Teacher assessments versus standardized tests: is acting "girly" an advantage?*

Adriana Di Liberto
University of Cagliari,
IZA and CRENoS

Laura Casula
University of Cagliari
and CRENoS

May 30, 2016

Abstract

We study if Italian teachers do apply gender discrimination when judging students. To this aim, we use a difference-in-difference approach that exploits the availability of both blind and non-blind scores in math and language that Italian students receive during the school year. Using data for all sixth graders, descriptives show that in both scores girls are better than boys in the language scores, while in math boys perform better than girls in the blind test. Moreover, our analysis suggest that boys are always discriminated by teachers in both subjects. This result holds also when we control for students noncognitive skills, gender specific-attitude towards cheating, possible cultural differences towards gender attitudes in math or language and the presence of family role models.

Keywords: Gender stereotypes, Discrimination, Schooling outcomes.
J.E.L. Classification: L2, I2, M1, O32.

PRELIMINARY DRAFT, PLEASE DO NOT CITE

*Contacts: Adriana Di Liberto, diliberto@unica.it, Dipartimento di Scienze Economiche e Aziendali, University of Cagliari, via S. Ignazio 17, 09123, Cagliari, ITALY.

“”

.

1 Introduction

Are school teachers influenced by students gender when evaluating them at the exams? And, if so, in what way? Do they discriminate against a specific gender and also on specific subjects? These questions are of a great interests as teachers stereotypical perceptions when assessing academic results may have long lasting consequences in students school performance and, through this, on their following labor market outcomes.

The literature suggests the presence of different channels that link gender stereotypical perceptions when evaluating students with their economic and social outcomes. First, evidence shows that the highest performing education systems are those that combine educational quality with equity, and teachers gender biased (mis)judgment may affect many educational outcomes. In particular, dropping out is considered the result of a complex process of student disengagement and teachers gender discrimination is detrimental for misjudged students (OECD, 2012; Lyche, 2010). Data also suggest that there exist a significant difference between boys and girls in educational attainment, with boys more likely to repeat school years than girls and being predominate among early school leavers (Eurydice, 2010). If these results were driven by the presence of teachers gender biased evaluations we should find that at school boys are discriminated against girls.

Second, teachers gender stereotypes could be also differentiated by subject, such as “boys are good in math and science, while girls in literature and poetry” (Lavy and Sands, 2015). This kind of teacher’s stereotypes could cause lower/higher grades for girls in math/language and it would result in a misallocation of talents and skills: science-gifted women would invest more in “girly” studies that are also less profitable in terms of labor market outcomes, while the opposite would be true for men. Overall, if this teachers stereotypes behavior represents the rule rather than an exception in an educational system, misallocation processes may significantly affect a nation labor force productivity and it may harm its growth perspectives. Again, international data show that women with the same educational attainment as men are under-represented in many scientific and technical degrees, which typically lead to better paid occupations. Recent evidence also suggests

that this gap is narrowing quickly in nations that pursue gender equality policies (Campbell et al., 1999, Machin and Pekkarinen 2008, Guiso et al. 2008).

In this study we focus on the Italian school system. With respect to other industrialized countries, the Italian educational system performs poorly: its mean performance at OECD-PISA tests in all subjects is below the OECD average. Moreover, boys outperform girls in mathematics by an average of 18 points, and this gap has remained stable since 2003. Conversely, girls outperform boys in reading by an average of 39 score points. The gender gap observed in both subjects is similar to that observed across OECD countries. Finally, considering school dropouts, as in most industrialized countries, girls outperform boys in Italian schools, with dropout rates among boys (17,7%) significantly higher than that observed for girls (12,2%).[1]

In order to test for the existence of gender stereotyping and discrimination by Italian teachers, we follow Lavy (2008) and use a difference-in-difference approach that exploit the presence in the dataset of both blind and non blind results in two different subjects, math and language. In other words, we assume that the blind score may be used as the counterfactual measure to the non-blind score, which may be influenced by the teachers discrimination.[2] Data are provided by the INVALSI, the Italian institute in charge of evaluating school performance, and include information on both blind (standardized tests carried out by INVALSI) and non blind (teachers evaluation) students results, together with many additional information on students and school characteristics.

The INVALSI standardized tests are compulsory for all Italian schools and students, both public and private, attending specific grades of schooling. Unlike other studies, this enables us to exploit within country data on all students during the schooling year 2010-11 and for different grades. Moreover, the INVALSI dataset provides a very rich set of information on student characteristics that includes not only a full set of demographics, but also information on noncognitive skills, such as students attitude towards learning by subject.

Our main result supports the evidence found in other studies and suggest that teachers discrimination act against male students.[3] This result is also robust to different model specifications.

[1] Eurostat, LFS 2014).

[2] On the use of the systematic difference between blind and non-blind tests across groups as a method to underline discrimination see also the work of Blank (1991) and Goldin and Rouse (2000).

[3] Together with Lavy (2008), see also Bjorn et al. (2011), Hanna and Linden (2012), Cornwell et al. (2013).

First, we control for noncognitive skills using different measures of self-assessed ability and propensity for studying Math or Language. Cornwell et al. (2013) suggest that excluding these skills from the analysis would produce biased results: they that the teachers gender discrimination vanishes when noncognitive skills are taken into account.

Second, we also perform the regression analysis for the subset of classes where external inspectors invigilate students during the blind standardized test. In fact, cheating is a well-known phenomenon during Italian schooling exams, and girls may have a different attitude towards cheating than boys. Thus, using this subsample we are confident a) that cheating is not an issue and b) since all the steps of the INVALSI testing protocol has been fulfilled, that the blind score is likely to be free of any bias that might be caused by teachers attitude.

Finally, we exploit a specific feature of our Italian sample. In fact, unlike most within-country data sets there exists a deep, persistent duality in Italy between the developed North-Center and the less developed South. This substantial geographical heterogeneity is also present in both education and gender roles. For the former, both quantitative (educational attainments) and qualitative (cognitive skill tests results) educational outcomes stress a large gap between the two areas, and the same North-South gap is observed for gender roles. As suggested by many labor market outcome indicators, women's traditional role of wife and mother is more persistent in the South.[4] Thus, in order to take into account for cultural factors and gender roles that may differently affect the choice of how much to invest in studying specific subjects, we have also performed the analysis separately for the Northern and Southern regions of the country.

Overall, all our robustness checks confirm that boys who perform equally as well as girls on language and math blind tests are graded less favorably by their teachers.

