Is There a Bias Towards Girls in Non Anonymous Evaluation?

Preliminary and Draft

Catarina S. Angelo [†]

Nova School of Business and Economics

May 2014

Abstract

This paper investigates the gender gap in high school students achievement in Portugal and explores to what extent this can be driven by the method of evaluation. In class evaluation results not only from measures of cognitive skills but also from non cognitive skills, like ability to pay attention in class, to organize class materials, to keep track of home-work or class materials in which male students underperform relative to girls. This is in contrast with national exam tests that focus only on cognitive skills. We compare results from national exams, required for higher education admission, with results from in class evaluation for the same student. The data set used has the exam and in class scores for Portuguese students that took the national exams between 2007 and 2012. By examining the difference between results from in class evaluation and national exams, controlling for age, type of school and vocational track we can identify a gender gap in student achievement most notably in favor of girls in portuguese and mathematics and in both evaluation schemes. The analysis also shows that the difference between in classroom evaluation and the exam score positive and higher for female students. This suggests that there is a bias in favor of girls in classroom evaluation.

JEL Classification: I21, J16.
Keywords: Educational Evaluation Schemes, Gender Gap

[†]The author benefited from the financial support of *FCT — Fundação para a Ciência e Tecnologia*, reference SFRH/BD/70215/2010, which is gratefully acknowledged. I am grateful to my supervisor, Professor Ana Balcão Reis, for her sugestions and constant encouragement.

1 Introduction

The question of whether boys and girls differ in a significant way in their cognitive abilities fostered a considerable debate in recent years, and empirical research on the subject conveyed several regularities regarding the sign and size of the gender gap in student achievement. This gender gap has been observed throughout the world, and although female students generally outscore boys in languages, the results are not so compelling in mathematics (Machin and Pekkarinen, 2008, Hyde et al., 2008).

Scores obtained from teachers are not only determined by the results on tests taken in classroom but also a consequence of a set of behavioral attitudes that the teacher can reward, either explicitly or not. These are mainly driven by personality skills which are not necessarily the same for boys and girls. Girls are more amenable to the learning process than boys, which makes teachers to be more prone to reward them with a higher score (Cornwell et al., 2013). Moreover, the rewarding of elicited attitudes may be seen as a positive message from the teacher to the students that can influence their future achievement. Pavlova et al. (2010) showed that not only a positive message enhances performance whereas a negative message diminishes it, but also that this effect is more pronounced for girls, for whom a negative stereotype message has a stronger impact.

In this work I follow Falch and Naper (2013) to investigate whether the observed gender differences in student achievement in Portugal are to some extent related to the evaluation scheme, by making use of the final scores for Portuguese students at the end of each cycle of studies that comprise compulsory schooling. My purpose is to test the role of personality skills on a student's score. Under the hypothesis that personality skills are easily rewarded in humanities than in sciences, a higher difference between the teacher's and the exam score would be expected in languages, which would be consistent with a grading bias against boys. Moreover, if non-cognitive skills develop later in life for boys than for girls we would expect the latter difference to narrow at higher levels of education.

For Portugal, initial findings regarding portuguese language and mathematics show that, at the end of compulsory schooling, girls outperform boys in both subjects. This is true when we consider the scores obtained from in classroom evaluation and also for the scores obtained in national exams. Moreover, the difference between the scores obtained from their teachers and the scores obtained in the national exam is positive and higher for female students in portuguese and the opposite is true in mathematics, a fact that is consistent with the hypothesis that there is a grading bias against boys in languages. This is to some extent in line with the findings

of Lavy, 2008 in a study of Israeli public high-school students, according to which there is a grading bias from high-school teachers against boys.

2 Institutional Setting

Since 2009, Portugal has 12 years of compulsory education[1], divided into basic and secondary education, where children attend school from the year they turn six to the year they turn 18. In basic education there are three sequential cycles of study, with 4, 2 and 3 years long. At the year a child turns 6 she enters the first cycle, where she stays for at least 4 years. The second cycle comprises the fifth and the sixth years of compulsory schooling and finally the third cycle goes from the seventh to the ninth grade. Secondary schooling goes from the tenth to the twelfth grade and is the last cycle of studies before admission to university.

