effect of student-teacher demographic match on teachers' beliefs." *Economics of Education Review*, 52, 209-224.

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. "Gender Similarities Characterize Math Performance", *Science* 321(5888), 494–495.

Hyde, J.S., and S. Jaffe, 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)

Kane, T. J., and Staiger, D. O., 2008. Estimating Teacher Impacts on Student Achievement: An Experimental Evaluation, NBER Working Paper No. 14607.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., 2010 "Effects of Free Choice among Public Schools," *Review of Economic Studies*, Volume 77 Issue 3 (July): 1164-1191.

Lavy, V., 2016. "What Makes an Effective Teacher? Quasi-Experimental Evidence", *CESifo Economic Studies*, 62 (1): 88-125.

Lavy, V., 2016 "Long Run Effects of Free School Choice: College Attainment, Employment, Earnings, and Social Outcomes at Adulthood", NBER working Paper.

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.

Lavy, V. and E. Sand, 2017. "The Effect of Social Networks on Students' Academic and Non-Cognitive Behavioral Outcomes: Evidence from Conditional Random Assignment of Friends in School", *Economic Journal*, Forthcoming.

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.

Morris, C. 1983. "Parametric Empirical Bayes Inference: Theory and Applications". *Journal of the American Statistical Association* 78, 47-55.

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.

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", *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.Sadker, D., Sadker, M., and K. R. Zittleman, 2009. "Still Failing at Fairness: How Gender Bias Cheats Girls and Boys in School and What We Can Do About It". (New York: Charles Scribner).

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.

Washington, E. L., 2008. “Female Socialization: How Daughters Affect Their Legislator Fathers”, The *American Economic Review* 98(1), 311-332.

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	0.083
	(0.273)	(0.244)	(0.276)
Number of Students in Elementary Schools	867	1127	1017
Number of Elementary Schools	17	20	20
Number of Elementary Classes	33	41	38
Number of Middle Schools	5	7	5

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

Table 2: Means and Standard Deviations of National and Primary School Exams Scores and Difference Between them, by Gender

	Boy			Girl			Difference Between School and National Exams Scores of Boys and Girls
	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	-0.133	-0.135	0.003	0.138	0.139	-0.002	0.005
	(1.015)	(1.026)	(1.054)	(0.965)	(0.952)	(0.932)	
Math	0.025	-0.013	0.039	-0.025	0.013	-0.039	0.078
	(1.007)	(1.033)	(0.971)	(0.992)	(0.964)	(0.913)	
English	-0.074	-0.046	-0.028	0.076	0.047	0.029	-0.057
	(1.027)	(1.033)	(1.012)	(0.965)	(0.963)	(0.944)	
Number of Students	4245	4246	4245	4122	4123	4122	8367

Notes: The national exam scores and the primary school exam scores are standardized scores. The number of students refers to the number of students in all three subjects. The last column (column 7) 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 deviations 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	-0.078	0.076
	(1.061)	(0.908)	(1.023)	(0.970)
Math	0.012	-0.012	0.048	-0.046
	(1.043)	(0.952)	(1.052)	(0.943)
English	-0.081	0.085	-0.011	0.011
	(1.036)	(0.952)	(1.000)	(0.976)
Number of Students	1676	1618	1481	1516

Notes: The national exam scores are standardized scores. The number of students refers to the number of students tested in all three subjects. Matriculation test scores are weighted based on the number of credit units taken, as computed by the Ministry of Education. Standard deviations are reported in parentheses.