2 Data and descriptives

We constructed a database with rich information on student, school and area characteristics. Our main source of data is the database provided by the National Institute for the Evaluation of the Educational System of Instruction and Training (INVALSI henceforth), a government agency that carries out a yearly evaluation of student attainment in both Mathematics and Language. The

[4] On this see, for example, Del Boca (2005).

INVALSI standardized tests are compulsory for all Italian schools and students, both public and private, attending specific grades of schooling. In our analysis we focus on the 2010-11 school-year data for sixth grade lower secondary school students.[5] INVALSI enforces a protocol for the administration of the tests to reduce discretion and the possibility of teachers manipulations (Invalsi, 2011). First, the type of tasks that students have to complete include multiple choice and closed-format short answer questions. Second, the test is not administered by the class teachers but by other teachers of the school, who in general teach a different subject from the one that is being tested. [6]

The INVALSI questionnaire is designed in order to collect detailed information about the student background and family characteristics.[7] In our analysis we include the following additional students demographic information: gender, citizenship (native, first and second generation immigrant students), if she/he speaks a foreign language at home or an Italian dialect, her/his socio-economic background using the number of books at home, the number of siblings, and parental education.[8]

Second, the set of school characteristics includes the number of students per class and school, the proportion of female students per class, and the school-average ESCS index. The latter is calculated based on the parental occupational status, their educational attainment levels and different measures of household possessions including cultural possessions such as home educational resources and the number of books. This ESCS index for student socioeconomic background is analogous to the same one computed by OECD for the PISA test. The individual scores of this index are obtained by a principal component analysis, with normalized zero mean and unit standard

[5]Tests are carried out also by students attending the second and fifth grade (in primary schools), the sixth and eighth grade (in lower secondary) and the tenth grade (in upper secondary). The Italian school system starts at age six with five years of primary school (grades 1 to 5) followed by three years of lower secondary school (grades 6 to 8). Upper secondary education lasts three to five years depending on the type of school chosen.

[6]Moreover, all the school teachers are simultaneously involved in the transcription process, so that they cross-check each other, while the school principal who is responsible for the correct implementation of the protocol supervises the whole process. For more on this see also Lucifora and Tonello, 2015)

[7]Information is collected through a "Family Questionnaire" sent to each family before the test, a "Student Questionnaire" filled by each student the first day of the test and, finally a student general information part compiled from school administrative staff.

[8]First generation are students born abroad of foreign-born parents, while second generation students are native-born children of foreign-born parents. In using the variable number of books at home we follow Hanushek and Woessmann (2011) who argue that this is the best single predictor of students performance.

deviation.[9]

Third, we include different area characteristics: we control for macro-area dummies, plus we include a proxy for the wealth level of the school catchment area (per capita value added), a measure of the level of criminality, and a social capital indicator.[10] In fact, previous studies show that geographical location is an important determinant of Italian students test scores, with students in the Northern area usually outperforming those living in the South and differences in both economic and cultural factors may play a role.

Fourth, together with the INVALSI test results, the dataset also include a measure of the score in both language and math assigned by teachers during the first term. Given the way the test is implemented, the Invalsi tests may be considered "blind" assessments while, in contrast, the score assigned by student's teacher is clearly a "non-blind" assessment.[11] The complete list of variables is reported in Appendix A and Table 1 sums up the major characteristics of the variables used in regressions for our overall sample.

Since the INVALSI and teachers' votes are expressed in different scales, in order to compare the two set of students results (blind and non-blind) we firstly convert the scores to the same scale and we secondly we calculate the z-scores: that is, we standardize them to a distribution with zero mean and unit standard deviation. Figures 1 to 4 show the kernel-density distribution of the two types of scores (blind/INVALSI tests vs non blind/teachers) by gender and subject. For language (Fig.1 and Fig. 2) we observe a righward shift of the teacher-score distribution relative to the INVALSI-score distribution for both boys and girls. The opposite is true for girls in math (Fig. 3), while the distribution of scores in math for boys is very similar (Fig. 4).

Table 2 includes the non-standardized average results by gender achieved in both the blind and non-blind test, while Table 3 ALSO includes a measure of the teachers bias at student level by gender.[12] Table 3 shows that on average boys outperform girls in math blind test scores but girls outperform boys on teacher grades. For language the picture is different: girls are always better

[9]They are the scores for the first principal component. The index is calculated considering the whole sample of sixth grade lower secondary school Italian students. See also Invalsi (2011) for details.

[10]We identify the following dummy variables: North-East, North-West, Centre, South, South-Islands.

[11]It is difficult to find a comparison of blind and non-blind grading of the exact same tests. One example is in Hinnerich et al. (2010).

[12]In Table 2 we compare the non-standardized scores, with both blind and non-blind test results rescaled from 0 to 100.

than boys in both types of tests. Moreover, compared to the blind test, teachers evaluations are still lower for boys. Overall, comparing the results in the two subjects, these numbers suggest that teachers discriminate boys in both subjects, and that the gender discrimination gap is higher for math.

Finally, data on noncognitive skills are reported in Table 4. The INVALSI questionnaire includes several indicators related to students drive and motivation in studying a specific subject. In particular, it asks different questions designed to measure the self-assessment of boys and girls about their ability in Math (Q3) and Language (Q5) studies. During the survey, Italian students are asked to indicate how much they agree with five different statements about mathematics and language studies.[13] The specific questions asked and the results by gender all show that the subject specific propensity for learning and achieving is very different between boys and girls. The former are more confident and enjoy more studying math, while girls are more confident in language studies.

3 Results

Following Lavy (2008), we use the data pooled over the two types of scores, one blind and the other non-blind, in the two subjects (Math and Language) and use a difference-in-difference regression setting of the form:

$$y_{ijb} = \alpha + \beta Male_i + \gamma NB_{ijb} + \delta(Male_i \times NB_{ijb}) + \upsilon_{ijb} \tag{3.1}$$

where y_{ijb} is an indicator of performance of student i attending school j for both blind and non-blind scores b, $Male_j$ is the gender dummy (equal to one if male), and NB is the dummy identifying the teachers (non-blind) scoring procedure. Thus, the intercept is the average score obtained by female students on blind tests, β captures the score difference of male students in both types of tests, and γ measures the teachers effect, that is, the average differences in scores due to the type of tests. The parameter of interest is on the interaction term, δ, that measures the difference in scores obtained by male students due to teachers. Given the INVALSI testing protocol, we may

[13] INVALSI uses the following scale: 1-moderately disagree, 2-moderately disagree, 3-somewhat agree, 4 strongly agree.

assume that the blind score is free of any bias that might be caused by stereotyped discrimination. Conversely, the non-blind score may possibly reflect biases teachers' gender stereotypes.