In the end of basic education students can choose between different tracks of secondary education. The academic track targets students who want to pursue a university degree. The students enrolled in this track represented, in the 2011/2012 school year, roughly 51% of the secondary students. Within this track there are four major areas of studies, Sciences and Technology, Economics, Languages and Humanities and Visual Arts. The other major track is the professional one, with 28% of the students enrolled in secondary education, which is tailored to the students who want to obtain professional qualification that allows them to enter the labor market. Together with these coexist others where the 21% remaining students are enrolled. Regardless of the track chosen by the students, they can always decide to pursue university degree.

Private and public schools coexist in all levels of education. In the school year of 2010/2011 there were 5422 basic schools, 356 secondary schools and 319 schools with basic and secondary education. At the end of a grade a student will apply to the school he wishes to attend in the next year. If after applying the criteria for admission the number of applicants exceeds the number of vacancies, the school has discretionary power to decide on a student's application. Given that the *best* schools' applicants often exceed its vacancies the student curriculum can be decisive for admission.

Students are tested at the end of each cycle of studies by means of national exams prepared by the Office for Educational Evaluation, *IAVE*, a central service from the Ministry of Education. These tests are meant to provide a measure for the level of knowledge of students in core subjects of the curricula and are graded anonymously, meaning that the teacher who is grading the test

[1]This applies to the students enrolled in the 7th grade in 2009/2010

has no knowledge about the student who took it besides the one he can infer from the answers he gives.

Portuguese and mathematics are the subjects tested at the end of each cycle of basic schooling whereas in the end of secondary education the national exam on portuguese is the only that is common and mandatory for all students. Secondary students are also tested in one or more other subjects, which are specific of their study track, in the end of the eleventh and the twelfth grade. The final score of a student in a subject for which the national exam is mandatory is determined in at least 30% by the exam score whereas the last 70% are determined by in classroom evaluation.

The non-anonymous component of the score is the responsibility of the student's teacher, and is not driven solely by the results on written tests taken in the classroom. Teachers can reward, either explicitly or not, the student's effort and commitment towards their course, which is basically a way to evaluate a range of personality skills (like emotional maturity, empathy, interpersonal skills and verbal and non-verbal communication) that can influence the behavior of a student. Overall, national exams and teacher grading are based on the same curricula[2] in what concerns measuring of cognitive skills and level of knowledge but differ substantially in the personality traits that can determine a good result.

3 Data and descriptive statistics

The data sets used are made available by *IAVE* and contain yearly information, from 2007 to 2012, on the results of the students that take the national exams. As I want to compare in classroom evaluation with exam scores I only consider internal students[3]. For this students, we know the school they were enrolled in the school year they take the exam and its geographical location, their age, gender and study track. Given that the identifier in the data set is the exam, we are not able to know if the student took one or more exams, but for each exam we know the exam score, the score obtained from in classroom evaluation and the final score for the course. Given that internal students can only take the national exam on a subject if the score assigned by their teacher in that subject is equal or above 8, the data set excludes the students in the left tail of the scores distribution.

For secondary education, the grading scale for in classroom evaluation and course final score is

[2]The curriculum is defined nationally.

[3]An internal student is someone who in the year he takes the exam, is enrolled in the regular academic track and attends the classes for the subject on which he takes the national exam.

from 1 to 20 an for the course national exam is 0 to 200. For convenience and in order to work with the same scale for both evaluations, the exam score was converted into a 1 to 20 scale. The analysis is performed on a first stage only for the students that are enrolled in the academic track of secondary education.

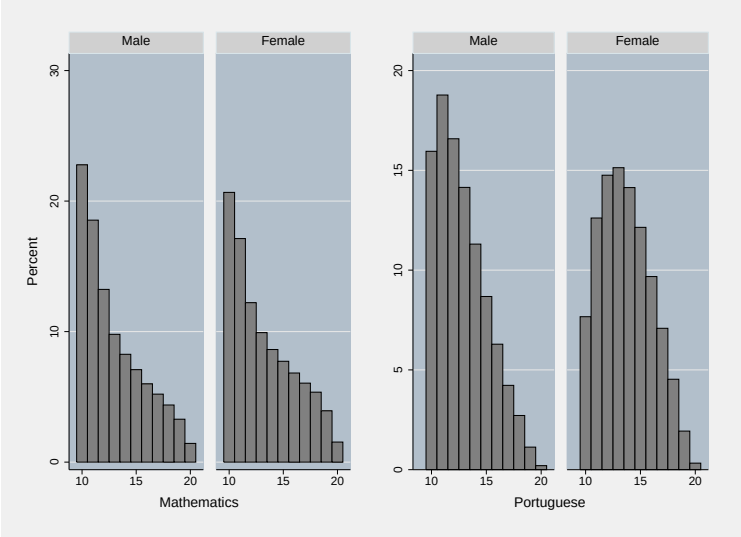

Figure 1: Percentage distribution of teacher's in classroom evaluation

It is clear from *Figures 1* and *2* that the scores obtained by female students are better than the ones obtained by male students, both for in classroom evaluation and national exams. Female students clearly perform better than male students in portuguese and mathematics. The difference is however less pronounced in mathematics, a fact that is in line with previous findings regarding boys and girls performance in this subject. Despite this, the gap narrows when we observe the scores obtained in national exams.

