WORKING PAPER

#640 | 2022

Gender Gaps in Different Assessment Systems: The role of teacher gender

CATARINA ANGELO, ANA BALCÃO REIS

NOVA SCHOOL OF BUSINESS & ECONOMICS

Gender Gaps in Different Assessment Systems: The role of teacher gender

Catarina Angelo[ab] Ana Balcão Reis[b]

Abstract

Previous research has identified a gender gap in the difference between teacher grading and scores on national exams at the end of secondary school. We go a step further and look at how teacher characteristics may influence this gender gap. We find that exams are relatively more favorable for boys, regardless of the teacher gender or the gender matching. Results suggest that having a male teacher tends to increase the assessment gap for all students through a greater decrease from teacher grades to exam scores, the impact being less for boys.

JEL Classification: I21, I24, J16.

Keywords: Student Achievement, Gender Gap, Grading Practices, Teacher Gender

---

[a]The author has benefited from the financial support of *FCT-Fundação para a Ciência e Tecnologia*, reference SFRH/BD/70215/2010, which is gratefully acknowledged.

[b]This work was funded by Fundação para a Ciência e a Tecnologia (PTDC/IIM-ECO/6813/2014, UID/ECO/00124/2019, UIDB/00124/2020, UIDP/00124/2020, PTDC/EGE-ECO/4764/2021 and Social Sciences DataLab - PINFRA/22209/2016), POR Lisboa and POR Norte (Social Sciences DataLab, PINFRA/22209/2016).

Ana Balcão Reis is an Associate Professor at Nova School of Business and Economics. Catarina Angelo is a graduate student at Nova School of Business and Economics.

The authors declare to have no relevant or material financial interests that relate to the research described in this paper.

1 Introduction

Differences between academic performance of male and female students continue to raise concerns as they may be a source of an unequal representation of men and women in both higher education and the labor market.

Even though the last three decades have been prolific with respect to the analysis of the root causes of the gender gap in achievement, the results are far from being settled. One of the causes pointed out for the gender gap in achievement is the type of assessment, with differences between teacher and exam assessment starting early in primary school (Cornwell et al.(2013)). Another potential cause is the effect of teacher gender, with mixed findings concerning its effect on student achievement (Dee (2007); Ammermueller and Dolton (2006); Cho (2012); Winters et al. (2013); Paredes (2014); Sansone (2017)). This paper looks at the interaction between these two factors.

In many educational systems academic assessment is not only carried out systematically and by different means throughout a student's academic path but also used to determine the choice of the track of studies and/or to rank students as they apply for university. Because scores obtained by students are important determinants of career choices and opportunities, it is important to understand if different types of assessment lead to systematic differences between boys and girls and what the drivers of these differences are.

In this paper we analyze the gender gap in different types of assessment - teacher grading and national exams - for upper secondary school students and the extent to which this gap is driven by teachers' gender. Our analysis focuses on how assignment to a same gender teacher affects the gender gap in the scores on assessments for Portuguese public-school students in the academic track of secondary education between 2008/09 and 2016/17.

This paper contributes to the literature on the causes of the gender gap in achievement by connecting two branches of research that have so far been treated separately. We jointly analyze the effect of teacher gender and grading systems on the gender gap in achievement and show that different grading systems affect boys' and girls' achievement differently, even when teachers' gender is taken into account. We also show that teachers' gender has an impact on the assessment gap.

Our results indicate that although both boys and girls on average earn lower scores on national exams than their grades obtained from teacher assessment, boys have on average a lower assessment gap than girls, meaning that the scores they obtain on national exams decrease less with respect to the teacher-assigned score than do girls' scores, i.e., the fall in scores is greater among girls. Moreover, students who had male teachers have a greater assessment gap than students who had female teachers, meaning that the scores they obtain on national exams decrease more with respect to teacher grading than when they were taught by female teachers, but this effect is smaller for boys.

For a better understanding of the mechanism behind these results we also investigate if there are systematic differences between male and female teachers when they assign a final score to a student, and also if there are systematic differences between the national exams' scores of students that have been taught by male or female teachers. Regression analysis using the score assigned by teachers as dependent variable shows that the effect of being assigned to a male teacher is significant in only two of the subjects analyzed and of opposite sign. Thus, we reject the hypothesis that male teachers are more prone to assign higher scores to their students than female teachers. Using as dependent variable the exam score we obtain that the effect of being assigned to a male teacher, when significant, is always negative. Thus, results suggest that on average male teachers' students are not as well prepared for exams as those who had female teachers.

The remainder of this paper is organized as follows. The next section provides a review of the literature, and Section 3 gives an overview of the Portuguese educational system. In Section 4 we present the data used and some descriptive statistics. Section 5 discusses the methodology used and the main results. In Section 6 we investigate a possible mechanism for the main results. The final section concludes.