Table 5 and 6 show the results for the Language and Math scores respectively. For Language (Table 5), model 1 include the results of eq. 1, our most parsimonious specification. We find that all coefficient are significant. Moreover, we also find that on average, female students perform better than boys, the mean difference between the teachers scores and the INVALSI scores is positive and significant, while our parameter of interest, the coefficient of the interaction term, is negative suggesting that teachers discrimination may act against male students.

In the following models, we exploit a rich set of variables that control for student characteristics, including self-assessed ability and propensity for studying both Math and Language, and for school and area characteristics. Model 2 introduce different variables that control for students demographics, while in Model 3 we increase the specification with more family characteristics in order to take into account for the student's socioeconomic background. Model 4 includes the school average socio-economic background (calculated by the ESCS index), the school size and the proportion of girls in each class. The latter variable should control for gender peer effects and it has been found to be an important determinant in these analysis (Lavy et al., 2011).[14] In model 5 we also introduce two different dummy variables that should capture students noncognitive abilities: the dummies "good in math/language" are equal to one for students that show a strong propensity for studying both Math and Language.[15] Finally in model 6 we also control for area characteristics, including total value added per capita in 2001 that represents a standard proxy of an area economic performance, the rate of extortions over 1000 inhabitants, and a measure of social capital.[16] All these additional indicators should capture cultural features that may differently affect boys and girls students outcomes.

Overall, the estimated coefficients on our additional controls all show the expected signs and still confirm that boys who perform equally as well as girls on language and math blind tests are graded less favorably by their teachers.

[14]Lavy et al., 2011, find that an increase in the proportion of girls improves boys and girls cognitive outcomes.

[15]Their answer in Q3A, Q3C, Q3D, Q3E, Q5A, Q5C, Q5D and Q5E is strongly agree, while they strongly disagree in Q3B and Q5B. See Table 4.

[16]To this aim we use a synthetic social capital index at regional NUTS3 level, provided by Cartocci (2007), which merges data on 1) blood donations, 2) sport participation, 3) dissemination of newspaper and 4) voter turnout.

4 Robustness checks

In this section we perform a set of robustness checks of the basic results discussed above.

First, our dataset enable to identify a representative and random sample of monitored classrooms where external inspectors invigilate students during the test and also help to both compute results and prepare the documentation relative to the test. Existing evidence shows that students in the non-monitored classroom received a more benevolent supervision, allowing student cheating behavior more easily (Lucifora and Tonello, 2015). Thus, even if, due to the INVALSI test protocol, we consider this hypothesis implausible, it is possible that teachers in non monitored classes, even during the blind test, may manipulate the scores and discriminate by gender. Moreover, it is also possible that the attitude towards cheating is different by gender. In order to control for these hypothesis, in Panel A of Tables 7 and 8 we show the results when we replicate the analysis for the sub-sample of classes with the presence of an inspector.

Second, unlike most studies on teachers discrimination, we are able to control for different measures of self-assessed ability and propensity for studying both Math and Language. We have seen above in Table 4 the different *subject specific* propensity for achievement: boys indicate a more positive attitude than girls in studying math, while the opposite is true for language studies. We cannot investigate what cause the observed differences between genders, but we replicate our analysis splitting the sample between (self-assessed) high (Panel B results) and low achieving (Panel C results) students. This improves our identification because it allows us to get rid of some of the differences in noncognitive abilities that have been found to be an important determinant in this analysis (Cornwell et al., 2013).

Third, previous studies show that geographical location is an important determinant of Italian student test scores (Cipollone et al., 2010) and that there exists a deep, persistent duality in Italy between the developed North-Center and the less developed South also in terms of culture and gender roles.[17] Therefore, in Panel D and E in replicate the analysis for the subsample of northern and southern regions respectively.

All our robustness checks confirm the result that male students face discrimination in each

[17] On this, see Del Boca (2005) and Di Liberto and Sideri (2015).

subject.

Further issues will be soon investigated. First, we will check if family role models play a role in our analysis. To this aim we will examine if results are different a) for girls who are lacking positive female role at home dividing the sample of girls by mother's education and b) for boys diving the sample for fathers education. Second, our results on lower secondary students will be compared with those obtained using the INVALSI primary school dataset. This will enable us to check if the presence of teachers discrimination is also affecting different schooling levels.

5 Conclusions

To be completed.

References

[1] Bjorn, T.H., Hoglin, E., Johannesson, E. (2011), Are boys discriminated in Swedish high schools? *Economics of Education Review*, Vol. 30(4), 682-690.

[2] Blank, R. M. (1991), The Effects of Double-Blind versus Single-Blind Reviewing: Experimental Evidence from The American Economic Review *American Economic Review*, Vol. 81(5), 1041-1067.

[3] Cartocci, R. (2007), Mappe del tesoro. Atlante del capitale sociale in Italia. Il Mulino, Bologna.

[4] Cipollone, P., Montanaro, P. and Sestito, P. (2010), Value-Added Measures in Italian High Schools: Problems and Findings, *Giornale degli Economisti* 69 (2):81114.

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

[6] Del Boca D. (2005), Editorial Foreword, *Labour*, vol. 19(s1), pp. 1-4.

[7] Di Liberto, A. and M. Sideri (2015), Past dominations, current institutions and the Italian regional economic performance, *European Journal of Political Economy* 38, 12-41

[8] Eurydice (2010), Gender Differences in Educational Outcomes: Study on the Measures Taken and the Current Situation in Europe, Brussels.

[9] Eurostat, Labour force survey 2014.

[10] Goldin, C., Rouse C. (2000), Orchestrating Impartiality: The Impact of "Blind" Auditions on Female Musicians. *American Economic Review*, Vol. 90(4), 715-741

[11] Guiso, L., Monte F., Sapienza, P. and Zingales L. (2008), Culture, Gender, and Math, *Science* 320: 11641165.

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

[13] Hanushek, E.A., Woessmann L. (2011), The Economics of International Differences in Educational Achievement, in: Hanushek E.A., Machin S., Woessmann L. (eds.) *Handbook of the Economics of Education*, Vol. 3, Amsterdam: North Holland.

[14] Hinnerich, B. T., Hoglin, E., Johannesson, M., (2011), Are boys discriminated in Swedish high schools? *Economics of Education Review*, Vol. 30(2011), 682-690.

