On The Origins of the Gender Human Capital Gap: Short and Long Term Effect of Teachers' Stereotypes*

October 2014

Victor Lavy
Hebrew University, University of Warwick, and NBER

Edith Sand
Bank of Israel

Abstract

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

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. We also benefited from comments from seminar participants at the Hebrew University, and Bank of Israel.

1. Introduction

Over the past decades there has been a large increase in female human capital investments and labor force participation. The ratio of male to female college graduates decreased continuously and in recent years it was even reversed in many countries (Goldin et al. (2006), and Becker et al. (2010)). This trend is partly due to the rise in graduation rate of women in what used to be historically men dominated fields such as math, science and engineering. The math test score gender gap is of special interest because it was shown to be a good predictor of future income (Murnane et al. (1995) and Paglin and Rufolo (1990)). However, there is still a considerable gender gap in academic outcomes and in employment in math and science intensive fields. For example, evidence based on recent PISA testing[1] show that in most countries girls outscore boys in reading while being outscored in math (Machin and Pekkarinen (2008)). This gap is shown to grow during early years of schooling (Fryer and Levitt (2010)), and is larger at the upper tail of the test scores distribution (Ellison and Swanson (2010), Hyde et al. (2008)).

What explains these gender disparities in cognitive performance and in math scores in particular is still an open question. Some emphasize the role of biological gender differences in determining gender cognitive differences[2] while others emphasis 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 are raising their children[3]; by affecting educational environment at school and teachers' attitudes[4]; and by determining social and

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

[2] This approach suggests that the difference in chromosomal determinants (Vandenberg (1968)), hormones 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.

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

[4] Stereotypical attitudes of teachers towards girls and boys in class are said to affect students' self-image and self-confidence, and influence substantially their future educational outcomes. Such mechanisms have been

cultural norms[5]. This debate is based on limited credible scientific evidence because it is difficult to disentangle the impact of biological gender dissimilarities from environmental conditions and because it is also difficult to measure stereotypical attitudes and test their causal implications.

In this paper we measure and test the effect of gender stereotypes in a schooling environment. We estimate the impact of primary school teachers' stereotypes about boys' and girls' math and languish ability on cognitive outcomes in later grades. We measure teacher's stereotypes by comparing his average marking of boys' and girls' in a "non-blind" classroom exam to the respective means in a "blind" national exam marked anonymously. We show that there is large variation within school in this measure of stereotypical bias/attitude and that it has a significant effect on the academic achievements of both genders during middle school and high school, in math, science and languish and also on the choice of the level of math and science study programs in high school. This high stake choice determines practically whether a student will be able to meet requirements of admission to science and math studies at universities later in life.

The systematic difference between non-blind and blind assessment across groups as a measure of discrimination or stereotypes was pioneered in economics by Blank (1991) and Goldin and Rouse

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

5 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 comparing to cultural environment. They show that there is a positive correlation between gender equality and gender gap in mathematics achievements according to data from OECD's international tests (PISA 2003) and data that measure gender equality taken from the World Economic Forum's Gender Gap Index (GGI). Moreover, they show that these results are not driven by biological differences across countries by using a genetic distance measurement between the populations. Pope and Sydnor (2010) and Fryer and Levitt (2010) replicate this methodology for different sets of countries.

(2000).[6] In economics of education, this approach was first implemented in Lavy (2008) to measure gender bias in grading by teachers and it was followed in other studies, for example, Björn, Höglin, and Johannesson (2011), Hanna and Linden (2012), Cornwell, Mustard, Van Parys (2013) and Burgess and Greaves (2013), that followed the same methodology using data from other countries and getting overall similar evidence about the effect of teachers' stereotypes/biases. In the present paper, however, we go beyond measuring teachers' stereotypical attitudes and focus on their implications for gender differences in human capital formation. We think this paper is perhaps the first to link stereotypical attitudes to the origin of the gender gap in human capital.[7] We test whether teachers' bias towards one of the sexes, as reflected by a more positive evaluation on the "non-blind" tests relative to the "blind" tests of this group, reinforce this group's future achievements and affect their educational choices.

We use in this paper data that allows us to evaluate the impact of gender biases of teachers on students' cognitive outcomes in later years, by following three cohorts of 6th grade students between the years 2002-2004 in Tel-Aviv, Israel. Tracking students from primary school to high school enables us to both assess students' exposure to gender stereotypes of teachers in primary school, and estimate its effect on both 8th grade test scores in national tests and on the matriculation tests scores at the end of high school, over six years after the timing of exposure to stereotypical discrimination. In addition, we are able to examine whether this measure of teachers' stereotypical attitude is correlated with certain teachers' characteristics, such as age, ethnicity, marital status and gender composition of own children.

[6] Blank (1991) shows that the probability of papers being accepted to economic journals depends on authors' affiliation. Goldin and Rouse (2000) examine sex-biased hiring patterns in orchestras by comparing blind and non-blind auditions.

[7] Alesina, Giuliano and Nunn (2013) examine the historical origins of existing cross-cultural differences in beliefs and values regarding the appropriate role of women in society. They find that the descendants of societies that traditionally practiced plough agriculture today have less equal gender norms, measured using reported gender-role attitudes and female participation in the workplace, politics, and entrepreneurial activities. Also related is Reuben, Sapienza, and Zingales (2014) who study the effect of stereotypes in an experimental market, where subjects were hired to perform an arithmetic task that, on average, both genders perform equally well. They find that when the employer had no information other than candidates' physical appearance, women were only half as likely to be hired as men, while revealing information on the candidate arithmetic ability reduced the degree of discrimination, but did not eliminate it completely.

For identification, we rely on the conditional random assignments of teachers and students to classes within a given grade and a primary school. Comparing students that attended the same primary school but were randomly assigned to different teachers, who might have different degree of stereotypical attitudes, account for any observed and unobserved teacher, class and school characteristics. An analysis based on within school variation insures that the measurement of teachers' stereotyping attitude is not correlated with students' observables. We test directly for randomness of class assignment within a given school and we also present robustness tests regarding the correlation between our measure of teachers' stereotypical attitude and students' and classes' predetermined characteristics.