2 Literature Review

Two branches of literature relate closely to this paper. The first focuses on differences between types of assessment. Using data on Norwegian student scores in the 10th grade (Falch and Naper (2013)), and Swedish student data for 9th graders (Lindahl (2016)), findings indicate that an evaluation system that relies heavily on teacher grading lowers boys' scores relative to girls'. Marcenaro-Gutierrez and Vignoles (2015) compare teacher and test-based assessments in reading and Mathematics, for two cohorts of Spanish students, aged 11 and 15. They find evidence that in Mathematics the difference between teacher assessment and test-based results is significantly greater for girls than for boys, meaning that teacher evaluation benefits girls. In reading they obtain no significant difference between genders. For the US, using data for students from kindergarten through the 5th grade, Cornwell et al. (2013) show that although boys perform as least as well as girls on math tests, and significantly better on science tests, they are graded less favorably by their teachers. Similar results are obtained by Angelo and Reis (2020) using data on Portuguese students from the 4^{th} to the 12^{th} grade. They find that scores from teacher grading are on average higher than those obtained on central exams and that girls are favored by the former type of assessment.

Several mechanisms could be driving the results obtained for the differences between boys and girls in different types of assessment. Considering the different results presented in the literature from different educational systems, teachers' gender bias is one of the channels that may contribute to the gender gap (Lavy (2008); Hinnerich et al. (2011); Carlana (2019); Terrier (2005); Lavy and Sand (2018)). However, given that in most cases the two evaluations being compared are dissimilar in nature, other factors may contribute to explain this gap. Different responses of boys and girls to stressful situations, as is the case of national exams, is a possible explanation. While Falch and Naper (2013) do not seem to find evidence of a different response, most studies conclude in favor of the existence of gender differences in response to stressful situations, with girls performing relatively more poorly than boys when the stakes are high (Ors et al. (2013); Jurajda and Münich (2011); Cai et al. (2018); Azmat et al. (2016)). Introducing another perspective, Pekkarinen (2015) studied gender differences in the strategies adopted when answering multiple choice university entrance exams, finding evidence that boys and girls tend to adopt different strategies and the ones adopted by girls under more competitive pressure contribute to their lower performance relative to their results in less competitive evaluations. There is also some research analyzing the effect of grading schemes on gender gaps, namely comparing relative with criterion-referenced grading (Czibor et al. (2014)) and different weights of coursework elements in students' evaluation (Machin and McNally (2005)). The evidence provided sustains the hypothesis that these channels may also influence the gap in assessment.

The second branch of research relating closely to our work concerns the effect of teacher gender in student achievement. While earlier studies focused on tertiary education, the first compelling results concerning earlier grades are reported by Dee (2007). He uses several different 8^{th} grade student outcomes (test scores, teacher perception of student performance and student perception of a given academic subject) for two subjects of the curriculum and finds evidence that the effect of assignment to same-gender teacher differs by subject and is stronger for girls. Girls benefit from assignment to a female teacher in History but their scores are lower when the same-gender assignment happens in Mathematics. For boys the effect of assignment to a female teacher is always negative. Following the same approach as Dee (2007), Ammermueller and Dolton (2006) find no effect of same-gender teacher assignment at age 9. However, for the 8^{th} grade these authors report a positive effect in the teaching of Mathematics in England but not in the USA. Cho (2012) investigates same-gender teacher assignment in 15 OECD countries using four waves of TIMSS data for lower secondary school students. She finds no significant impact on test scores of same gender teacher assignment in eight countries, a positive impact for boys in four, and a positive impact for girls in the remaining three. Winters et al. (2013) use panel data on students in grades 3 to 10 enrolled in Florida public schools over a period of five years and find no significant effect of assignment to same-gender teacher in elementary grades. For middle and high school grades they are able to identify a positive effect from assignment to a female teacher, regardless of the gender of the student, although it is greater for girls.

Different results obtained for students of different age and gender need not be taken as contradictory. According to Martin and Ruble (2004), at the age of 11 to 13 girls start to understand and apply gender stereotypes. It is around this age that the gender difference in mathematics performance becomes more significant and the subject is seen as a predominantly male domain (Ambady et al. (2001); Hyde et al. (1990)).

The literature offers alternative explanations for the results obtained to date. According to a role-model effect, same-gender teachers may work as a better role model with students benefiting from same-gender teacher assignment through higher engagement (Ammermueller and Dolton (2006); Carrington et al. (2008); Francis et al. (2008); Krieg (2005); Nixon and Robinson (1999)). But it may also be the case that student achievement is influenced by teacher's expectations concerning the student's performance (Ehrenberg et al. (1994); Carrington et al. (2008); Nixon and Robinson (1999); Rosenthal and Jacobson (1968)).

Paredes (2014) explores the mechanisms through which teacher–student gender matching may have a positive effect on student achievement. Using data for 8^{th} graders in Chile, she obtains that being assigned to a female mathematics teacher increases the average scores of girls by approximately 0.04 standard deviations. Moreover, she finds no evidence of an effect from assignment to a male teacher on boys' scores. She also provides evidence that the positive effect found for girls is due to a role model effect rather than to a teacher bias effects. Using US data from the HSLS:09 and extending Paredes (2014) by including information on teachers' behavior, Sansone (2017) finds no effect for teacher gender once the teacher's attitudes and behaviors are taken into account. We look at the impact of teacher gender on the gender assessment gap, thereby combining the two branches of research.

3 Institutional Setting

The Portuguese educational system comprises four cycles of compulsory education, with students attending school from the ages of 6 to 18.[1] Herein we focus on upper secondary education, which comprises the 10th, 11th and 12th grades.