[15] INVALSI (2011), Rapporto tecnico sulle caratteristiche delle prove INVALSI 2011, *Technical Report, INVALSI.*

[16] Lavy, V. (2008), Do gender stereotypes reduce girls' or boys' human capital outcomes? Evidence from a natural experiment, *Journal of Public Economics*, Vol. 92, 2083-2105.

[17] Lavy, V., Schlosser A. (2011), Mechanisms and Impacts of Gender Peer Effects at School, *American Economic Journal: Applied Economics*, 3: 133.

[18] Lavy, V., Sand E. (2015), On The Origins of Gender Human Capital Gaps: Short and Long Term Consequences of Teachers Stereotypical Biases, NBER Working Paper No. 20909.

[19] Lyche, C. (2010), Taking on the Completion Challenge: A Literature Review on Policies to Prevent Dropout and Early School Leaving, OECD Education Working Papers, No.53, OECD, Paris.

[20] Lucifora, C. and Tonello M. (2015), Cheating and social interactions. Evidence from a randomizedexperiment in a national evaluation program, *Journal of Economic Behavior and Organization*, 115, 4566.

[21] Machin S., Pekkarinen T. (2008), Global Sex Differences in Test Score Variability, *Science*, Volume 322(5906): 1331-2.

[22] OECD (2012), Equity and Quality in Education: Supporting Disadvantaged Students and Schools, OECD Publishing.

[23] OECD (2014),PISA 2012 Results: What Students Know and Can Do, OECD Publishing.

[24] Paredes, V., (2012), A teacher like me or a student like me? Role model versus teacher bias effect *Economics of Education Review*, Vol. 39(2014), 38-49.

A Data sources

Description of Variables:

Dependent Variables:

- **Language_test**: INVALSI language test scores
- **Math_test**: INVALSI Math test scores
- **Language_Teacher**: Teacher's language scores
- **Math_Teacher**: Teacher's Math scores

Student and family characteristics:

- **Males**: dummy=1 if male
- **Q3_dummy**: dummy=1 if good in math
- **Q5_dummy**: dummy=1 if good in language
- **n_brothers**: number of siblings (4 indicates 4 or more)
- **manybooks**: dummy=1 if more than 100 books at home
- **degree_m**: dummy=1 if mother with a degree
- **degree_f**: dummy=1 if father with a degree
- **high_m**: dummy=1 if mother with a high school diploma
- **high_f**: dummy=1 if father with a high school diploma
- **housewife**: dummy=1 if mother housewife, dummy=0 otherwise
- **Dialect**: dummy=1 if language spoken at home is a dialect
- **Foreign language**: dummy=1 if language spoken at home is not Italian
- **Foreign 1st generation**: dummy=1 if students born abroad of foreign-born parents

- **Foreign 2nd generation**: dummy=1 if native-born abroad of foreign-born parents

School and Class characteristics:

- **stud_class**: number of students per class
- **f_m_ratio_class**: percentage of females in the total class size
- **school_size**: number of students per school
- **escs_school**: Average School Level ESCS Index. ESCS refers to the PISA index of economic, social and cultural status
- **Campione**: dummy=1 if class selected for external monitoring by Invalsi

All these variables are from INVALSI.

Area characteristics:

- **lnvapop09**: Total value added per capita, constant prices (base year 2000), 2001 data. Source: Fondazione Istituto Tagliacarne (2006). http://www.tagliacarne.it.
- **mean_est_99_02**: Extortions (1999-2001): average rate of extortions over 10,000 inhabitants. Source: Fiaschi, D., Gianmoena, L. and Parenti, A. (2011)
- **putnam**: Social capital indicator. Source: Cartocci (2007).