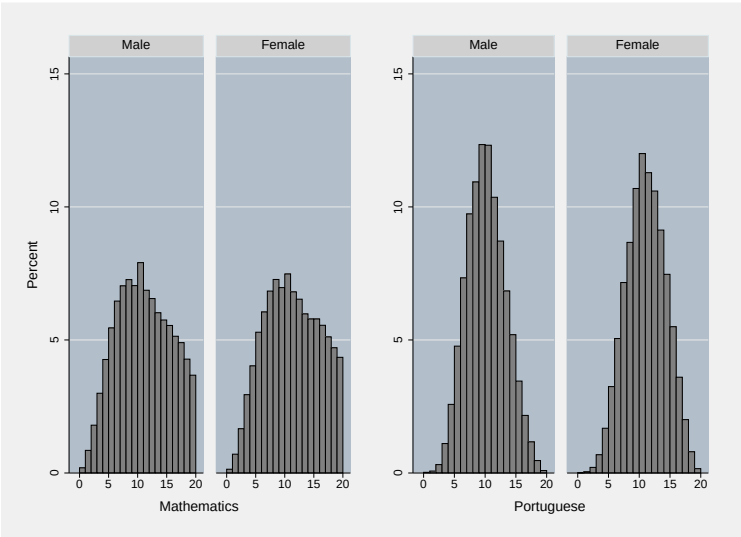

Figure 2: Percentage distribution of exams' scores

Regarding the scores from national exams, we can observe from *Table 1* that girls obtain on average a score 0.23 higher than boys in mathematics and of 0.89 in portuguese. This difference increases for mathematics for scores given by teachers and decreases for portuguese, with girls outperforming relative to boys on average by 0.26 and 0.85, respectively. All the differences are statistically significant.

Table 1: Mean comparison tests by gender and evaluation scheme (2007-2012)

	Mathematics			Portuguese		
	Teacher score	Exam score	Difference	Teacher score	Exam score	Difference
All	13.09	11.15	1.94	13.35	10.70	2.65
	[2.79]	[4.63]	[3.21]	[2.35]	[3.21]	[2.54]
Females	13.21	11.26	1.95	13.69	11.06	2.64
	[2.82]	[4.65]	[3.20]	[2.34]	[3.20]	[2.54]
Males	12.95	11.03	1.92	12.84	10.17	2.68
	[2.75]	[4.61]	[3.23]	[2.28]	[3.14]	[2.54]
Difference	-0.26	-0.23	-0.03	-0.85	-0.89	0.03
(Male-Female)	(-22.36)	(-11.80)	(-2.40)	(-1.1e+02)	(-82.34)	(4.53)

Note. Standard deviations in brackets and t-values in parentheses. All differences are statistically significant at at least $\alpha = 0.01$

The results for portuguese language are in line with the ones obtained from PISA tests, with girls outperforming boys at all levels. This is however not true when it comes to mathematics. In PISA, on average, boys perform better than girls, with a statistically significant difference of 11 points. However, the results are not contradictory. While PISA tests 15 year-old students, irrespectively of the grade they attend, and aims at having a representative sample of students, in this work I look at students that take the national exams in the 12th grade and are enrolled in the academic track. This difference on itself should suffice to justify the difference obtained.

4 Bias in non-anonymous evaluation

4.1 Empirical strategy

I estimate the following model,

$$Sd_{ijt} = S_{E=1,it} - S_{E=0,it} = \alpha + \rho S_{E=1,it} + \delta G_i + \sigma G_i * S_{E=1,it} + \beta X_{ijt} + \gamma_j + \mu_t + \epsilon_{ijt} \quad (1)$$

Where Sd_{ijt} is the difference between student i's scores for in classroom evaluation ($E = 1$) in school j and national exam ($E = 0$), in time t. Sd_{ijt} is assumed to be a function of the students' gender, G ($G = 1$ for female and $G = 0$ for male). The difference also depends on the score obtained from teacher's evaluation, T_{ijt}. This accounts for the fact that the performance in the exam relative to in classroom of high and low achieving students can be different. An interaction effect between the dummy variable for female students and teacher grading is also included, $\sigma G_i * S_{E=1,it}$.