When enrolling in secondary education a student must choose between the different tracks available. We consider only the academic track, which targets students who wish to pursue a university degree, as these are the students for whom national exams are mandatory. Students who choose this track represent more than 50% of all students enrolled in secondary education for the years under analysis. Within the academic track the student chooses which field of studies to attend from the four available: Sciences and Technology, Socio-Economic Sciences, Languages and Humanities, and Visual Arts.

These four fields of study share a common set of courses: Portuguese Language, one foreign language, Philosophy and Physical Education.[2] For each field of studies there is also a specific component which includes one mandatory triennial course attended by the student during the three years of secondary education, two biennial courses to be attended in the 10th and, 11th grade and two annual courses attended in the 12th grade.

The students in the academic track of secondary education must take national exams in at least four subjects to obtain their diploma. At the end of the 11th grade a student must take national

[1]In 2009, a law was approved extending compulsory education to 12 years or until the student is 18. After 2016, all students in the Portuguese educational system face 12 years of compulsory education, therefore including secondary education.

[2]The foreign languages available are English, French, German and Spanish.

exams in two biennial courses of the specific component of the field of studies she/he is enrolled in or in one of the biennial courses and Philosophy. At the end of the 12^{th} grade all students take the national exam in Portuguese Language and in the triennial subject mandatory for their field.

Exams take place in June, one to two weeks after the end of classes and after the publication of teacher grades. They are curriculum based, include both multiple choice and open-ended questions, and last from 90 to 150 minutes depending on the subject. The evaluation of the exam is based on the answer key and grading scheme provided by an agency of the Ministry of Education. Exams are graded anonymously by teachers from a school different from the one attended by the student. They are identified only by an ID-number (the student's name does not appear on the exam paper).[3] Grades are assigned based on absolute or criterion-referenced evaluation. There is no "curving" or relative grading. Exams from earlier years and their answer keys are available online to teachers and students and are often used as study materials.

Teachers' grades are based on several coursework elements that include in-class tests but also homework, oral presentations, class participation, and student behavior. The weight of the different elements varies across schools and subjects. In exam years it is known that teachers prepare their students for the exams and usually written tests follow the exams' model. Although there is no regulation on the number of written tests, it is most common to have two tests per trimester leading to around six written tests per year. In the Portuguese educational system curving is not used, and as in national exams, teacher grades are based on absolute criteria.

[3]The ID-number that identifies the exam is not known to the student.

A student can take the national exam in a subject as an internal or as an external student. To be considered an internal student it must be the case that he/she attended classes in that school for the subject on which he/she takes the national exam. Moreover, students must obtain a final score from teacher grading of at least 10 (on a scale of 1 to 20) to be able to take the exam as internal students. External students are any others.[4] For students who take exams as internal students, 30% of the final score in a subject is determined by his/her exam score, and the remaining 70% is determined by teacher grading. Access to higher education depends on the weighted average of high-school GPA and scores on national exams, where the weight of the latter varies between 40% and 50% depending on the university and the degree.

4 Data and Descriptive Statistics

The sources of the data used in this study are databases managed by the Portuguese Ministry of Education. One set of databases (JNE) includes student results on national exams, age, gender, nationality and school attended. The other database used (MISI) is a rich set of administrative data at the individual level. From MISI we have information regarding student grade and track of studies, socio-economic background, and school and teacher characteristics.[5]

The databases allow us to link each student's score in the national exam with the score in the same subject attributed by her/his teacher. We also have access to the scores obtained by the

[4]Students can cancel their enrolment in a subject and take the national exams as external students, in which case their final score is determined entirely by the exam score. Unfortunately, the data do not allow us to identify the students who cancel their enrolment.

[5]The information is anonymous and available for students attending public schools in the Portuguese mainland.

student in the 9th grade national exams of Portuguese Language and Mathematics.

Table 1 shows the values of the variables used for the sample of students in our analysis. These students are the ones enrolled in the academic track of secondary education for which we can match a score in a national exam with the respective teacher-assigned grade. In terms of the variables considered, our sample characteristics are in accordance with the values for the population of students in the academic track. Girls account for almost 60% of the students. At this level of education 24% of the students benefit from social support (government subsidies) and almost 40% come from families in which at least one of the parents has a level of education equal to or above upper secondary.

Table 1: Socio-economic background of students enrolled in the academic track of secondary education in the school years 2008/09 to 2016/17 for which we identify the assigned teacher for at least one of the courses in which the student takes a national exam.