Table 4: Correlations between Bayesian Biases of Teachers by Subjects of Instruction

	Same Teachers	Different Teachers	Overall		Within School	
	Teachers Biases in Hebrew	Teachers Biases in Hebrew	Teachers Biases in Hebrew	Teachers Biases in Math	Teachers Biases in Hebrew	Teachers Biases in Math
	(1)	(2)	(3)	(4)	(5)	(6)
Teachers' Biases in Math	0.486***	0.214	0.363***		0.268**	
	(0.135)	(0.182)	(0.084)		(0.110)	
Teachers' Biases in English			0.286***	0.260***	0.160	0.178
			(0.089)	(0.093)	(0.131)	(0.135)
Number of Observations	36	42	112		112	

Notes: The table presents the estimated correlation coefficient of Bayesian teachers' biases measures by subjects of instruction. The (OLS) estimated coefficient in column 1 is between biases measures of the same teachers who instruct students from the same class both math and Hebrew; and the (OLS) estimated coefficient in column 2 is between biases measures of different teachers who instruct students from the same class in both math and Hebrew. Both last estimated coefficients are from using separate OLS regressions, The estimates in each row in columns 3-4 are the correlation coefficients between bias measures using the sample of all teachers (same or different teachers for each two subjects), from separate OLS regressions. The estimated coefficients in each row in columns 5-6 are similar to those in columns 3-4, but primary school fixed effects are included in the regressions. Standard errors are reported in parentheses. Significance level of regressions are reported as follows: "***"=1% level, "**"=5% level, and "*"=10% level.

Table 5: Correlations of Teachers' Biases Measure with Characteristics of Teachers

	Age Dummy (dummy=1 if Older than the Median)	Ethnicity Asia/Africa	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)
6th Grade School Fixed Effects	0.075	0.556*	0.326	-0.465*	0.065	-1.013*	-0.556
	(0.313)	(0.329)	(0.302)	(0.262)	(0.147)	(0.449)	(0.517)
Number of Teachers	113	114	112	112	112	108	108

Notes: The table presents the estimated correlation between several teachers' characteristics and Bayesian teachers' stereotypical bias measure. Each regression includes school and subject and year fixed effects. The estimates in each column in columns 1-6 are from a separated regression. Standard errors are reported in parentheses. Significance level of regressions are reported as follows: "***"=1% level, "**"=5% level, and "*"=10% level.

Table 6: Estimated Effect of Bayesian Teachers' Biases on Test Scores

	Basic Teacher Bias Measure		Jackknife Teacher Bias Measure	
	Boy	**Girl**	**Boy**	**Girl**
	(1)	(2)	(3)	(4)
A. 8th Grade GEMS Test Scores				
OLS	0.093***	-0.033	0.052*	0.011
	(0.030)	(0.035)	(0.029)	(0.034)
6th Grade School by Subject Fixed Effects	0.088***	-0.039	0.046**	0.003
	(0.021)	(0.024)	(0.021)	(0.025)
6th Grade School by Subject Fixed Effects and Student Characteristics	0.094***	-0.038	0.054**	0.005
	(0.021)	(0.025)	(0.022)	(0.026)
Number of Observations	2938			
B. Matriculation Test Scores				
OLS	0.121***	-0.035	0.080*	0.005
	(0.044)	(0.049)	(0.042)	(0.048)
6th Grade School by Subject Fixed Effects	0.090***	-0.052**	0.048*	-0.015
	(0.026)	(0.026)	(0.026)	(0.026)
6th Grade School by Subject Fixed Effects and Student Characteristics	0.091***	-0.064***	0.052**	-0.026
	(0.025)	(0.023)	(0.025)	(0.024)
Number of Observations	2682			
C. Pooled 8th Grade GEMS and 12th Grade Matriculation Test Scores				
6th Grade School by Subject Fixed Effects and Student Characteristics	0.092***	-0.051***	0.052**	-0.011
	(0.021)	(0.020)	(0.021)	(0.021)
Number of Observations	5620			

Notes: The table reports the estimated effect of Bayesian teachers' gender biased behavior on students' academic achievements, based on estimating equations 1 and 2. The estimates of the effect of Bayesian teachers' biases according to the basic teacher bias measure is presented in column 1-2 and the estimates according to the jackknife teacher bias measure is presented in column 3-4. Each regression includes a dummy for boy, the teacher bias measure and their interaction, subject and year fixed effects and student's 5th grade test score. The estimated effects on boys are presented in columns 1 and 3, and the estimated effects on girls are presented in columns 2 and 4. Panels A and B show 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 stacked. Panel C reports the estimated effect of Bayesian teachers' biases on both 8th grade test scores and matriculation test scores, where the scores in all three subjects and in all tests (8th grade test scores and matriculation test scores) are stacked and a dummy variable for type of test (GEMS or matriculation tests) is added to the regression. All test scores are standardized scores, by year and subject. Standard errors are clustered by class and student and are reported in parentheses. Significance level of regressions are reported as follows: "***"=1% level, "**"=5% level, and "*"=10% level.