B Figures and Tables

B.1 Figures

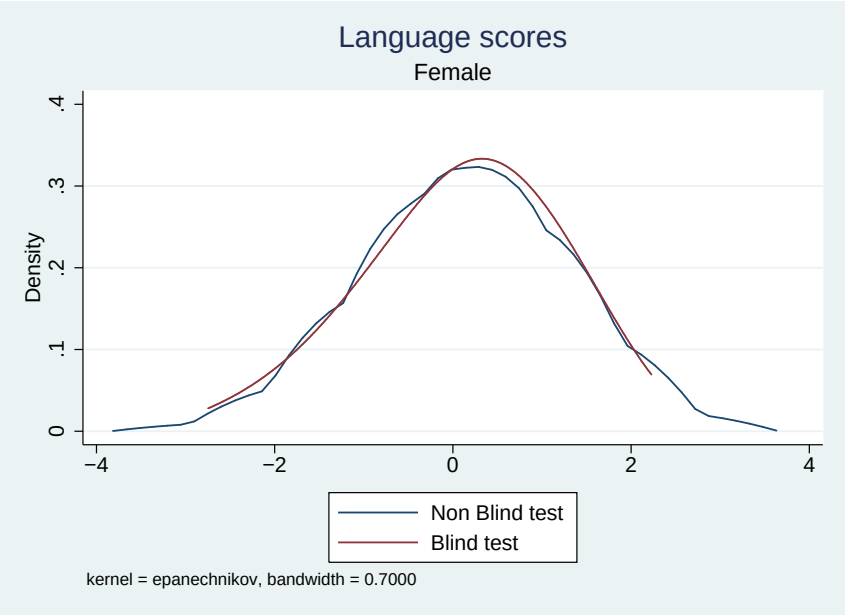

Figure 1: Language scores - Girls

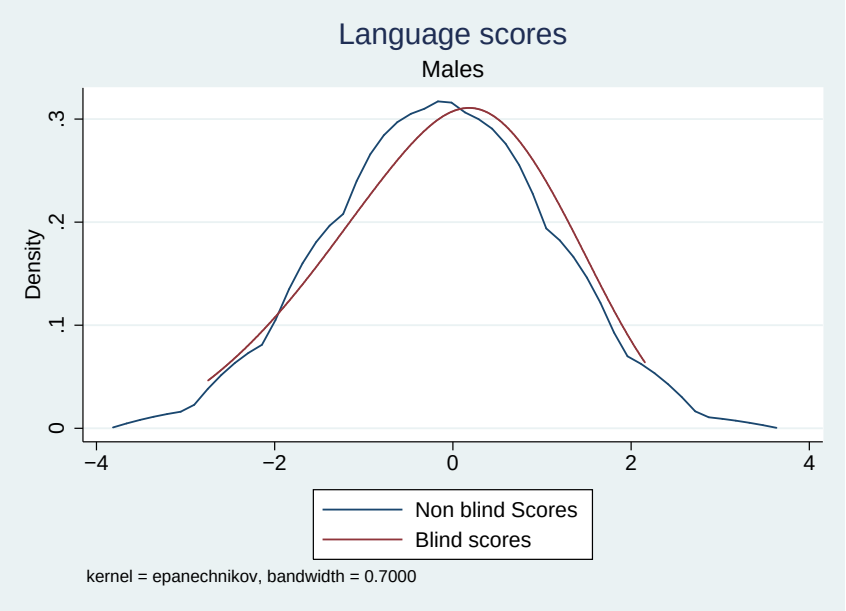

Figure 2: Language scores - Boys

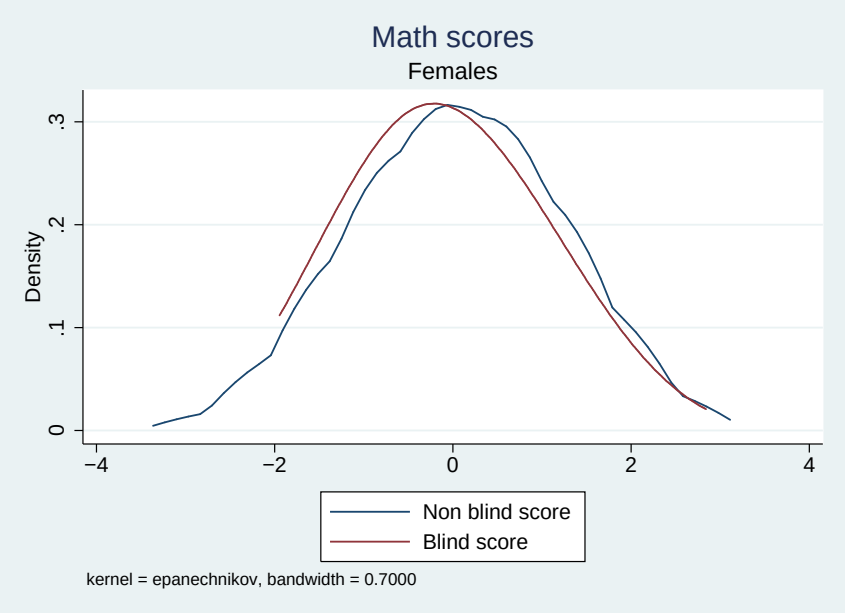

Figure 3: Math scores - Girls

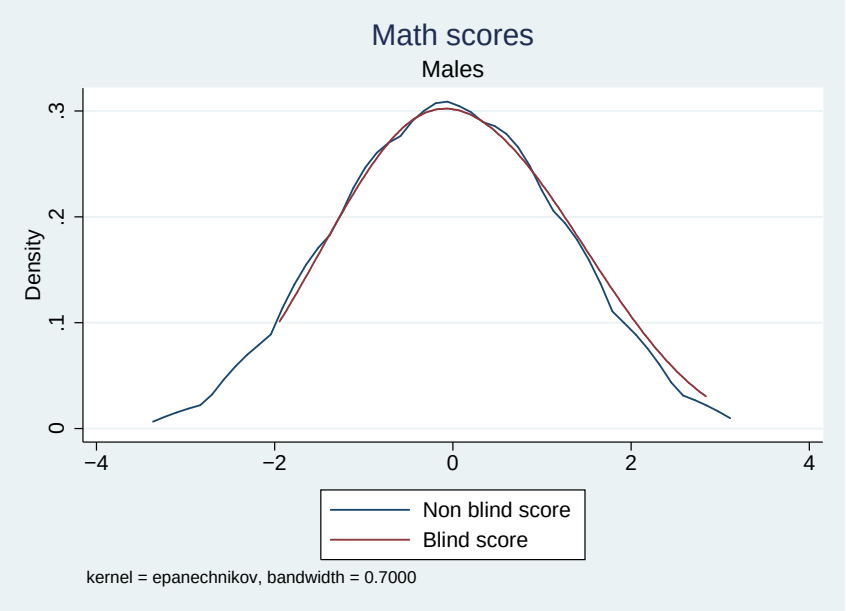

Figure 4: Math scores - Boys

B.2 Tables

Table 1: Descriptives statistics: overall sample

Variable	Obs	Mean	Std. Dev.	Min	Max
Dependent variables					
Language Test	498824	55.21	20.10	0.00	100.00
Math Test	498824	40.65	20.89	0.00	100.00
Language Teacher	498824	55.21	20.10	0.00	100.00
Math Teacher	498824	40.65	20.89	0.00	100.00
Student and family characteristics					
males	498824	0.51	0.50	0.00	1.00
good in math	492657	0.51	0.50	0.00	1.00
good in language	492172	0.48	0.50	0.00	1.00
n_brothers	462457	1.24	0.91	0.00	4.00
manybooks	498707	0.32	0.47	0.00	1.00
degree_m	418947	0.12	0.33	0.00	1.00
degree_f	412435	0.12	0.32	0.00	1.00
high_m	418947	0.38	0.49	0.00	1.00
high_f	412435	0.33	0.47	0.00	1.00
housewife_m	424056	0.40	0.49	0.00	1.00
dialect	467149	0.16	0.37	0.00	1.00
Foreign language	467149	0.07	0.26	0.00	1.00
Foreign 1st generation	498824	0.06	0.23	0.00	1.00
Foreign 2nd generation	498824	0.04	0.18	0.00	1.00
School and Class characteristics					
no stud class	498824	21.74	3.86	1.00	34.00
f m ratio (class)	498824	0.46	0.11	0.00	1.00
no stud school	498824	147.14	77.54	1.00	417.00
escs_school	486597	-0.01	0.47	-2.39	1.78
campione	498824	0.08	0.27	0.00	1.00
Area characteristics					
lnvapop09	498824	10.04	0.29	9.50	10.47
mean_est_ 02	498824	6.50	3.74	1.71	19.45
putnam	498824	-0.69	3.16	-6.43	5.47
North_West	498824	0.25	0.43	0.00	1.00
Centre_North	498824	0.18	0.38	0.00	1.00
Centre_South	498824	0.23	0.42	0.00	1.00
Islands_South	498824	0.16	0.37	0.00	1.00