	Mean			Std.Dev.			Min		Max	
Variable	Overall	Boys	Girls	Overall	Boys	Girls	Boys	Girls	Boys	Girls
Girl	0.57	-	-	0.49	-	-	-	-	-	-
Age	16.65	16.67	16.64	0.80	0.82	0.78	12.8	12.9	21.7	21.7
Social Support Beneficiary	0.26	0.23	0.29	0.44	0.42	0.45	0	0	1	1
Portuguese Born	0.95	0.95	0.95	0.22	0.21	0.23	0	0	1	1
Portuguese National	0.97	0.98	0.97	0.16	0.15	0.16	0	0	1	1
Average 9th grade Exam Scores										
Portuguese Language	61.98	60.06	63.44	14.59	14.48	14.51	0	1	100	100
Mathematics	58.07	60.06	56.56	22.71	22.01	23.07	0	0	100	100
Mother Education										
College	0.21	0.24	0.19	0.41	0.43	0.39	0	0	1	1
Upper Secondary	0.20	0.21	0.20	0.40	0.40	0.40	0	0	1	1
Father Education										
College	0.15	0.18	0.14	0.36	0.38	0.34	0	0	1	1
Upper Secondary	0.19	0.20	0.18	0.39	0.40	0.38	0	0	1	1
Computer at home	0.71	0.71	0.71	0.45	0.45	0.45	0	0	1	1
Internet at home	0.65	0.66	0.65	0.47	0.47	0.48	0	0	1	1
Number of students	442,853	190,612	252,241							

In **Table 2** we summarize the variables available to characterize teachers. As in most countries, in the Portuguese Educational System female teachers comprise the majority, with a share of

75%. The average teacher is in her/his late 40's with teaching experience of around 20 years. Most teachers have a bachelor's degree and only a few have undertaken graduate studies.

Table 2: Characterization of the teachers in the sample.

	Mean			Std.Dev.			Min		Max	
Variable	Overall	Male	Female	Overall	Male	Female	Male	Female	Male	Female
Female Teacher	0.75	-	-	0.43	-	-	-	-	-	-
Age	46.54	47.44	46.24	7.75	7.71	7.74	24.7	22.9	67	69.2
Tenure (in years)	19.70	19.93	19.62	9.64	9.52	9.68	0	0	42.92	41.74
Education										
Phd	0.01	0.01	0.01	0.09	0.12	0.08	0	0	1	1
Masters	0.12	0.14	0.11	0.32	0.35	0.31	0	0	1	1
Bachelor's	0.86	0.82	0.87	0.35	0.38	0.34	0	0	1	1
2 Year College Programs	0.01	0.01	0.01	0.12	0.15	0.11	0	0	1	1
Number of teachers	21,444	5,433	16,011							

In **Table 3** we see that the assessment gap (Teacher Score - Exam Score) is positive in all subjects for both boys and girls (columns 9 and 10), meaning that on average students lower their score on National Exams with respect to the score assigned by their teacher. Also, whenever significant, girls' assessment gap is on average greater than boys' (column 11). It is worth mentioning that this bias is less pronounced in the usually analyzed subjects of the curriculum, Native Language (here Portuguese) and Mathematics, which is an indication that it may be important to include other subjects in the analysis.

Regarding the differences between scores obtained by students when taught by a male or a female teacher, it is clear from the analysis of **Table 3** that students assigned to a male teacher obtain on average lower scores on National Exams than the grades that were assigned by a female teacher (columns 14 and 15). Moreover, whenever the difference between the assessment gap of students taught by female or male teachers is significant it is smaller for female teachers (column 18), meaning that students with female teachers lower their scores on National Exams less than do students with male teachers. This is observed for the majority of the subjects analyzed.

Given that in some subjects the number of male teachers was too low to study the impact of teacher gender we were forced to restrict our sample concerning the subjects analyzed. To include a subject we required that for each year it was taught by at least 50 male teachers. Based on this criterion we exclude five subjects from the analysis. These subjects coincide with those for which the yearly average of the number of exams taken by students is below 2000. Thus, our analysis herein focuses on the 12 subjects from **Table 3**, still a broader set than in most previous studies.

Table 3: Means and estimated standard errors of teachers' and national exams' scores in secondary education (2009-2017) - Mean comparison tests by student gender and by teacher gender