Each student is observed only at one point in time. The model includes as co-variates the students age and the moment in time he takes the exam (X_{ijt}), and year and school fixed effects, μ_t and γ_j respectively. I estimate the model separately for each subject and also for both subjects together.

4.2 Results

Table 2 presents the estimation results of the model specification described in Eq.(1). The first model includes both subjects, while the rest of the table presents the results separately for each subject. The table reports only the parameters of main interest.

The interaction effect between the dummy variable for female students and teacher score is rather small, but negative, both for portuguese and mathematics, and positive when we look at both subjects together. This would indicate that, on average, higher achieving female students are rewarded significantly worst by their teachers.

However, the coefficient for female students is positive, meaning that the difference between teacher and exam score is higher for girls, which suggests that, on average, girls perform worst than boys in the exam relative to in classroom evaluation. A plausible reason for the gender difference in scores differences is that at some levels of the scores distribution teachers give better grades to a girl than a boy of apparent similar ability. This supports the hypothesis that there is a bias

Given that both assessments, by the teachers and in national exams, have a similar high stakes component, it is not straightforward to interpret these results as girls performing worst than boys in a more competitive environment. Moreover, even if we consider that the national exam has a higher high stakes component, in a recent work Falch and Naper, 2013 find evidence that does not support the hypothesis that girls perform relatively worst than boys when stakes are high.

Table 2: Estimation results

Variable	**All Subjects**	**Portuguese**	**Mathematics**
Age	0.283**	0.361**	0.232**
	(0.007)	(0.008)	(0.012)
Teacher Score	-0.006*	0.211**	-0.207**
	(0.002)	(0.003)	(0.003)
Female	-0.215**	0.206**	0.228**
	(0.037)	(0.047)	(0.053)
Female x Teacher Score	0.020**	-0.027**	-0.008*
	(0.003)	(0.003)	(0.004)
Exam 1st	0.106**	0.423**	-0.178**
	(0.008)	(0.010)	(0.014)
Private School	0.041**	-0.038**	0.225**
	(0.011)	(0.013)	(0.018)
Intercept	-2.389**	-6.509**	2.411**
	(0.136)	(0.166)	(0.797)

Significance levels : † : 10% * : 5% ** : 1%

5 Concluding remarks

The preliminary results obtained so far indicate a significant bias in favor of girls when they are assessed by their teachers which can be due to the fact that the skills under evaluation in the classroom are different from the ones that determine a good score in the national exam. The difference between the teacher and national exam score is higher for girls than for boys, when we account for its other determinants . Systematically over grading girls may be responsible for the increase in female enrollment rates at higher levels of education with potential impact in labor market characteristics and economic growth.

Further empirical research is required in order to determine the causes for this bias. It is important to acknowledge whether it results from the different nature of the tests taken in classroom and nationally or if instead, the bias is driven by student teacher interaction, through which teachers systematically reward girls' effort. This reward of girls' effort can be positive for girls but may have a potential negative impact on boys.

References

(2012). Diário da República.

Bedard, K. and Cho, I. (2010). Early gender test score gaps across OECD countries. *Econ. Educ. Rev.*, 29(3):348–363.

Bonesrø nning, H. (2008). The Effect of Grading Practices on Gender Differences in Academic Performance. *Bull. Econ. Res.*, 60(3):245–264.

Borghans, L., Meijers, H., and Ter Weel, B. (2008). the Role of Noncognitive Skills in Explaining Cognitive Test Scores. *Econ. Inq.*, 46(1):2–12.

Cornwell, C. M., Mustard, D. B., and Parys, J. V. (2013). Non-cognitive Skills and the Gender Disparities in Test Scores and Teacher Assessments : Evidence from Primary School. *J. Hum. Resour.*, Volume: 48(1):236–264.

Cunha, F., Heckman, J., and Schennach, S. (2010). Estimating the Technology of Cognitive and Noncognitive Skill Formation. *Econometrica*, 78(3):883–931.

Dee, T. S. (2005). Teachers and the Gender Gaps in student Achievement.

Dee, T. S. (2006). Teachers and the Gender Gaps in Student Achievement. *J. Hum. Resour.*, (September 2006).