Our results suggest that teachers' over-assessment of boys in a specific subject has a positive and significant effect on boys achievements in that subject national test administered during middle and high school and it has an asymmetric significant negative effect on girls. In addition, primary school math teachers favoring boys over girls affect also the choices made by students while in high school regarding their enrollment in high level math studies: they encourage boys and discourage girls from engagement in advanced math courses offered in school; since these courses are pre-requisites for admission to higher education in these subjects, such teachers' stereotypical attitudes contributes to the gender gap in academic degrees in fields like engineering and computer science and by implication they also contribute to the gender gap in related occupations. Our evidence also points to spillover effects of stereotypical attitudes of teachers across subjects, implying that a teacher bias against girls or boys in one subject can have a broader influence on students' achievements in other subjects. We also show that these effects have interesting patterns of heterogeneity by parental years of schooling, parental education gap, ethnicity and birth order of children.

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

2. Data

We use in this study data from the Tel-Aviv municipality's school authority administrative records. The baseline sample is sixth-grade students in the city's schools in 2002-2004. We follow these students until 12th grade using an individual identifier and build a panel data set that includes test scores from school exams held in 6th grade, national exams that are held in 5th and 8th grade (GEMS[8] - Growth and Effectiveness Measures for Schools - Meizav in Hebrew), and information on matriculation exams held in 11th-12th grade in high school. These data are merged with Israel Ministry of Education students' registry files that include student, class and school identifiers, students' demographic information[9] (gender, ethnicity, number of siblings, and parents' education). The matriculation tests scores data[10] has also information on the level of studies of each student in math, Hebrew, English, physics and computer science. The GEMS surveys in 5th grade and 6th grade also include questions about various aspects of class environment and teaching practices.

To construct a measure of teachers' stereotypical attitude we combine the GEMS 5th grade external exam's scores with the internal exams scores held in the mid of 6th grade. The GEMS test scores is a "blind" assessment since the GEMS exams are graded by an independent agency where at no stage are the identity and gender of the student revealed. In contrast, the internal exam is graded by the students' teacher and therefore it is a "non-blind" assessment. Note that although the non-blind school exams are held in 6th grade while the blind exam are held in mid of 5th grade, the GEMS individual student's tests scores and her class mean are not revealed to teachers and other school staff and only school mean scores by subject are sent to schools. Therefore it is safe to assume that the "blind" test scores cannot influence the classroom teacher non-blind assessment of each individual student. We also note here that both exams are not high stakes tests because they are not used for matters important directly to students.

[8] During 2001-2006 the GEMS was administered by the Division of Evaluation and Measurement of the Ministry of Education. For more information the GEMS see the Division of Evaluation and Measurement website (in Hebrew): http://cms.education.gov.il/educationcms/units/rama/odotrama/odot.htm. The proportion of students tested is above 90 percent, and the rate of questionnaire completion is roughly 91 percent.

[9] Additional demographic information (the place of birth of the student's grandparents and the birth order of children) is available to us from the Population Registry. We access this data at the research lab of the National Insurance Institute.

[10] The matriculation exams scores and number of study units are available at the Ministry of Education lab which we access remotely. In some cases the matriculation test score for a given subject is based on more than one exam and the final score is a weighted average of the multiple scores, the weights being the credit units that correspond to each exam. The composite score for a subject with multiple tests is computed by the Ministry of Education.

We test the impact of teachers' stereotypical attitude on students' cognitive outcomes in 8th grade (GEMS national tests) and in 12th grade (matriculation national exams). We use the matriculation exams scores to estimate the long run effects of the teachers' stereotypical attitude. In addition, we are able to examine whether teachers' stereotypical attitudes are correlated with various teaching practices differentially by gender, and with teachers' characteristics. Teaching practices are measured based on the students' GEMS survey questions in 5th and 6th grade.[11] Considering the survey replies of boys and girls separately permit distinguishing teachers' teaching practices targeted differentially towards boys and girls. In addition, for some of the teachers[12], we are able to add information from the Population Registry at NII on demographic background (gender, age, marital status, ethnicity and number and gender of children). We use these additional data to examine their correlation with teachers' stereotypical attitude.

The final merged dataset includes the national external test scores (blind) in the 5th grade, the school test scores (non-blind) in the 6th grade, GEMS surveys questions in 5th and 6th grade, national exam GEMS test scores in 8th grade, matriculation exams scores and units of study at the end of high school for 2001-2008, 2002-2009 and 2003-2010, and student characteristics.[13] In addition, we also observe teachers' characteristics for a sub-sample of the schools.

Table 1 presents descriptive statistics, and information about sample size, number of schools, and number of classes for the three sixth grade cohorts that we use: 2002, 2003 and 2004. The panel data includes around 20 secular elementary schools and 5 secular middle schools.[14] The sample from the 2002 cohort includes 867 students (in 33 classes), 1,027 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.

[11] Note that the GEMS questionnaires are taken in fifth and sixth grades. Since the teachers in elementary school are assigned to the same classes for two years consecutively, the teachers in both 5th and 6th grades are the same.
[12] We observe teachers' identifier for only a sub-sample of teachers who are home class teachers and teach at least one of the relevant subjects: math, Hebrew or English.
[13] The raw test scores uses a 1-to-100 scale that we transform into z-scores for each year and for each subject, in order to facilitate interpretation of the results.
[14] The number of middle schools presented in the paper refers only to middle schools with GEMS test scores.

Table 2 presents descriptive statistics for the sub-sample of teachers for whom we have additional demographic information, sample size, and subject of instruction. The sample includes 13 math teachers, 29 Hebrew teachers and 36 teachers who teach both math and Hebrew. English teachers are not part of this sample because none of them also served as a classroom home teacher.

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

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

Next we examine whether the apparent gap between non-blind and blind test scores of boys relative to girls is statistically significant, using the same estimation framework used in Lavy (2008). We assume that the students' scores depend on his gender, type of test (non-blind test=1) and their interaction term. Appendix Table A1 presents estimates based on this basic specification. We first run a regression that includes individuals' characteristics and year, subject and class fixed effects, and then a second regression that includes year, subject and students fixed effects. The estimated coefficient of the interaction term, which measures the difference between the non-blind scores of boys and those of girls, is positive in math, negative in English, and practically zero in Hebrew. While the estimates in Hebrew and English were not statistically different from zero in both regressions, the positive estimates in math were statistically different from zero in the first regression, and marginally significant in the second. These results suggest that there is systematic bias against girls in math exams marking.

Table 4 presents the means of both middle schools and high schools test scores in external exams, separately for boys and girls. The gender gap in favor of girls in Hebrew and English external exams persists to a large extend also in middle school and high school: the gap in Hebrew is 0.3 in middle school and 0.2 in high school, while the gap in English is 0.035 in middle school and in high school. On the other hand, the gender gap in favor of girls in math external exams in primary school is reversed in middle school and in high school: from a gap of 0.028 in favor of girls, to a gap of 0.025 in favor of boys.