				Students							Teachers						
	Number of			Mean Teacher Score		Mean Exam Score		Mean Difference		T test a)	Mean Teacher Score		Mean Exam Score		Mean Difference		T test b)
Subject	Observations	Schools	% Girls	Boys	Girls	Boys	Girls	Boys	Girls	[(10)-(9)]	Male	Female	Male	Female	Male	Female	[(17)-(16)]
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)	(15)	(16)	(17)	(18)
Humanities																	
Geography	132960	451	63	13.0	13.2	11.2	10.6	1.8	2.6	0.8***	13.2	13.1	10.7	10.9	2.5	2.2	-0.3**
				(0.03)	(0.03)	(0.04)	(0.04)	(0.08)	(0.06)	[7.96]	(0.06)	(0.03)	(0.06)	(0.05)	(0.09)	(0.05)	[-2.79]
History	102971	439	70	12.6	13.0	10.5	10.5	2.0	2.5	0.5***	12.8	12.9	10.2	10.6	2.6	2.2	-0.4**
				(0.03)	(0.03)	(0.06)	(0.05)	(0.11)	(0.07)	[3.60]	(0.04)	(0.03)	(0.08)	(0.06)	(0.10)	(0.07)	[-3.22]
History of Culture and Arts	22429	223	68	12.6	13.3	9.5	10.1	3.0	3.2	0.1	13.0	13.1	9.6	10.1	3.3	3.0	-0.3
				(0.05)	(0.06)	(0.09)	(0.08)	(0.17)	(0.12)	[0.60]	(0.07)	(0.06)	(0.12)	(0.10)	(0.16)	(0.12)	[-1.62]
Philosophy	46689	472	65	13.2	13.9	9.5	10.4	3.7	3.5	-0.2	13.6	13.7	9.9	10.2	3.7	3.5	-0.2
				(0.04)	(0.04)	(0.09)	(0.07)	(0.13)	(0.10)	[-1.14]	(0.05)	(0.04)	(0.10)	(0.09)	(0.14)	(0.10)	[-1.03]
Portuguese	454237	475	58	12.8	13.6	10.0	11.0	2.7	2.7	-0.1	13.4	13.2	10.5	10.6	2.9	2.7	-0.2*
				(0.03)	(0.03)	(0.04)	(0.04)	(0.05)	(0.05)	[-1.20]	(0.05)	(0.03)	(0.05)	(0.04)	(0.09)	(0.04)	[-2.28]
Sciences																	
Biology and Geology	286245	472	56	13.5	13.8	9.6	9.7	3.9	4.1	0.2**	13.7	13.6	9.5	9.7	4.2	3.9	-0.2*
				(0.03)	(0.03)	(0.04)	(0.04)	(0.06)	(0.05)	[2.96]	(0.05)	(0.03)	(0.05)	(0.04)	(0.09)	(0.04)	[-2.44]
Descriptive Geometry	51705	328	50	14.4	14.1	11.0	9.6	3.5	4.4	0.9**	14.2	14.3	10.0	10.8	4.2	3.5	-0.7*
				(0.05)	(0.05)	(0.16)	(0.15)	(0.20)	(0.21)	[3.24]	(0.06)	(0.06)	(0.17)	(0.21)	(0.19)	(0.22)	[-2.40]
Mathematics	308026	473	53	12.9	13.3	10.1	10.5	2.7	2.8	0.0	13.0	13.1	10.1	10.4	2.9	2.7	-0.28*
				(0.03)	(0.03)	(0.07)	(0.07)	(0.09)	(0.08)	[0.37]	(0.04)	(0.03)	(0.09)	(0.08)	(0.11)	(0.07)	[-2.09]
Mathematics SS	61874	408	70	12.7	13.4	10.1	10.7	2.6	2.7	0.1	13.1	13.2	10.2	10.6	2.9	2.6	-0.3
				(0.03)	(0.04)	(0.08)	(0.07)	(0.13)	(0.08)	[0.99]	(0.059	(0.03)	(0.11)	(0.08)	(0.13)	(0.08)	[-1.82]
Physics and Chemistry	300917	471	52	13.0	13.3	8.9	9.0	4.0	4.3	0.3**	13.2	13.2	8.8	9.0	4.4	4.1	-0.3*
				(0.03)	(0.03)	(0.06)	(0.06)	(0.07)	(0.07)	[2.82]	(0.04)	(0.03)	(0.08)	(0.06)	(0.10)	(0.06)	[-2.34]
Other																	
Drawing	34006	226	67	14.6	15.3	12.4	12.8	2.2	2.6	0.4*	15.1	15.1	12.6	12.7	2.5	2.4	-0.1
				(0.05)	(0.05)	(0.06)	(0.06)	(0.12)	(0.09)	[2.39]	(0.07)	(0.05)	(0.08)	(0.07)	(0.11)	(0.09)	[-0.64]
Economics	46394	317	48	13.8	14.3	11.5	11.6	2.3	2.6	0.3	14.1	14.0	11.3	11.7	2.7	2.3	-0.4*
				(0.05)	(0.06)	(0.07)	(0.08)	(0.12)	(0.12)	[1.88]	(0.08)	(0.06)	(0.10)	(0.09)	(0.14)	(0.11)	[-2.30]

Note. The grading scale is 0 to 20. The mean difference in columns (9) and (10) is the assessment gap, defined as mean teacher score minus mean exam score. a) In column (11) the difference tested is column (10) minus column (9), girls' assessment gap minus boys' assessment gap. b) In column (18) the difference tested is column (17) minus column (16), students with female teachers' assessment gap minus students with male teachers' assessment gap. The T test is for the difference in mean differences between girls' and boys' scores and the T statistic in square brackets reflects standard errors in parentheses that are corrected for clustering at the school level. The number of clusters corresponds to the number of schools in column (3). The significance of the assessment gap was estimated using Stata's clttest command.
* significant at $p < 0.05$, ** significant at $p < 0.01$, *** significant at $p < 0.001$

5 Empirical Results

5.1 Empirical strategy

We assume a linear specification for the assessment gap equation to estimate the following model.

$$G_{ijkst} = \alpha + \beta B_i + \delta M_k + \tau B_i M_k + \rho X_i + \nu T_k + \gamma_s + \mu_t + \epsilon_{ijt} \quad (1)$$

where G_{ijkst} is the gap between assessments for student i in subject j with teacher k at school s and time t. G_{ijkst} is assumed to be a function of the student's and the teacher's gender. $B_i = 1$ if the student is a boy and $B_i = 0$ otherwise and $M_k = 1$ if the teacher is a man and $M_k = 0$ if the teacher is a woman. The interaction term $B_i M_k$ is thus equal to 1 for boys being taught by a male teacher.