Table 7: Estimated Effect of Bayesian Teachers' Biases on Other Educational Outcomes

	Basic Teacher Bias Measure		Jackknife Teacher Bias Measure	
	Boy	**Girl**	**Boy**	**Girl**
	(1)	(2)	(3)	(4)
A. Probability of Receiving a Matriculation Diploma				
OLS	0.053***	-0.019	0.041**	-0.003
	(0.018)	(0.022)	(0.018)	(0.020)
6th Grade School by Subject Fixed Effects	0.039***	-0.035***	0.026**	-0.018*
	(0.010)	(0.011)	(0.011)	(0.011)
6th Grade School by Subject Fixed Effects and Student Characteristics	0.035***	-0.037***	0.022**	-0.020*
	(0.010)	(0.011)	(0.010)	(0.010)
Number of Observations	2798			
B. Total Number of Successfully Completed Matriculation Exams' Units				
OLS	1.445**	-0.631	1.012*	-0.146
	(0.569)	(0.517)	(0.555)	(0.492)
6th Grade School by Subject Fixed Effects	1.237***	-0.699**	0.766**	-0.252
	(0.344)	(0.280)	(0.343)	(0.267)
6th Grade School by Subject Fixed Effects and Student Characteristics	1.095***	-0.821***	0.654*	-0.338
	(0.338)	(0.264)	(0.341)	(0.258)
Number of Observations	2851			

Notes: See Table 6. The table reports the estimates of Bayesian teachers' biases on other educational outcomes: Panel A shows results of the estimated effect of Bayesian teachers' biases on the probability of receiving a matriculation diploma (the estimates are marginal effects at the means from separate logistic regressions) and Panel B shows results of the estimated effect of Bayesian teachers' biases on the total number of successfully completed matriculation exams units. Standard errors are clustered by class and student and are reported in parentheses. Significance level of regressions are reported as follows: "***"=1% level, "**"=5% level, and "*"=10% level.

Table 8: Estimated Effect of Bayesian Teachers' Biases on Pooled 8th Grade GEMS and 12th Grade Matriculation Test Scores, With Alternative Control Variables

	Basic Teacher Bias Measure		Jackknife Teacher Bias Measure	
	Boy	**Girl**	**Boy**	**Girl**
	(1)	(2)	(3)	(4)
Difference Between Boys' and Girls' 5th Grade GEMS Scores	0.094***	-0.048**	0.055**	-0.008
	(0.021)	(0.020)	(0.021)	(0.021)
6th Grade Class Fixed Effects and Student Characteristics and the Difference Between Boys' and Girls' 5th Grade GEMS Scores	0.076***	-0.059**	0.036	-0.019
	(0.023)	(0.026)	(0.025)	(0.027)
Number of Observations	5620			

Notes: The specification is the same as in Table 6 Panel C, but in the first row include the difference between boys' and girls' 5th grade GEMS scores and in the second row include additionally 6th grade class fixed effects instead of 6th grade school by subject fixed effects. Standard errors are clustered by class and student and are reported in parentheses. Significance level of regressions are reported as follows: "***"=1% level, "**"=5% level, and "*"=10% level.