Table 5 presents the distribution of students across matriculation exams' units of studies, for boys and girls separately. The proportion of boys that were tested in each one of the three subjects (Hebrew, math and English) ranges from 80.9% to 88.2%, while for girls the range is from 84.3% to 91.3%. However, boys outnumber girls in the advance courses in these subjects. The proportion of boys enrolled in the 5 units course in math is 21.1%, while the respective proportion of girls is only 14.1%. In Science courses this gender gap is even larger: for example 21.6% of boys enrolled in and advanced physics and 13% in an advanced computer science. The respective figures for girls are only 8.1% and 4.5%.

We will further test whether these large differences in achievements, especially in math scores and in enrollment in advance math and science courses, can be explained by the degree of exposure to teachers' stereotypes.

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

3. Identification and Estimation

The main goal of this paper is to investigate how teachers' biases towards one of the sexes reinforce this group's future achievements. As noted, the data allows us to track students from primary school, where students were exposed to different teachers' stereotypical attitudes, through middle and then high school and 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 also choices of math and science studies in the high school matriculation program. Our main identification strategy relies on the conditional random assignments of students and teachers in classes within a school.[15] Using within school analysis (primary school fixed effect framework), we compare students that study in the same school but were randomly exposed to different teachers with perhaps different stereotypical attitudes.

We first test the randomness of class composition in our sample by a series of Pearson Chi-Square (χ^2) tests that check whether the student's characteristics and the class he is assigned to is statistically independent. Based on 37 elementary schools (with two or more classes) and eight characteristics (gender, ethnicity, number of siblings, and level of parents' education) we find that out of 296 p vales only 14 were equal or lower than 5 percent. This implies that for only 5% of the classes we cannot reject that there is non-random assignment. In addition, of the 37 middle schools in our sample, in only two schools the p value was equal or lower than 5%. We therefore conclude that in our sample of students and classes there is no evidence of systematic non-random formation of classrooms with respect to students' characteristics.[16] The implication of this evidence is that since all classes within a school are not different in terms of average students' ability or any unobserved characteristic, teachers assignment to classes are also unrelated to observed and unobserved students background. As a result it is safe to assume that teachers' stereotypical attitude is also not correlated with students' and classes' observable or unobservable characteristics.

[15] The randomness of class composition results from the fact that students' assignments in primary and middle school into class based on ability, family background or any other characteristics of the students are forbidden by law (Integration Law 1968) in Israel and this law is strictly enforced.

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

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

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

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

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

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

where AS_{icjt} is the average of the teachers' stereotypical attitude in all three subjects. The coefficient of interest in this case is β_1 which captures the average-subject effects of teacher's stereotypical attitude in all three subjects on the outcome in subject *j*.[17] The average-subject effect captures the overall stereotypical environment that students' are exposed to in primary school.

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

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

whether our measure of teachers' stereotypical attitude is correlated with students' predetermined characteristics.

4. Results: Effect of Teachers' Stereotypes

A. Main Results

Table 6 reports the estimated effect of teachers' stereotypical attitude on students' academic achievements based on estimating equations 1 and 2. The test scores in all three subjects (math, English, and Hebrew) are pooled together, and each regression has subject and year fixed effects. We present the estimates of the direct-subject effect of teachers' stereotypical attitude and of their cross-subject effect where both are included jointly in a regression.[18] The estimates based on the sample of boys are presented in columns 1-2 and based on the sample of girls in columns 3-4. In columns 5-6 we present the estimated coefficient of the average over the three subjects of the teacher's stereotypical attitude. Panel A shows results for the middle school GEMS test scores and Panel B shows results for the matriculation tests scores.

Short term effects

In panel A of Table 6, we report results from three different specifications. The simple OLS estimates (first row) are positive and marginally significant for boys for the direct-subject effect (column 1), the cross-subject effect (column 2) and the average-subject effect (column 5); for girls on the other hand, these estimates are considerably lower and are not significantly different from zero. Adding primary school fixed effects to the regressions (second row) do not changes the estimates for boys but it lowers the estimated standard errors and as a result the estimated effects are now significantly different from zero. The estimates for girls in this second specification are now all negative, but only the cross-subject effect and the average-subject effect estimates are significantly different from zero. Remarkably, adding students' characteristic leaves the estimates for boys and for girls unchanged, implying that pupil's characteristic are not correlated with the teacher's stereotypes measure once we control for primary school.

[18] Appendix Table A2 reports the estimates of the direct-subject effect of teachers' stereotypical attitude and of their cross-subject effect from two separat regressions. The estimates are very similar to that reported in table 6, suggesting that the direct-subject and cross-subject effects are not correlated.

The estimated effect of teachers' stereotypes on boys' outcomes are positive; This indicate that teachers' over-assessment of boys' test scores increases their achievements at a later age. The estimate of the direct-subject effect for boys is 0.116 (SE=0.058), the respective estimate of the cross-subject effect is 0.139 (SE=0.083) and the average-subject effect is 0.254 (SE=0.113). Calibrating the effect size, increasing a teacher stereotypical attitude in a specific subject from zero (no gender bias) to one (the maximal value observed in the sample), will increase boys' test score in that subject by 0.116 of a standard deviation. Increasing the average stereotypical measurements in the other two subjects, from zero to one, will similarly increase boys' test score in that subject by 0.139 of a standard deviation. If in this scenario, we change the exposure of a male student in all three subjects from no gender bias to the highest stereotypical attitude observed in the data, his test score in that subject will improve by 0.254 of a standard deviation.

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

Long term effects

In panel B of Table 6, we present evidence of teachers' gender biases on the test score in the high school matriculation exams in the three subjects. These exams are taken at end of 12th grade, more than 6 years since the end of 'exposure' to teachers' gender stereotypes. Similar to the pattern we found in Panel A, only few OLS estimates are significant, though the estimates for boys are positive, while for girls they are negative. Adding primary school fixed effects to the regressions, does not change the boys' estimates much, while lowering further girls estimates. However, the within school estimation lowers again the estimated standard errors which makes all estimates statistically significant. Adding student characteristics as additional controls in the regressions leaves again the estimated effect unchanged.