The model considers a vector of co-variates X_i including students' observable characteristics and socio-economic background. The variables considered are the student's age, gender, and country of birth. We control for students' ability by including the scores obtained on the 9th grade national exams of Portuguese and mathematics. Also, to account for socio-economic background we include the student's social support beneficiary status, availability of computer and internet at home, parents' education, and parents' employment status. We also take into account if the student is taking the national exam in order to apply for university. T_k is a vector of teacher characteristics, namely age, and education. The model includes school and year fixed effects.

Given that the dependent variable is a difference between scores in the same subject for the same student, the model accounts for the unobserved student characteristics that affect the two components of the dependent variable similarly, the teacher and the exam scores. This is our main identification strategy. **Table 4** summarizes the coefficients of interest, taking as the baseline a girl with a female teacher.

To account for the potential selection into the sample that may arise from considering only internal students, we will also estimate a Heckman selection model in which the outcome equation is the one for Model 1 and the selection equation is as follows.

$$Prob(Internal) = \alpha + \beta B_i + \delta M_k + \tau B_i M_k + \rho X_i + \nu T_k + \psi Enrolled_{it} + \mu_t + \tau_s + \epsilon_{ijt} \quad (2)$$

For identification purposes we included the indicator variable $Enrolled_{it}$. This variable takes the value 1 if student i took at least one national exam as internal student in year t, and takes the value 0 otherwise. Students may decide to take one exam as internal and the other as external to improve their grade, or be forced to take one exam as external if the teacher grade is below 10.

Table 4: Effects schematics.

	Girl	Boy
Female Teacher	Default	β
Male Teacher	δ	$\beta + \delta + \tau$

5.2 Results

Table 5 shows the results for Model 1. The model was estimated using ordinary least squares pooling all observations from the available repeated cross sections for each subject. The estimates presented are obtained using school and year fixed effects and control for student, family, and teacher characteristics.

In column 5 we present the estimated coefficient for β, the difference between boys' and girls' assessment gap when they are both assigned to a female teacher. Column 6 shows the estimated impact on girls' assessment gap of having a male instead of a female teacher, δ. The coefficient associated to the interaction term, B_iM_k, is shown in column 7. In column 8 we have the estimated assessment gap for a boy of being assigned to a male instead of a female teacher, given by $(\delta + \tau)$. The total effect in column 9 concerns the difference between the assessment gap of a girl assigned to a female teacher and a boy assigned to a male teacher.

We obtain that the assessment gap is on average smaller for boys than for girls in almost every subject, indicating that boys lower their scores on national exams with respect to teacher assigned scores less than do girls. This is the case for both Humanities and Sciences' subjects. For the default case of female teachers this is given by β, and the size of the effect ranges from -0.8 in Geography to -0.09 in Mathematics A.[6] For the case of male teachers we should look at $(\beta + \tau)$, but as τ is always negative except for drawing, in which case it is smaller than β, the conclusion is maintained and even reinforced.

Regarding teacher gender, the estimates obtained for δ, column 6, indicate that a student with a

[6] Philosophy is the only subject for which the effect is positive, though with a significance level of only 0.05.

Table 5: Estimated Assessment Gap in Secondary Education, by subject. Dependent variable is the difference $TeacherScore - ExamScore$.

Subject Type/Subject	Number of Observations	% Boys	% Male Teachers	Boy	Male Teacher	Boy with Male Teacher		
				β	δ	τ	$\delta+\tau$	$\beta+\delta+\tau$
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Humanities								
Geography	126,771	37.1%	26.2%	-0.80***	0.20***	-0.06*	0.14***	-0.67***
				(0.014)	(0.029)	(0.028)	(0.031)	(0.031)
History A	90,215	29.6%	34.1%	-0.62***	0.22***	-0.08*	0.14*	-0.48***
				(0.023)	(0.049)	(0.039)	(0.055)	(0.053)
History of Culture & Arts	19,332	31.5%	34.7%	-0.24***	-0.35	-0.11	-0.45	-0.69
				(0.048)	(0.358)	(0.082)	(0.361)	(0.360)
Philosophy	45,682	34.9%	32.4%	0.08*	0.18***	0.00	0.18**	0.26***
				(0.032)	(0.047)	(0.057)	(0.057)	(0.054)
Portuguese	394,318	41.5%	14.6%	-0.00	0.13***	-0.06**	0.06**	0.06**
				(0.008)	(0.016)	(0.021)	(0.018)	(0.018)
Sciences								
Biology & Geology	276,398	43.5%	20.2%	-0.25***	0.20***	-0.08***	0.12***	-0.13***
				(0.009)	(0.016)	(0.020)	(0.018)	(0.018)
Descriptive Geometry	46,883	49.8%	55.8%	-0.64***	0.19*	-0.09	0.10	-0.53***
				(0.049)	(0.085)	(0.066)	(0.087)	(0.086)
Mathematics A	270,047	47.5%	27.4%	-0.09***	0.17***	-0.10***	0.07**	-0.02
				(0.012)	(0.019)	(0.022)	(0.020)	(0.020)
Mathematics Applied to SS	57,748	29.5%	26.4%	-0.42***	0.14	-0.15**	-0.00	-0.42***
				(0.028)	(0.097)	(0.056)	(0.103)	(0.101)
Physics & Chemistry	284,347	47.7%	23.4%	-0.28***	0.22***	-0.09***	0.15***	-0.15***
				(0.009)	(0.016)	(0.019)	(0.017)	(0.016)
Other								
Drawing	28,804	32.0%	39.3%	-0.53***	0.41***	0.12*	0.54***	0.00
				(0.039)	(0.091)	(0.062)	(0.098)	(0.094)
Economics	43,246	52.3%	34.2%	-0.44***	0.82***	-0.15**	0.67***	0.23*
				(0.026)	(0.109)	(0.045)	(0.108)	(0.109)