Table 2: Blind vs non blind test: average results by gender

	Gender	Obs	Mean	Std. Dev.	Min	Max
Language - non blind	Male	255032	49.12	16.56	0.00	100.00
	Female	243792	54.02	16.14	0.00	100.00
Mathematic - non blind	Male	255032	51.58	20.01	0.00	100.00
	Female	243792	53.45	19.32	0.00	100.00
Language - blind	Male	255032	53.16	20.71	0.00	98.53
	Female	243792	57.36	19.21	0.00	100.00
Mathematic - blind	Male	255032	41.92	21.39	0.00	100.00
	Female	243792	39.31	20.28	0.00	100.00

Notes:

Table 3: Means and Standard Deviations of Blind (B) and Non Blind (NB) test and Teachers' Biases Measure at the Student Level by Gender

	Males			Females			
	Non Blind Test	Blind Test	Difference Between N_B and B test	Non Blind Test	Blind Test	Difference Between N_B and B Test	Teachers' Biases Measure (Student Level)
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Math	-0.046	0.107	-0.107	0.049	-0.064	0.112	-0.219
	(-1.016)	0.956	(0.870)	(0.981)	(0.971)	(0.865)	
Language	-0.145	-0.102	-0.043	0.151	0.061	0.045	-0.088
	(1.001)	(1.030)	(0.891)	(0.976)	(1.024)	(0.856)	
Number of Students	255032	255032	255032	243792	243792	243792	498824

Notes: The Blind and Non Blind scores are rescaled and standardized scores. The teachers' biases measured at the student level (column 7) are equal to the difference between boys' blind and non blind scores (column 3) less the difference between girls' blind and non blind scores (column 6). Standard errors are reported in parentheses.

Table 4: Ability in Math and Language studies: boys vs girls self-assessment

Please indicate how much you agree with the following statements or how true it is about you (mathematics) using the following scale: 1-moderately disagree, 2-moderately disagree, 3-somewhat agree, 4 strongly agree	**MALE**	**FEMALE**	**M vs F**
Q3_A - I am good at maths - I am proficient in maths	3.07	2.90	0.17
Q3_B - Studying math is more difficult for me than for most of my classmates	1.93	2.05	-0.12
Q3_C - It is easy for me to learn maths	3.05	2.87	0.18
Q3_D - Studying mathematics is fun	2.81	2.67	0.14
Q3_E - I would like to study more math at school	2.41	2.24	0.17
Please indicate how much you agree with the following statements or how true it is about you (Language) using the following scale: 1-strongly disagree, 2-moderately disagree, 3-moderately agree, 4 strongly agree	**MALE**	**FEMALE**	**M vs F**
Q5_A - I am good at language/Italian - I am proficient in Language/Italian	2.90	3.08	-0.18
Q5_B - Studying Language is more difficult for me than for most of my classmates	2.03	1.79	0.24
Q5_C - It is easy for me to learn Italian/Language	2.96	3.20	-0.24
Q5_D - Studying Italian/Language is fun	2.57	2.92	-0.35
Q5_E - I would like to study more Italian at school	2.15	2.47	-0.33

Notes:

Table 5: Teachers gender bias in Language

Dependent Variable: Test results in Language (Blind and Non blind)						
	(1)	(2)	(3)	(4)	(5)	(6)
Male	-0.209***	-0.166***	-0.171***	-0.168***	-0.176***	-0.176***
	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)
Non-blind score	0.045***	0.040***	0.046***	0.047***	0.046***	0.046***
	(0.004)	(0.004)	(0.004)	(0.004)	(0.004)	(0.004)
Interaction	-0.088***	-0.090***	-0.091***	-0.092***	-0.092***	-0.092***
	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)
dialect		-0.348***	-0.170***	-0.169***	-0.152***	-0.148***
		(0.004)	(0.004)	(0.004)	(0.004)	(0.004)
for_language		-0.296***	-0.256***	-0.257***	-0.253***	-0.254***
		(0.007)	(0.008)	(0.008)	(0.008)	(0.008)
foreign1		-0.633***	-0.443***	-0.443***	-0.445***	-0.449***
		(0.008)	(0.009)	(0.009)	(0.009)	(0.009)
foreign2b		-0.431***	-0.267***	-0.267***	-0.274***	-0.279***
		(0.008)	(0.009)	(0.009)	(0.009)	(0.009)
n_brothers			-0.101***	-0.100***	-0.093***	-0.092***
			(0.002)	(0.002)	(0.002)	(0.002)
manybooks			0.213***	0.210***	0.173***	0.171***
			(0.003)	(0.003)	(0.003)	(0.003)
degree_m			0.419***	0.411***	0.387***	0.389***
			(0.005)	(0.005)	(0.005)	(0.005)
degree_f			0.332***	0.322***	0.302***	0.306***
			(0.005)	(0.005)	(0.005)	(0.005)
high_m			0.309***	0.306***	0.289***	0.290***
			(0.003)	(0.003)	(0.003)	(0.003)
high_f			0.225***	0.221***	0.208***	0.210***
			(0.003)	(0.003)	(0.003)	(0.003)
housewife_m			-0.025***	-0.022***	-0.028***	-0.024***
			(0.003)	(0.003)	(0.003)	(0.003)
no_stud_class				0.001**	0.001**	0.001**
				(0.001)	(0.001)	(0.001)
f_m ratio (class)				0.069***	0.079***	0.085***
				(0.021)	(0.021)	(0.021)
no_stud_school				-0.000***	-0.000***	-0.000***
				(0.000)	(0.000)	(0.000)
escs_school				0.039***	0.059***	0.055***
				(0.006)	(0.005)	(0.006)
good in math					0.275***	0.276***
					(0.003)	(0.003)
good in language					0.176***	0.178***
					(0.003)	(0.003)
lnvapop09						-0.033*
						(0.019)
mean_est_99_02						-0.006***
						(0.001)
social capital (putnam)						0.018***
						(0.001)
Constant	0.107***	0.408***	0.143***	0.106***	-0.080***	0.227
	(0.003)	(0.005)	(0.005)	(0.016)	(0.016)	(0.200)
Regional controls		YES	YES	YES	YES	YES
Observations	997,648	934,298	706,764	689,110	686,406	686,406
R-squared	0.016	0.093	0.193	0.194	0.222	0.223
No. classes	25819	25661	22928	22354	22350	22350