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

We turn next to discuss results from several robustness tests. We examine how sensitive are the treatment estimates to adding some classroom level controls. These controls include the proportion of boys in class and the difference between the boys' and girls' in class mean blind test scores.

Appendix Table A3 presents the estimates of the average-subject effect of teachers' stereotypical attitude on both 8th grade GEMS test scores (Panel A) and matriculation test scores (Panel B), separately for boys and girls. The test scores in all three subjects (math, English, and Hebrew) are pooled together, and each regression includes students' characteristics and year and school fixed effects. In the first row, of both Panels, the proportion of boys in class is added as a control; in the second row, the difference between boys' and girls' in class mean blind test scores is added as control; whereas in the third row both controls are added jointly to the regression. The estimated effects of the average stereotype measure on boys' and on girls' outcomes are very similar to the estimates of our preferred specification (last row of Table 6) in both panels. Adding controls to the regression leads to only minor changes in the estimates. We think that these robustness tests reduce the likelihood that our results are derived by selection bias.[19]

[19] Appendix Table A4 presents further evidence that the measure of teacher's stereotype attitude is a characteristic of the teacher and not of the classroom: the table presents the correlations between stereotypical attitudes of teachers by subjects of instruction. The estimates presented suggest that the correlation between math teachers' stereotype attitude and Hebrew teachers' stereotype attitude is the highest. This result reinforce our assumption regarding teachers' stereotypical attitude not being correlated with students' and classes' observable

B. Estimated Effects by Subject

Short term effects

In this section we present and discuss results of estimating the effect for each subject separately. In Table 7, we present evidence based on estimating a separate regression for each subject, using the specification of a regression that includes students' characteristics, year and school fixed effects. As before, we present the estimates of the direct-subject effect of teachers' stereotypical attitude and of their cross-subject effect from one joint regression, for boys and for girls separately. The last two columns show the estimated coefficient of the average effect of teachers' stereotypes in all three subjects from separate regressions for boys and girls. In Panel A the depended variable is 8^{th} grade GEMS test scores whereas in Panel B the depended variable is matriculation test scores.

In Panel A of Table 7, the estimates of the direct subject effect of teachers' stereotypes are relatively small and not significantly different from zero for both genders, except in mathematics. The estimated effect of math teachers' stereotypical attitude on boys' 8^{th} grade math test scores is the largest direct-subject effect and is positive and significant, 0.374 (SE=0.142). The estimated effect of math teachers' stereotypical attitude on girls' math test scores is also the largest direct-subject effect though it is not precisely measured, -0.135 (SE=0.143). This result suggests that the 8^{th} grade students' test scores in math are mainly affected by their math teachers' stereotypical attitude. In Hebrew and English[20], the cross-subject effects are larger than the direct effect on students' 8^{th} grade test scores for both genders. The average-subject effect is significant in three of the six estimates and marginally so in two others, indicating that the overall stereotype environment is positive and significant for boys in math and Hebrew; while the opposite is true for girls in English.

Long term effects

In Panel B of Table 7, we present the estimated effect on the matriculation score by subject. Focusing on the overall effect presented in columns 5-6, we note that the estimated effect of teachers' stereotypes on boys' test scores in all subjects are positive and significantly different from zero at the

---

or unobservable characteristics since according to Table 2, most math teachers instruct also Hebrew and no English teachers instruct the other two subjects (Hebrew/math).

[20] Note that the estimated effect of teachers' stereotypical attitude on English outcomes is based on a smaller sample (311 observations) because of missing GEMS test scores in English in eighth grade in 2006.

5 percent level of significance. The respective effect on girls is negative in all subjects and it is significantly different from zero at the 5 percent level of significance for the Hebrew and math test scores. These estimates indicate that increasing the average stereotypical bias against girls from zero to its maximal values of one generates a 0.46 standard deviation gap in math test scores in the matriculation exam in favor of boys (0.263+0.195). The respective magnitudes for Hebrew and English are equally large.

The evidence on the effects of teachers' stereotypical attitude on students' enrollment in science, math and English advance courses in high school (equivalent to honor classes in the US) are presented in Table 8 and Table 9. Table 8 presents the estimated effect of teachers' stereotypes on enrollment in such courses and Table 9 presents the estimated effect of teachers' stereotypes on the total number of matriculation credits a student gained in each of these advanced courses. We note that an advance class yields 5 matriculation credits and a basic class yields only 3 matriculation credits. In science we included computer science honor courses and physics honor courses. Both tables present evidence based on estimating a separate regression for each subject, using the specification that includes students' characteristics, year and school fixed effects. As in earlier tables, we present the estimates of the direct-subject effect of teachers' stereotypical attitude and of their cross-subject effect, for boys and for girls separately. In columns 6-7 we present the estimates for the overall exposure in all three subjects.

Table 8 presents estimates from linear probability regressions. We also estimated logit regressions and we present the marginal effects estimated from this model in Appendix Table A5. Since these marginal effects are very similar to the estimates obtained from the linear probability regressions, we focus our discussion here on the later estimates in Table 8. The estimated effect of math teachers' stereotypical attitude on enrollment in advance studies in math (4 or 5 credit units programs) is positive and significant for boys (0.093, SE=0.049) and negative and significant for girls (-0.073, SE=0.044). The respective estimates in English and Science are not precisely measured though they are positive for boys and negative for girls. The estimated average-subject-effects are positive and significant for boys in English and in math and they are negative but not different from zero for girls. In order to assess the magnitude of the effect for boys, assuming that a boy is moved from a neutral teachers' stereotypical environment to one with a boys' bias of one, this will increase the enrollment rate of boys in advanced math studies by 11 percent and in advanced English program by 5.8 percent.

Table 9 present the estimated effect of teachers' stereotypes on students' total number of matriculation credits gained in these study programs. The average-subject estimated effects on math credits are significant for both boys' and girls'. For boy's the estimated effect is also significant on English matriculation units. As before, we can simulate the impact of moving from a neutral teachers' stereotypical environment to one with a boys' bias of one. Such change will increase boys 'number of matriculation units in math by 0.338, and decrease girls' number of math units by 0.291.