Note. The model includes school and year fixed effects and controls for student, family, and teacher characteristics. Standard errors in parentheses are robust. Significance levels are *** p<0.001, ** p<0.01, * p<0.05.

male teacher has on average a larger gap between assessments than one with a female teacher, meaning that on average students with male teachers lower their score on the exam with respect to teacher grading more than students of female teachers. The size of the impact ranges from 0.13 in Portuguese Language to 0.82 in Economics and is non-significant only in History of Culture and Arts.

Although being assigned to a male teacher increases students' assessment gap, the impact is greater for girls than for boys, as can be seen from the sign of the estimated value of τ. Whenever significant, the parameter takes on negative values in all but one of the subjects analyzed, still with only a 0.05 significance.

Taking advantage of the fact that each student is enrolled and evaluated in several subjects, possibly with male teachers in some and female teachers in others, we extended our analysis including student fixed effects.

Table 6 presents results from the estimation of Equation 1 with student covariates replaced by student fixed effects, and considering all subjects simultaneously. Subject fixed effects are also considered. The model was estimated including all students (column 2), and also separately for boys and girls (columns 3 and 4, respectively). The results obtained confirm those in **Table 5**. Being assigned to a male instead of a female teacher increases the assessment gap for both boys and girls, the effect being higher for the latter.

The results shown in **Table A1** in the Appendix confirm those obtained when selection into the sample is taken into account. Not only is the sign of the estimated effects maintained, but also its magnitude.

Table 6: Estimated Assessment Gap in Secondary Education, considering student, subject, and year of exam fixed effects. Dependent variable is the difference $TeacherScore - ExamScore$. Results from full regression and for separate regressions for boys and girls.

Variable	Coefficient			Confidence Intervall		
	Overall	Boys	Girls	Overall	Boys	Girls
(1)	(2)	(3)	(4)	(5)	(6)	(7)
Teacher						
Male	0.27***	0.15***	0.24***	[0.22,0.32]	[0.09,0.20]	[0.19,0.29]
	(0.026)	(0.029)	(0.026)			
Age	0.00	0.00	0.00	[-0.00,0.00]	[-0.00,0.01]	[-0.00,0.00]
	(0.002)	(0.002)	(0.002)			
Education						
Master	-0.02	0.01	-0.05	[-0.21,0.17]	[-0.21,0.24]	[-0.23,0.14]
	(0.096)	(0.113)	(0.094)			
Phd	0.14	0.15	0.13	[-0.14,0.42]	[-0.16,0.45]	[-0.14,0.41]
	(0.141)	(0.156)	(0.139)			
Student						
University Application	-0.25***	-0.23***	-0.24***	[-0.33,-0.16]	[-0.33,-0.12]	[-0.34,-0.14]
	(0.041)	(0.053)	(0.049)			
Boy with Male Teacher	-0.16***	-	-	[-0.20,-0.12]	-	-
	(0.019)					
Constant	3.31***	3.19***	3.38***	[3.03,3.58]	[2.87,3.50]	[3.10,3.66]
	(0.140)	(0.160)	(0.143)			

Note. The model includes fixed effects for student, subject and year. Standard errors are robust. Significance levels are *** p<0.001, ** p<0.01, * p<0.05.

Table 7: Mechanism - Dependent variable is Teacher Score.

Subject Type/Subject	Grade	Number of Observations	% Boys	% Male Teachers	Boy	Male Teacher	Boy with Male Teacher	Total Effect
					β	δ	τ	$\beta+\delta+\tau$
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Humanities								
Geography	11	126,771	37.1%	26.2%	-0.12***	0.04	0.07**	-0.01
					(0.012)	(0.024)	(0.023)	(0.026)
History A	12	90,215	29.6%	34.1%	-0.15***	-0.11**	0.09**	-0.18***
					(0.018)	(0.037)	(0.030)	(0.040)
History of Culture & Arts	11	19,332	31.5%	34.7%	-0.45***	0.50*	0.13	0.15
					(0.040)	(0.249)	(0.065)	(0.261)
Philosophy	11	45,682	34.9%	32.4%	-0.69***	-0.10**	0.06	-0.73***
					(0.023)	(0.032)	(0.039)	(0.037)
Portuguese	12	394,318	41.5%	14.6%	-0.78***	0.02	-0.00	-0.77***
					(0.005)	(0.011)	(0.014)	(0.012)
Sciences								
Biology & Geology	11	276,398	43.5%	20.2%	-0.12***	-0.03*	0.12***	-0.03*
					(0.008)	(0.013)	(0.016)	(0.015)
Descriptive Geometry	11	46,883	49.8%	55.8%	-0.12**	0.07	0.07	0.03
					(0.036)	(0.060)	(0.047)	(0.060)
Mathematics A	12	270,047	47.5%	27.4%	-0.31***	-0.01	0.04*	-0.28***
					(0.009)	(0.015)	(0.018)	(0.016)
Mathematics Applied to SS	11	57,748	29.5%	26.4%	-0.46***	-0.38***	0.08*	-0.76***
					(0.022)	(0.070)	(0.041)	(0.073)
Physics & Chemistry	11	284,347	47.7%	23.4%	-0.20***	0.01	0.09***	-0.10***
					(0.008)	(0.014)	(0.017)	(0.014)
Other								
Drawing	12	28,804	32.0%	39.3%	-0.67***	0.13	0.16**	-0.37***
					(0.032)	(0.069)	(0.049)	(0.073)
Economics	11	43,246	52.3%	34.2%	-0.32***	0.32***	0.09*	0.09
					(0.023)	(0.093)	(0.038)	(0.093)