Delisle, M.-N., Guay, F., Senécal, C., and Larose, S. (2009). Predicting stereotype endorsement and academic motivation in women in science programs: A longitudinal model. *Learn. Individ. Differ.*, 19(4):468–475.

Ellison, G. and Swanson, A. (2010). The Gender Gap in Secondary School Mathematics at High Achievement Levels: Evidence from the American Mathematics Competitions. *J. Econ. Perspect.*, 24(2):109–128.

Falch, T. and Naper, L. R. (2013). Educational evaluation schemes and gender gaps in student achievement. *Econ. Educ. Rev.*, 36:12–25.

Giorgi, G. D., Bharadwaj, P., Hansen, D., and Neilson, C. (2012). The Gender Gap in Mathematics.

Guiso, L., Monte, F., Sapienza, P., and Zingales, L. (2008). Culture, Gender and Math. *Science (80-.).*, 320.

Harlen, W. (2005). Trusting teachers judgement: research evidence of the reliability and validity of teachers assessment used for summative purposes. *Res. Pap. Educ.*, 20(3):245–270.

Heckman, J., Pinto, R., and Savelvey, P. (2013). Understanding the Mechanisms Through Which an Influential Early Childhood Program Boosted Adult Outcomes. *Am. Econ. Rev.*, 103(6):2052–2086.

Hinnerich, B. T., Höglin, E., and Johannesson, M. (2011). Are boys discriminated in Swedish high schools? *Econ. Educ. Rev.*, 30(4):682–690.

Husain, M. and Millimet, D. L. (2009). The mythical boy crisis? *Econ. Educ. Rev.*, 28(1):38–48.

Hyde, J. S., Lindberg, S. M., Linn, M. C., Ellis, A. B., and Williams, C. C. (2008). Gender Similarities Characterize Math Performance. *Science (80-.).*, 321(July):494–495.

Jacob, B. a. (2002). Where the boys aren't: non-cognitive skills, returns to school and the gender gap in higher education. *Econ. Educ. Rev.*, 21(6):589–598.

Kleinfeld, J. (1998). The Myth That Schools Shortchange Girls : Social Science in the Service of Deception University of Alaska. Technical report.

Lavy, V. (2008). Do gender stereotypes reduce girls' or boys' human capital outcomes? Evidence from a natural experiment. *J. Public Econ.*, 92(10-11):2083–2105.

Machin, S. and Pekkarinen, T. (2008). Global Sex Differences in Test Score Variability. *Science (80-.).*, 322(November).

Mechtenberg, L. (2009). Cheap Talk in the Classroom : How Biased Grading at School Explains Gender Differences in Achievements, Career Choices and Wages. *Rev. Econ. Stud.*, 76(4):1431–1459.

Mendoza-Denton, R., Park, S. H., and OConnor, A. (2008). Gender stereotypes as situationbehavior profiles. *J. Exp. Soc. Psychol.*, 44(4):971–982.

Ministério da Educação e Ciência (2007). *Exames Nacionais do Ensino Secundário - Guia Geral de Exames 2007.*

Ministério da Educação e Ciência (2008). *Exames Nacionais do Ensino Secundário - Guia Geral de Exames 2008.*

Ministério da Educação e Ciência (2009). Exames Nacionais do Ensino Secundário - Guia Geral de Exames 2009.

Ministério da Educação e Ciência (2010). *Exames Nacionais do Ensino Secundário - Guia Geral de Exames 2010.*

Ministério da Educação e Ciência (2011). *Exames Nacionais do Ensino Secundário - Guia Geral de Exames 2011.*

Ministério da Educação e Ciência (2012). *Exames Nacionais do Ensino Secundário - Guia Geral de Exames 2012.*

Ministério da Educação e Ciência (2013). Exames Nacionais do Ensino Secundário - Guia Geral de Exames 2013.

Ministério da Educação e Ciência (2014). Exames Nacionais do Ensino Secundário - Guia Geral de Exames 2014.

Mulligan, C. B. (2004). The Closing of the Gender Gap as a Roy Model Illusion.

Pavlova, M. a., Wecker, M., Krombholz, K., and Sokolov, A. a. (2010). Perception of intentions and actions: gender stereotype susceptibility. *Brain Res.*, 1311:81–5.

Reemployment, P., Experiments, B., and Pennsylvania, T. (2002). INFERENCE ON THE QUANTILE REGRESSION PROCESS By Roger Koenker and Zhijie Xiao 1. 70(4):1583–1612.