The estimates of the direct-subject effect in math are of special interest because of the considerable gender gap in math achievements and its impact on future labor market outcomes.[21] Based on our evidence in Appendix Table A1, simulating a 0.07 decrease in the math teachers' stereotypes attitudes (see Table A1 in the appendix) will decrease boys' math achievements in middle school by 0.026 standard deviation, and as a result will also eliminate the positive gender gap in favor the boys in math achievements in middle school (0.024). Such change will also decrease the enrollment of boys in advanced math studies in high school by 0.7 percent and will increase enrollment of girls in high level math by 0.5. As a result, the gender gap in studying math at high level in high school will decline from 3 to 1.8. A more drastic decline in math teachers' stereotypes, say a decrease of one standard deviation of the math teachers stereotypical attitude (0.4), will reverse the gender gap in math achievements in middle school from a positive gap of 0.024 SD in favor the boys to a negative gap of 0.126 SD in favor of the girls. The same change will also change the gender gap in enrollment in advance math studies from 3 in favor of the boys, to 3.6 in favor of girls.

C. Heterogeneous Treatment Effects of Teachers' Stereotypes on Test Scores

In order to gain more insight on the effects of teachers' stereotypical attitude on students' academic success, we explore these heterogeneous effects across different dimensions. In Table 10 we present the estimated effect of the average over the three subjects of the teacher's stereotypical attitude

[21] See Murnane et al. (1995) and Paglin and Rufolo (1990).

on GEMS test scores[22] for boys and for girls separately, based on different stratifications of the full sample. We use the specification that includes students' characteristics, year and school fixed effects. The first part reports the heterogeneous treatment effects of teacher stereotypical attitude by mother's education level (whether mothers' years of schooling is above the median – 12 years); the second reports the heterogeneous treatment effects by the gap in parental education (high parental education gap is defined as a non-negative gap between the fathers' and the mothers' years of schooling); the third reports the heterogeneous treatment effects by ethnicity (whether grandparents' place of birth is Asia/Africa) and the last reports the heterogeneous treatment effects by the child birth order (whether the student is a firstborn child).

The estimates presented in panel A of Table 10 indicate that the overall stereotype environment is positive and significant only for boys with high parental education; while the opposite is true for girls with low parental education. The estimated average-subject effect for boys with high parental education is 0.492 (SE=0.193), while the estimated average-subject effect for girls with low parental education is -0.407 (SE=0.212). In contrast, the average-subject effects for boys with low parental education and for girls with high parental education are of similar signs but are not significantly different from zero. These results seem partly counterintuitive since one would expect that the overall stereotypical environment to which students are exposed to would be more influential among students with low parental education for both gender and not only for girls. Moreover, the relevant sociology and psychology literature suggest mothers' employment status is correlated with a more egalitarian gender role attitude[23], students of educated mothers should be less influenced by teachers' stereotypical attitudes. Therefore, we also consider the heterogeneous treatment effects of teachers' stereotypical attitude based on a slightly different stratification of the sample where we divide students based on the within family parental education gap. We postulate that a child from a

[22] Appendix Table A6 presents the estimates of the average-subject effect on matriculation test scores based on only two stratifications of the full sample (by parental education and by the gap in parental education), since the matriculation test scores that are available at the Ministry of Education lab could not be merged with the additional demographic information from the population registry (available only at National Insurance Institute lab). Comparing the estimates of Table 10 to the estimates of Table A6, reveals similar effects of teachers' stereotypical attitude on both GEMS and matriculation tests scores by parental education; though the effect of teachers' stereotypical attitude on matriculation test scores is more pronounced for girls with low parental education gap, rather than with high parental education gap as is the case according to GEMS tests scores.

[23] See for example Hoffman (1977) and Herzog et al (1983)

family with lower parental education gap could be less susceptible to the effect of external gender stereotypes in school.

The treatment effects of teachers' stereotypical attitudes by parental education gap panel B of Table 10. Indeed, the table indicates that the overall stereotype environment effect is more important for students with high parental education gap, in particular when the student's father has more education then the mothers. This is true for boys and girls. The estimated average-subject effect is positive and significant for boys with a high parental education gap (0.298, SE=0.158); while it is negative and significant for girls with a high parental education gap (-0.431, SE=0.168). On the other hand, the average-subject effect for students with low parental education gap is not significantly different from zero. These results suggest that students from more gender equality home environment are indeed less influence by their teachers' stereotypical attitudes.

Panel C in Table 10 presents the heterogeneous treatment effects by ethnicity (ethnic origin Asia-Africa versus all others which includes mainly Europe-North America origin). The estimated effects by ethnicity have a similar pattern to those by parental education. This is not surprising because the Asia-Africa ethnic group has much lower level of parental education in comparison to the North America/Europe ethnic group. The estimated average-subject effect is positive and significant boys of Asia-Africa ethnicity (0.289, SE=0.135); while the opposite is true for girls from Asia-Africa ethnicity (-0.556, SE=0.215).

Panel D of Table 10 reports estimated treatment effects by child birth order. This stratification is not common, but much has been argued about the impact of birth order on children personality and behavior, especially with regard to firstborn children who are said to a more sociable, dependent and conforming.[24] We therefore posit that a stereotype environment could affect children differently by their birth order. Interestingly, the results suggest that firstborn children of both sexes tend to be less influenced by teacher stereotypical attitude: the average-subject effect is significant only among non-firstborn children. The estimated effect on non-firstborn boys is positive and significant (0.313, SE=0.185)); whereas the estimated effect on non-firstborn girls is negative and significant (-0.417, SE=0.189)).

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

D. Teacher’s Characteristics and Teaching Practices

Using administrative data from NII we are able to explore further the measure of teachers' stereotypical attitude and its correlation with characteristics of teachers. In Table 11 we examine the correlations of teachers' stereotypical attitudes with teachers' characteristics based on a sub-sample of teachers that we can match to NII data. The estimates are from separate regression for each of the teachers’ characteristics that we have, using a simple OLS regression with year and subject fixed effects.

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

In Table 12 we test whether teachers’ over-assessment of boys according to our stereotype measurement is correlated with teaching practices that differ across gender. The information about the use of different teaching practices for boys and for girls in the classroom are based on students’ report of classroom environment and teachers’ attitudes (of all teachers) available from to five GEMS questionnaire items. In these questions, students are asked rank from 1 (strongly agree) to 6 (strongly disagree) the extent to which they agree with a series of statements. These are the items we use in this analyis: (1) “Teachers help students to learn their subjects of interest”; (2) “Sometimes teachers insult me”; (3) “There are good relations between teachers and students”; (4) “The relations between students and teachers are of mutual respect”; and (5) “I am satisfied in school”.