Table 6: Teachers gender bias in Mathematics

Dependent Variable: Test results in Math (Blind and Non blind)						
	(1)	(2)	(3)	(4)	(5)	(6)
Male	0.125***	0.163***	0.161***	0.166***	0.088***	0.088***
	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)
Non-blind score	0.112***	0.113***	0.130***	0.131***	0.130***	0.130***
	(0.003)	(0.003)	(0.004)	(0.004)	(0.004)	(0.004)
Interaction	-0.219***	-0.221***	-0.223***	-0.224***	-0.224***	-0.224***
	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)
dialect		-0.314***	-0.148***	-0.150***	-0.131***	-0.127***
		(0.004)	(0.004)	(0.004)	(0.004)	(0.004)
for_language		-0.222***	-0.183***	-0.186***	-0.192***	-0.194***
		(0.007)	(0.008)	(0.008)	(0.008)	(0.008)
foreign1		-0.494***	-0.314***	-0.314***	-0.317***	-0.321***
		(0.008)	(0.009)	(0.010)	(0.009)	(0.009)
foreign2b		-0.361***	-0.209***	-0.209***	-0.216***	-0.221***
		(0.008)	(0.010)	(0.010)	(0.009)	(0.009)
n_brothers			-0.072***	-0.071***	-0.066***	-0.064***
			(0.002)	(0.002)	(0.002)	(0.002)
manybooks			0.209***	0.207***	0.171***	0.168***
			(0.003)	(0.003)	(0.003)	(0.003)
degree_m			0.398***	0.393***	0.361***	0.362***
			(0.006)	(0.006)	(0.005)	(0.005)
degree_f			0.317***	0.313***	0.284***	0.288***
			(0.006)	(0.006)	(0.005)	(0.005)
high_m			0.286***	0.284***	0.261***	0.262***
			(0.003)	(0.004)	(0.003)	(0.003)
high_f			0.210***	0.207***	0.188***	0.191***
			(0.003)	(0.004)	(0.003)	(0.003)
housewife_m			-0.040***	-0.038***	-0.038***	-0.033***
			(0.003)	(0.003)	(0.003)	(0.003)
no_stud_class				0.001**	0.002***	0.002***
				(0.001)	(0.001)	(0.001)
f_m ratio (class)				0.086***	0.072***	0.078***
				(0.021)	(0.021)	(0.020)
no_stud_school				-0.000***	-0.000***	-0.000***
				(0.000)	(0.000)	(0.000)
escs_school				0.019***	0.041***	0.036***
				(0.006)	(0.006)	(0.006)
good in math					0.556***	0.558***
					(0.003)	(0.003)
good in language					-0.097***	-0.095***
					(0.003)	(0.003)
lnvapop09						-0.018
						(0.019)
mean_est_99_02						-0.007***
						(0.001)
social capital (putnam)						0.019***
						(0.001)
Constant	-0.064***	0.253***	-0.025***	-0.068***	-0.228***	-0.079
	(0.003)	(0.005)	(0.006)	(0.016)	(0.015)	(0.199)
Regional controls		YES	YES	YES	YES	YES
Observations	997,648	934,298	706,764	689,110	686,406	686,406
R-squared	0.003	0.070	0.160	0.161	0.238	0.240
No. classes	25819	25661	22928	22354	22350	22350

Table 7: Robustness checks: Language

Dependent Variable: Test results in Language (Blind and Non blind)						
	(1)	(2)	(3)	(4)	(5)	(6)
Panel A: sub-sample of inspected schools						
Male	-0.208***	-0.170***	-0.175***	-0.172***	-0.181***	-0.182***
	(0.011)	(0.010)	(0.011)	(0.011)	(0.011)	(0.011)
Non-blind score	0.015	0.013	0.019	0.019	0.019	0.019
	(0.012)	(0.012)	(0.013)	(0.013)	(0.013)	(0.013)
Interaction	-0.108***	-0.112***	-0.112***	-0.112***	-0.112***	-0.112***
	(0.009)	(0.009)	(0.010)	(0.010)	(0.010)	(0.010)
Observations	*77,708*	*75,234*	*59,558*	*59,558*	*59,464*	*59,464*
Panel B: High achieving students (self-assessed)						
Male	-0.213***	-0.165***	-0.169***	-0.166***	-0.196***	-0.196***
	(0.006)	(0.006)	(0.006)	(0.007)	(0.006)	(0.006)
Non-blind score	0.213***	0.208***	0.215***	0.216***	0.216***	0.216***
	(0.005)	(0.005)	(0.005)	(0.005)	(0.005)	(0.005)
Interaction	-0.069***	-0.070***	-0.070***	-0.070***	-0.070***	-0.070***
	(0.005)	(0.005)	(0.006)	(0.006)	(0.006)	(0.006)
Observations	*219,690*	*209,422*	*162,272*	*157,922*	*157,768*	*157,768*
Panel C: Low achieving students (self-assessed)						
Male	-0.273***	-0.245***	-0.243***	-0.233***	-0.210***	-0.210***
	(0.017)	(0.017)	(0.018)	(0.018)	(0.018)	(0.018)
Non-blind score	-0.047***	-0.064***	-0.067***	-0.067***	-0.066***	-0.066***
	(0.012)	(0.013)	(0.015)	(0.015)	(0.015)	(0.015)
Interaction	-0.052***	-0.043***	-0.047***	-0.050***	-0.051***	-0.051***
	(0.015)	(0.015)	(0.018)	(0.018)	(0.018)	(0.018)
Observations	*36,600*	*33,858*	*23,944*	*23,376*	*23,344*	*23,344*
Panel D: Northern regions						
Male	-0.214***	-0.177***	-0.170***	-0.169***	-0.186***	-0.186***
	(0.004)	(0.004)	(0.004)	(0.004)	(0.004)	(0.004)
Non-blind score	0.141***	0.136***	0.134***	0.135***	0.134***	0.134***
	(0.005)	(0.005)	(0.005)	(0.005)	(0.005)	(0.005)
Interaction	-0.071***	-0.076***	-0.081***	-0.080***	-0.080***	-0.080***
	(0.004)	(0.004)	(0.004)	(0.004)	(0.004)	(0.004)
Observations	*428,176*	*405,330*	*312,018*	*305,218*	*304,234*	*304,234*
Panel E: Southern regions						
Male	-0.216***	-0.155***	-0.179***	-0.171***	-0.171***	-0.171***
	(0.005)	(0.005)	(0.006)	(0.006)	(0.006)	(0.006)
Non-blind score	-0.026***	-0.031***	-0.021***	-0.021***	-0.022***	-0.022***
	(0.006)	(0.006)	(0.007)	(0.007)	(0.007)	(0.007)
Interaction	-0.105***	-0.107***	-0.104***	-0.104***	-0.103***	-0.103***
	(0.005)	(0.005)	(0.005)	(0.006)	(0.006)	(0.006)
Observations	*392,322*	*363,858*	*275,138*	*267,764*	*266,522*	*266,522*
Panel F: North and sub-sample of inspected schools						
Male	-0.229***	-0.194***	-0.173***	-0.166***	-0.182***	-0.182***
	(0.017)	(0.016)	(0.016)	(0.017)	(0.017)	(0.017)
Non-blind score	0.080***	0.077***	0.073***	0.073***	0.073***	0.073***
	(0.017)	(0.017)	(0.018)	(0.018)	(0.018)	(0.018)
Interaction	-0.090***	-0.094***	-0.098***	-0.098***	-0.098***	-0.098***
	(0.014)	(0.014)	(0.015)	(0.015)	(0.015)	(0.015)
Observations	*29,512*	*28,708*	*22,856*	*22,856*	*22,824*	*22,824*
Panel G: South and sub-sample of inspected schools-						
Male	-0.214***	-0.152***	-0.194***	-0.192***	-0.193***	-0.194***
	(0.019)	(0.018)	(0.019)	(0.018)	(0.019)	(0.019)
Non-blind score	-0.011	-0.013	-0.007	-0.007	-0.007	-0.007
	(0.021)	(0.022)	(0.023)	(0.023)	(0.023)	(0.023)
Interaction	-0.126***	-0.129***	-0.117***	-0.117***	-0.118***	-0.118***
	(0.016)	(0.016)	(0.017)	(0.017)	(0.017)	(0.017)
Observations	*31,064*	*30,020*	*24,178*	*24,178*	*24,128*	*24,128*