Table 9: Estimated Effect of Bayesian Teachers' Biases on Pooled 8th Grade GEMS and 12th Grade Matriculation Test Scores, by Subject

	Basic Teacher Bias Measure		Jackknife Teacher Bias Measure	
	Boy	**Girl**	**Boy**	**Girl**
	(1)	(2)	(3)	(4)
Hebrew	0.079**	-0.042	0.040	-0.008
	(0.034)	(0.035)	(0.029)	(0.036)
Math	0.108***	-0.064**	0.067**	-0.011
	(0.030)	(0.031)	(0.033)	(0.030)
English	0.097***	-0.038	0.054**	-0.003
	(0.030)	(0.029)	(0.027)	(0.029)

Notes: See Table 6 Panel C. Each row present estimates from separate regressions for each subject. Standard errors are clustered by class and student and are reported in parentheses. Significance level of regressions are reported as follows: "***"=1% level, "**"=5% level, and "*"=10% level.

Table 10: Estimated Effect of Bayesian Teachers' Biases on Choice of Advanced Courses in Math and Science

	Basic Teacher Bias Measure		Jackknife Teacher Bias Measure	
	Boy	**Girl**	**Boy**	**Girl**
	(1)	(2)	(3)	(4)
A. Probability of Successfully Completing an Advanced Math Level Course in High School				
Math (dummy=1 if # units=5|4)	0.037**	-0.040**	0.025	-0.017
	(0.018)	(0.019)	(0.018)	(0.018)
B. Total Number of Successfully Completed Units in Science, Math and English Courses in High School				
English	0.136**	-0.012	0.061	-0.029
	(0.048)	(0.055)	(0.046)	(0.056)
Math	0.192***	-0.167***	0.149***	-0.060
	(0.048)	(0.055)	(0.052)	(0.052)
Sum of Number of Science Units (Math, Computer Science, Chemistry and Physics)	0.341**	-0.111	0.229*	0.101
	(0.125)	(0.133)	(0.131)	(0.125)

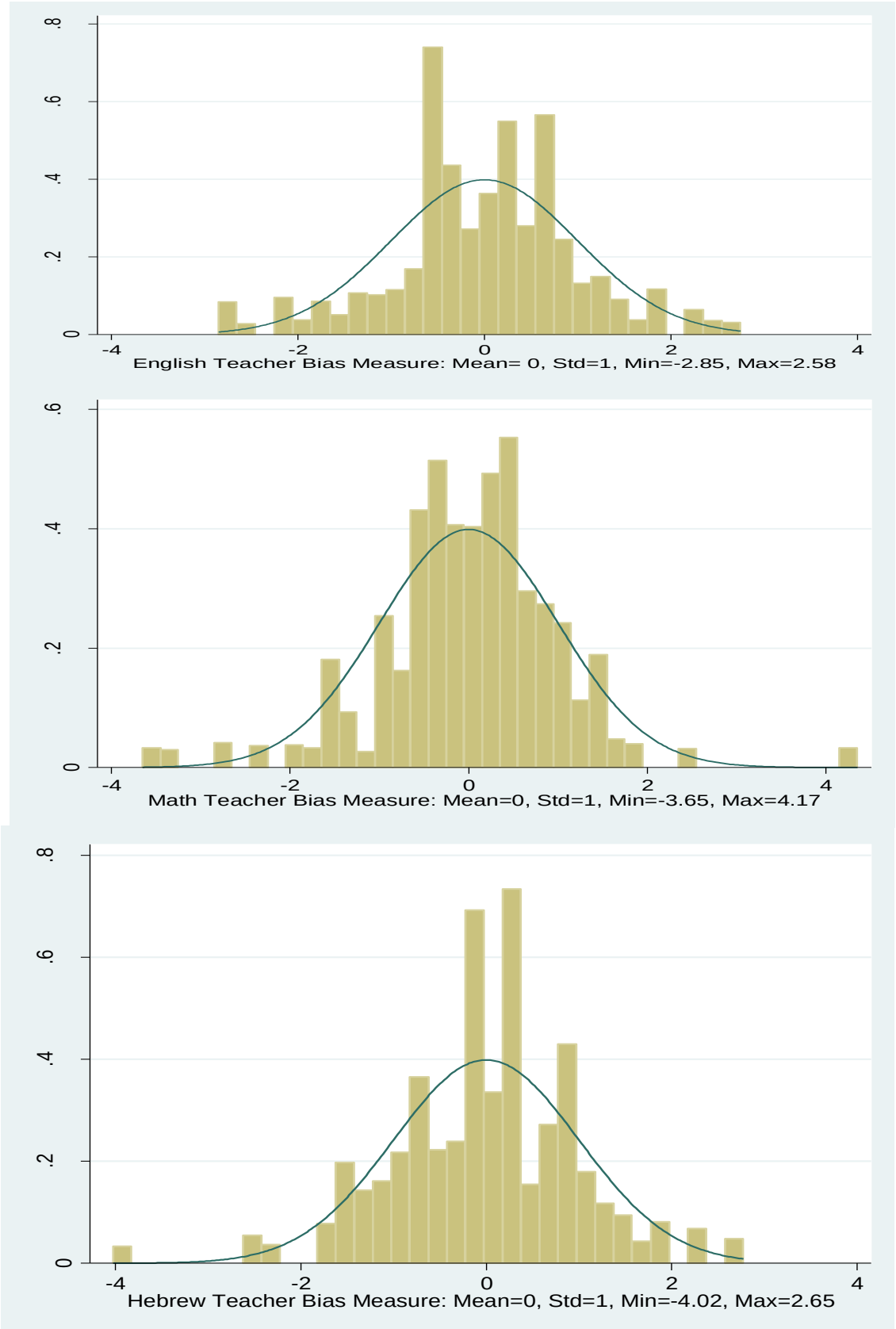
Notes: See Table 6 Panel C. Panel A presents the marginal effect of math teachers' biases at the means from a logistic regression. The dependent variable is discrete and equal one if the number of matriculation credits exceeds a certain level 4 credit units. In Panel B, each row present estimates from 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. The sum of number of science units is the number of units the student takes in math, physics, chemistry and computer science courses altogether. Standard errors are clustered by class and student and are reported in parentheses. Significance level of regressions are reported as follows: "***"=1% level, "**"=5% level, and "*"=10% level.