In Table 12 we report estimates of the effect of the overall classroom stereotype environment on each of these five behavioral outcomes, using regressions that include students' characteristics, and school and year fixed effects. We estimate these regressions for boys and girls separately. We note that the measure of the overall teachers’ stereotype capture only the behavior of only math, English and

Hebrew teachers while the survey questions guide the students to refer to all of their teachers. The estimates in Table 12 suggest teachers' biased attitude against girls lead girls to feel that they are getting less support from their teachers (-0.333, SE= 0.239), a bias against boys make them sense an insulting attitude from their teachers (-0.381, SE=0.262). The estimated effects on more general questions regarding good students-teachers relations and overall satisfaction in class are not significantly different from zero, for both genders.

5. Conclusions

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

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

The results we present suggest that teachers' over-assessment of boys in a specific subject has a positive and significant effect on boys overall achievements in that subject while having a significant negative effect on girls. We also provide evidence that suggest spillover effects across stereotypical attitudes of teachers of different subjects can also impact students' achievements in other subjects. These effects persist through middle school and high school and actually have dramatic implications for matriculation test scores and for enrollment at higher level of math and science courses in high school. Interestingly we find that teachers' stereotypes influence more students from families where student's father has more education then the mothers, as well as students who are youngest among their siblings; and girls with low parental schooling or from an Asia/Africa ethnicity.

We also find that boys favoritism among math and science teachers have an especially large and positive effect on boys math test score and on the likelihood of boys studying math and science at the highest level in high school; the respective effect on girls is negative and statistically significant. The estimates of the direct-subject effect in math are of special interest because of the considerable gender gap in math achievements and its impact on future labor market outcomes. Moreover, since this gap in math achievement partly results from teachers' discrimination of girls in mathematics, eliminating this discrimination will go a long way in reducing the math achievements gender gap, and it will also decrease the gender gap in enrollment in advance math studies.

6. References

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

Lavy, V., 2011, What Makes an Effective Teacher? Quasi-Experimental Evidence , *NBER Working Paper* 16885.

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.

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

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

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

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

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

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

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

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

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

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

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

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 Kristin Powell. Sex Differences in Test Performance: A Survey of Literature. No. 89. New York: College Entrance Examination Board, 1989

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	867	1127	1017
Number of Elementary Schools	17	20	20
Number of Elementary Classes	33	41	38
Number of Middle Schools	5	7	5
Number of Middle School Classes	26	32	31

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

Table 2: Summary Statistics of Teachers' Characteristics

Number of Identified Teachers	78
Number of Identified Teachers Instructing only Math	13
Number of Identified Teachers Instructing only Hebrew	29
Number of Identified Teachers Instructing both Hebrew and Math	36
Proportion of Teachers Older than 50 years old	0.257 (0.439)
Proportion of Teachers from Europe/America Ethnicity	0.471 (0.501)
Proportion of Married Teachers	0.681 (0.468)
Proportion of Single Teachers	0.115 (0.320)
Mean Number of Teachers' Offspring	2.354 (0.915)
Mean Proportion of Daughters among Teachers' Offspring	0.501 (0.329)
Proportion of Teachers with at Least One Daughter	0.834 (0.373)

Notes: Identified teachers are teachers who are home class teachers and teach at least one of the relevant subjects: math, Hebrew or English. Standard deviations are reported in parentheses.

Table 3: Means and Standard Deviations of National and School Primary School Exams Scores and Teachers' Stereotypes Measurement at the Student Level, by Gender

	Boy			Girl			
	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	Teachers' Stereotypes Measurement at the Student Level
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Hebrew	-0.099	-0.136	0.037	0.168	0.139	0.029	0.005
	(1.001)	(1.022)	(1.047)	(0.951)	(0.952)	(0.925)	(0.989)
Math	0.052	-0.014	0.066	0.003	0.014	-0.011	0.038
	(0.985)	(1.034)	(0.960)	(0.971)	(0.963)	(0.903)	(0.933)
English	-0.036	-0.047	0.010	0.113	0.049	0.064	-0.026
	(1.002)	(1.036)	(0.999)	(0.940)	(0.962)	(0.933)	(0.968)
Number of Students	4246	4246	4246	4122	4122	4122	8368

Notes: The national exams scores and the school scores are standardized scores. The number of students referees to the number of students in all three subjects. The teacher stereotype measurement at the student level is equal to the difference between boys' school and national exams scores less the difference between girls' school and national exams scores. Standard errors are reported in parentheses.

Table 4: Means and Standard Deviations of National Exams Scores in Middle School and High School at the Student Level, by Gender

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

Notes: The national exams scores are standardized scores. The number of students referees to the number of students in all three subjects. Matriculation test scores are weighted according to units of studies for each subject, as computed by the Ministry of Education. Standard errors are reported in parentheses.

Table 5: Number of Matriculation Exams Units and Proportion of Students in each Unit, by Gender

		Boy				Girl			
		(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Hebrew	Number of Units	2			Total	2			Total
	Proportion of Students	86.9			86.9	91.1			91.1
Math	Number of Units	3	4	5	Total	3	4	5	Total
	Proportion of Students	34.1	25.7	21.1	80.9	40.5	29.7	14.1	84.3
English	Number of Units	3	4	5	Total	3	4	5	Total
	Proportion of Students	10.4	17.3	60.5	88.2	11.7	21.4	58.2	91.3
Physics	Number of Units	1	3	5	Total	1	3	5	Total
	Proportion of Students	6.0	0.5	15.1	21.6	0.0	3.3	4.8	8.1
Computer Science	Number of Units	1	3	5	Total	1	3	5	Total
	Proportion of Students	0.6	1.1	11.3	13.0	0.4	0.9	3.2	4.5

Notes: Proportion of students referees to the proportion of students in each matriculation exam unit from the total number of students in the sample (percent), by gender.

Table 6: Estimated Effect of Teachers' Stereotypes on Test Scores in Math, English, and Hebrew

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

Notes: The test scores in all three subjects (math, English, and Hebrew) are pooled together. These test scores are standardized scores, by year and subject. Panel A shows results of the estimated effect of teachers' stereotypical attitude on 8th grade GEMS test scores and Panel B shows results of the estimated effect of teachers' stereotypical attitude on Matriculation test scores (matriculation test scores are weighted according to units of studies for each subject, as computed by the Ministry of Education). The first specification is a simple OLS regression with subject and year fixed effects; The second specification includes also elementary school fixed effects; The third specification includes additionally students' characteristics (gender, parental education, number of siblings, and dummies for four ethnicity groups). The direct-subject effect is the effect of teacher stereotype in a specific subject on the test scores in the same subject; the cross-subject effect is the impact of the average teacher stereotypes in the other subjects on the test scores in the referred subject. Average-subject effect is the impact of the average teacher stereotypes in all subjects. The estimates in each row in columns 1-2 are each from a joint regression and so are the estimates in columns 3-4. The estimates in each row in columns 5-6 are each from separate regressions. Standard errors are clustered by class and are reported in parentheses.