Table 8: Robustness checks: Mathematics

Dependent Variable: Test results in Math (Blind and Non blind)						
	(1)	(2)	(3)	(4)	(5)	(6)
Panel A: sub-sample of inspected schools						
Male	0.142***	0.175***	0.169***	0.173***	0.086***	0.085***
	(0.011)	(0.011)	(0.011)	(0.011)	(0.011)	(0.011)
Non-blind score	0.122***	0.125***	0.143***	0.143***	0.143***	0.143***
	(0.011)	(0.011)	(0.012)	(0.012)	(0.012)	(0.012)
Interaction	-0.244***	-0.245***	-0.249***	-0.249***	-0.249***	-0.249***
	(0.009)	(0.009)	(0.010)	(0.010)	(0.010)	(0.010)
Observations	77,708	75,234	59,558	59,558	59,464	59,464
Panel B: High achieving students (self-assessed)						
Male	0.097***	0.119***	0.128***	0.132***	0.132***	0.133***
	(0.006)	(0.006)	(0.006)	(0.006)	(0.007)	(0.007)
Non-blind score	0.261***	0.258***	0.267***	0.267***	0.267***	0.267***
	(0.005)	(0.005)	(0.006)	(0.006)	(0.006)	(0.006)
Interaction	-0.263***	-0.262***	-0.265***	-0.265***	-0.265***	-0.265***
	(0.005)	(0.005)	(0.006)	(0.006)	(0.006)	(0.006)
Observations	*244,350*	*233,552*	*181,242*	*176,658*	*176,416*	*176,416*
Panel C: Low achieving students (self-assessed)						
Male	0.037***	0.071***	0.067***	0.072***	0.042***	0.042***
	(0.011)	(0.012)	(0.013)	(0.014)	(0.013)	(0.013)
Non-blind score	-0.160***	-0.156***	-0.129***	-0.127***	-0.127***	-0.127***
	(0.008)	(0.008)	(0.009)	(0.010)	(0.010)	(0.010)
Interaction	-0.125***	-0.128***	-0.134***	-0.136***	-0.135***	-0.135***
	(0.011)	(0.011)	(0.014)	(0.014)	(0.014)	(0.014)
Observations	*45,728*	*42,554*	*29,998*	*29,200*	*29,130*	*29,130*
Panel D: Northern regions						
Male	0.132***	0.165***	0.171***	0.176***	0.080***	0.080***
	(0.004)	(0.004)	(0.005)	(0.005)	(0.005)	(0.005)
Non-blind score	0.171***	0.171***	0.180***	0.180***	0.180***	0.180***
	(0.005)	(0.005)	(0.005)	(0.005)	(0.005)	(0.005)
Interaction	-0.214***	-0.216***	-0.217***	-0.217***	-0.217***	-0.217***
	(0.004)	(0.004)	(0.004)	(0.004)	(0.004)	(0.004)
Observations	*428,176*	*405,330*	*312,018*	*305,218*	*304,234*	*304,234*
Panel E: Southern regions						
Male	0.105***	0.157***	0.143***	0.149***	0.095***	0.094***
	(0.005)	(0.005)	(0.006)	(0.006)	(0.005)	(0.005)
Non-blind score	0.075***	0.077***	0.095***	0.096***	0.096***	0.096***
	(0.006)	(0.006)	(0.007)	(0.007)	(0.007)	(0.007)
Interaction	-0.229***	-0.231***	-0.235***	-0.236***	-0.236***	-0.236***
	(0.005)	(0.005)	(0.005)	(0.005)	(0.005)	(0.005)
Observations	*392,322*	*363,858*	*275,138*	*267,764*	*266,522*	*266,522*
Panel F: North - sub-sample of inspected schools						
Male	0.121***	0.152***	0.163***	0.171***	0.065***	0.064***
	(0.017)	(0.017)	(0.017)	(0.018)	(0.017)	(0.017)
Non-blind score	0.127***	0.127***	0.141***	0.141***	0.141***	0.141***
	(0.017)	(0.017)	(0.018)	(0.018)	(0.018)	(0.018)
Interaction	-0.223***	-0.225***	-0.227***	-0.227***	-0.227***	-0.227***
	(0.014)	(0.014)	(0.016)	(0.016)	(0.016)	(0.016)
Observations	*29,512*	*28,708*	*22,856*	*22,856*	*22,824*	*22,824*
Panel G: South - sub-sample of inspected schools						
Male	0.145***	0.196***	0.164***	0.170***	0.104***	0.103***
	(0.017)	(0.017)	(0.018)	(0.018)	(0.017)	(0.017)
Non-blind score	0.163***	0.166***	0.180***	0.180***	0.180***	0.180***
	(0.020)	(0.020)	(0.022)	(0.022)	(0.022)	(0.022)
Interaction	-0.280***	-0.280***	-0.280***	-0.280***	-0.279***	-0.279***
	(0.015)	(0.015)	(0.017)	(0.017)	(0.017)	(0.017)
Observations	*31,064*	*30,020*	*24,178*	*24,178*	*24,128*	*24,128*