Table 11: Estimated Effect of Bayesian Teachers' Stereotypes on Pooled 8th Grade GEMS and 12th Grade Matriculation Test Scores, by Sub-Groups

	Basic Teacher Bias Measure				Jackknife Teacher Bias Measure			
	Boy	**Girl**	**Boy**	**Girl**	**Boy**	**Girl**	**Boy**	**Girl**
	(1)	(2)	(3)	(4)	(1)	(2)	(3)	(4)
	Low Parental Education		**High Parental Education**		**Low Parental Education**		**High Parental Education**	
	0.107***	-0.065**	0.073***	-0.033	0.060*	-0.026	0.037	0.002
	(0.033)	(0.031)	(0.024)	(0.021)	(0.035)	(0.030)	(0.024)	(0.024)
Number of Observations	2830	.	2790	.	2830	.	2790	.
	Mothers are More Educated than Fathers		**Mothers are Less/Equally Educated than Fathers**		**Mothers are More Educated than Fathers**		**Mothers are Less/Equally Educated than Fathers**	
	0.091***	-0.010	0.084***	-0.067**	0.059*	0.030	0.042*	-0.028
	(0.033)	(0.031)	(0.021)	(0.024)	(0.033)	(0.031)	(0.021)	(0.024)
Number of Observations	1716	.	4140	.	1716	.	4140	.

Notes: See Table 6 Panel C. The table reports the estimated effect of Bayesian teachers' gender biased behavior on students' academic achievements, by sub-groups. The scores in all three subjects (math, English, and Hebrew) and in all tests (8th grade GEMS and matriculation exams) are stacked and the regression includes a dummy for the type of exam (8th grade GEMS versus matriculation exams). High parental education is defined as more than 12 years of average parental schooling. Standard errors are clustered by class and student and are reported in parentheses. Significance level of regressions are reported as follows: "***"=1% level, "**"=5% level, and "*"=10% level.

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 exam scores (non-blind) and the national exam scores (blind), by subject.