Table 7: Estimated Effect of Teachers' Stereotypes on Test Scores, by Subject

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

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

Table 8: Estimated Effect of Teachers' Stereotypes on Students' Enrollment in Science, Math and English Advance Study Program in High School

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

Notes: See Table 7. Each row present estimates from separate linear probability regression for each subject (English /Math/Math oriented subjects). The dependent variables are discrete and equals one if the number of matriculation credit units exceed a certain level. Each regression includes students' characteristics, elementary school and year fixed effect. The estimates in each row in columns 1-2 are each from a joint regression and so are the estimates in columns 3-4. The estimates in each row in columns 5-6 are each from separate regressions. Standard errors are clustered by class and are reported in parentheses.

Table 9: Estimated Effect of Teachers' Stereotypes on Students' Total Number of Matriculation Credit Units in Science, Math and English Advance Study Program in High School

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

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

Table 10: Estimated Effect of Average-Subject Stereotypes on 8th Grade GEMS Test Scores, by sub-groups

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

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

Table 11: Correlation of Teachers' Stereotypes with Characteristics of Teachers

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

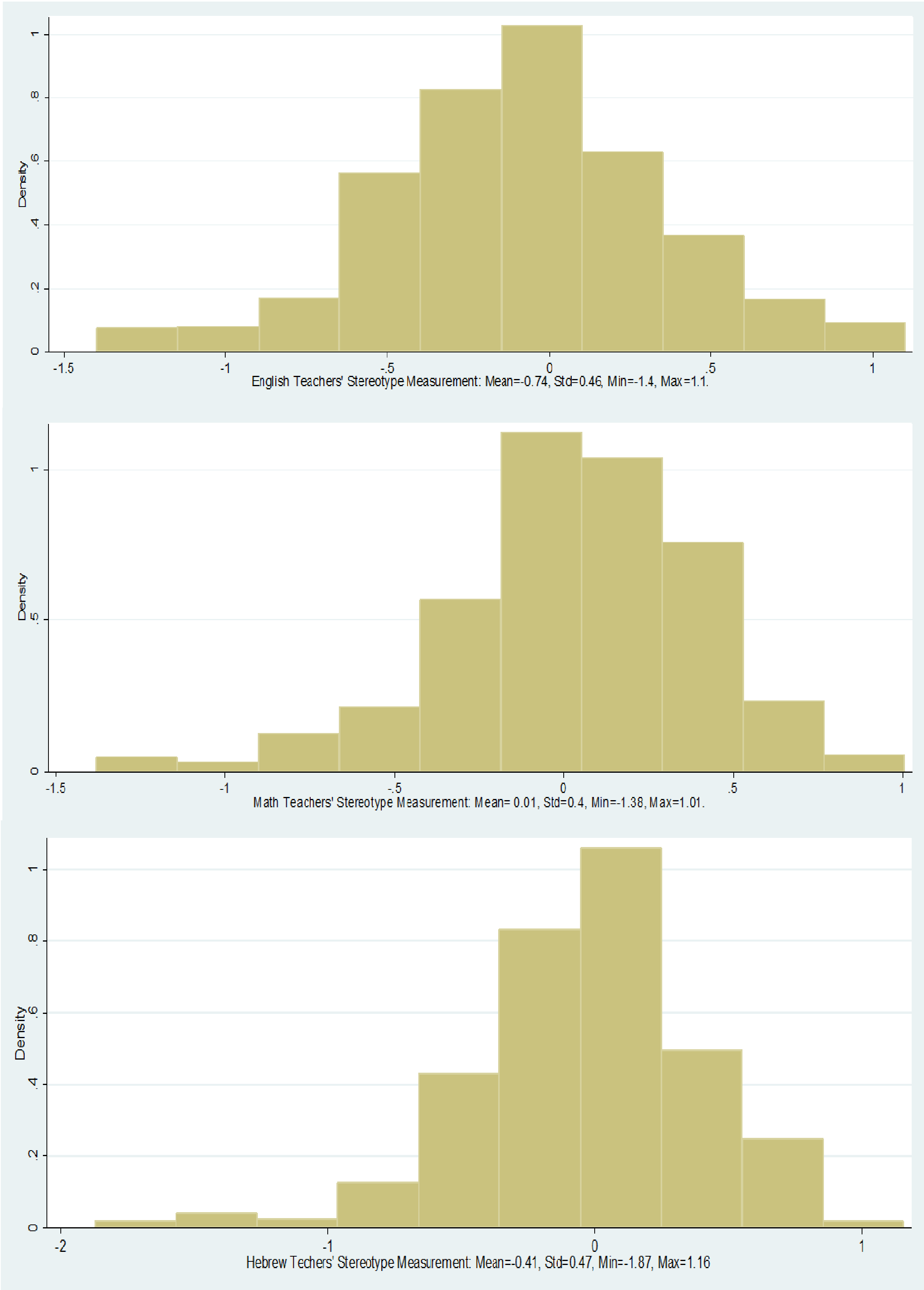
Notes: The specification is a simple OLS regression with subject and year fixed effects. The estimates in each row in columns 1-6 are each from a separated regression. Standard errors are clustered by class and are reported in parentheses.

Table 12: Estimated Effect of Average Teachers' Stereotypes on 5th Grade Behavioral Outcomes

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

Notes: Each regression includes students' characteristics, school and year fixed effects. The estimates in each row in columns 1-2 are each from a separated regression. Standard errors are clustered by class and are reported in parentheses.

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



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

Table A1: Estimated Gender Stereotypes by Subject at the Student Level

	Total		Hebrew		Math		English	
	OLS	Student Fixed Effect	OLS	Student Fixed Effect	OLS	Student Fixed Effect	OLS	Student Fixed Effect
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Male	-0.145		-0.288		-0.018		-0.125	
	(0.029)		(0.034)		(0.037)		(0.041)	
Non-blind score	0.004	0.017	0.006	0.029	-0.024	-0.011	0.035	0.064
	(0.048)	(0.049)	(0.050)	(0.073)	(0.046)	(0.067)	(0.056)	(0.078)
Male x (non-blind score)	0.007	0.010	0.005	0.009	0.064	0.076	-0.051	-0.054
	(0.031)	(0.033)	(0.042)	(0.059)	(0.039)	(0.053)	(0.046)	(0.062)
Number of Students	16428	17395	5479	5806	5488	5812	5461	5777

Notes: Dependent variables are standardized scores. The number of observations is twice the number of exam takers, since the datasets are stacked. The OLS regressions includes in addition to the regressors' presented in the table also pupil's characteristics (gender, parental education, number of siblings, and dummies for four ethnicity groups), year and subject dummies and class fixed effects. The Student Fixed Effect regression includes in addition to the regressors' presented in the table also year and subject dummies and student fixed effects. Standard errors are corrected for class level clustering and are presented in parentheses.

Table A2: Estimated Effect of Teachers' Stereotypes on 8th Grade Test Scores in Math, English, and Hebrew

	Boy		Girl	
	Direct-Subject Effect	Cross-Subject Effect	Direct-Subject Effect	Cross-Subject Effect
	(1)	(2)	(3)	(4)
OLS	0.130	0.140	0.027	-0.008
	0.099	0.150	0.117	0.174
6th Grade School Fixed Effects	0.112	0.097	-0.049	-0.242
	(0.057)	(0.087)	(0.083)	(0.100)
6th Grade School Fixed Effects and Student Characteristics	0.107	0.122	-0.058	-0.235
	(0.058)	(0.087)	(0.081)	(0.094)
Number of Students	1420	1317	1420	1317

Notes: See table 6. The estimates in each row in columns 1-4 are each from a separate regression. Standard errors are clustered by class and are reported in parentheses.

Table A3: Estimated Effect of Teachers' Stereotypes on Test Scores in Math, English, and Hebrew from Alternative Specifications

	Boy Average-Subject Effect (5)	Girl Average-Subject Effect (6)
A. 8th Grade GEMS Test Scores		
Proportion of Boys	0.226	-0.263
	(0.112)	(0.118)
Difference Between Boys' Grades and Girls' Grades in Class	0.274	0.000
	(0.117)	(0.000)
Both Controls	0.253	-0.284
	(0.110)	(0.121)
Number of Students	1420	1317
B. Matriculation Test Scores		
Proportion of Boys	0.248	-0.137
	(0.083)	(0.089)
Difference Between Boys' Grades and Girls' Grades in Class	0.209	-0.132
	(0.077)	(0.075)
Both Controls	0.213	-0.094
	0.087	0.074
Number of Students	3883	4033

Notes: See Table 6. Each regression includes students' characteristics, primary school, year and subject fixed effect. The first regression include as control, the proportion of boys in primary school class; the second regression include as control the differences between boys' grade to girls' grades in 5th grade national exams; and the third regression include both controls. Standard errors are clustered by class and are reported in parentheses.

Table A4 : Correlation Matrices of Teachers' Stereotypes Measurement Across Subjects

	Overall			Within School		
	Teachers' Stereotypes in Hebrew	Teachers' Stereotypes in Math	Teachers' Stereotypes in English	Teachers' Stereotypes in Hebrew	Teachers' Stereotypes in Math	Teachers' Stereotypes in English
	(1)	(2)	(3)	(4)	(5)	(6)
Teachers' Stereotypes in Math	0.508			0.315		
Teachers' Stereotypes in English	0.287	0.311		0.077	0.180	

Notes: The correlations in each row in columns 1-3 are the correlation between teachers' stereotypes indices across subject, from the overall sample. The correlations in each row in columns 4-6 are the correlation between teachers' stereotypes indices across subject, where the school means of teachers stereotypes in each subject are netted out.

Table A5: Estimated Effect of Teachers' Stereotypes on Students' Enrollment in Science, Math and English Advance Study Program in High School

		Boy		Girl		Boy	Girl
		Direct-Subject Effect	Cross-Subject Effect	Direct-Subject Effect	Cross-Subject Effect	Average-Subject Effect	Average-Subject Effect
		(1)	(2)	(3)	(4)	(5)	(6)
English (dummy=1 if # units=5|4)	Log of Odds Ratio	0.185	0.166	-0.114	-0.390	0.366	-0.497
		(0.132)	(0.220)	(0.195)	(0.285)	(0.249)	(0.280)
	Marginal Effect at the Mean	0.042	0.038	-0.028	-0.095	0.080	-0.120
Math (dummy=1 if # units=5|4)	Log of Odds Ratio	0.535	0.131	-0.324	-0.033	0.669	-0.373
		(0.244)	(0.284)	(0.211)	(0.284)	(0.320)	(0.278)
	Marginal Effect at the Mean	0.118	0.027	-0.024	-0.003	0.147	-0.028
Physics/Computer Science (dummy=1 if units=5)	Log of Odds Ratio	0.136	0.040	0.402	-0.080	0.181	0.363
		(0.313)	(0.404)	(0.461)	(0.586)	(0.329)	(0.499)
	Marginal Effect at the Mean	0.026	0.007	0.020	-0.003	0.035	0.017

Notes: See Table 7. Each row present log of odds ratios and the marginal effects at the means from separate logistic regression for each subject (English /Math/Math oriented subjects). The dependent variables are discrete and equals one if the number of matriculation credit units exceed a certain level. Each regression includes students' characteristics, elementary school and year fixed effect. The estimates in each row in columns 1-2 are each from a joint regression and so are the estimates in columns 3-4. The estimates in each row in columns 5-6 are each from separate regressions. Standard errors are clustered by class and are reported in parentheses.

Table A6: Estimated Effect of Average-Subject Stereotypes on Matriculation Exam Scores, by sub-groups

	Boy	Girl	Boy	Girl
	(1)	(2)	(3)	(4)
	Low Parental Education		**High Parental Education**	
Mother's Education Level	0.206	-0.348	0.216	0.053
	(0.123)	(0.120)	(0.094)	(0.114)
Number of Students	1943	2064	1761	1718
	Low Parental Education Gap		**High Parental Education Gap**	
Parental Education Gap (Father's Education Less Mothers' Education)	0.153	-0.369	0.211	-0.138
	(0.188)	(0.172)	(0.084)	(0.100)
Number of Students	1140	1112	2564	2670

Notes: See Table 10. The Table presents the estimated effect of teachers' stereotypical attitude on Matriculation test scores. Standard errors are clustered by class and are reported in parentheses.