Note. The model includes school and year fixed effects and controls for student, family, and teacher characteristics. Standard errors in parentheses are robust. Significance levels are *** p<0.001, ** p<0.01, * p<0.05.

5.3 Mechanism

Our results indicate that students with male teachers tend to lower their scores on national exams with respect to scores obtained from teacher assessment more than students who are assigned to female teachers. This may be explained either by male teachers assigning on average higher scores to their students or by male teachers' students obtaining on average lower scores on national exams.

In an attempt to understand the mechanism behind the results we analyze the scores from the two grading systems separately by replacing the outcome variable in **Model 1** with the score from each type of assessment.

Results in **Table 7** column 7 show that there is a significant difference between female and male teachers when assigning grades to their students in only 6 of the 12 courses analyzed. Moreover, the sign is positive in two of them and negative for all the rest. Thus, there is no clear association between teacher gender and teacher scores.

However, when the dependent variable is the exam score, we see in **Table 8** that male teachers' students obtain on average lower scores than students who were taught by a female teacher. This negative relationship is found for all but two of the subjects and the size of the coefficient ranges from -0.11 to -0.53. The difference between scores is not significant for only Descriptive Geometry, and has a positive sign for History of Culture and Arts, again only at a 0.05 significance level. These results suggest that the fact that students with male teachers tend to lower their scores on national exams with respect to scores obtained from teacher assessment more than students who are assigned to female teachers can be explained by male teachers' students obtaining on average lower scores on national exams.

Table 8: Mechanism - Dependent variable is the Exam Score

Subject Type/Subject	Grade	Number of Observations	% Boys	% Male Teachers	Boy	Male Teacher	Boy with Male Teacher	Total Effect
					β	δ	τ	$\beta+\delta+\tau$
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Humanities								
Geography	11	126,771	37.1%	26.2%	0.68***	-0.16***	0.13***	0.66***
					(0.016)	(0.031)	(0.030)	(0.034)
History A	12	90,215	29.6%	34.1%	0.47***	-0.34***	0.17***	0.30***
					(0.027)	(0.058)	(0.046)	(0.063)
History of Culture&Arts	11	19,332	31.5%	34.7%	-0.22***	0.79*	0.24*	0.80*
					(0.056)	(0.392)	(0.094)	(0.394)
Philosophy	11	45,682	34.9.0%	32.4%	-0.77***	-0.28***	0.06	-0.99***
					(0.038)	(0.054)	(0.066)	(0.062)
Portuguese	12	394,318	41.5%	14.6%	-0.78***	-0.11***	0.06	-0.83***
					(0.009)	(0.017)	(0.023)	(0.020)
Sciences								
Biology&Geology	11	276,398	43.5%	20.2%	0.13***	-0.23***	0.20***	0.10***
					(0.011)	(0.020)	(0.025)	(0.022)
Descriptive Geometry	11	46,883	49.8%	55.8%	0.52***	-0.12	0.16	0.56***
					(0.061)	(0.103)	(0.083)	(0.104)
Mathematics A	12	270,047	47.5%	27.4%	-0.22***	-0.19***	0.14***	-0.27***
					(0.016)	(0.026)	(0.030)	(0.027)
Mathematics Applied to SS	11	57,748	29.5%	26.4%	-0.05	-0.53***	0.23***	-0.35**
					(0.033)	(0.108)	(0.064)	(0.112)
Physics&Chemistry	11	284,347	47.7%	23.4%	0.08***	-0.23***	0.17***	0.02
					(0.013)	(0.021)	(0.026)	(0.022)
Other								
Drawing	12	28,804	32.0%	39.3%	-0.13**	-0.28**	0.03	-0.37***
					(0.043)	(0.099)	(0.068)	(0.103)
Economics	11	43,246	52.3%	34.2%	0.12***	-0.47***	0.24***	-0.11
					(0.033)	(0.133)	(0.056)	(0.132)

Note. The model includes school and year fixed effects and controls for student, family, and teacher characteristics. Standard errors in parentheses are robust. Significance levels are *** $p<0.001$, ** $p<0.01$, * $p<0.05$.

Table 9: Mechanism - Results from full regression and for separate regressions for boys and girls, considering student, subject, and year of exam fixed effects.

Independent/Dependent Variable	Teacher Score			Exam Score		
	Overall	Boys	Girls	Overall	Boys	Girls
(1)	(2)	(3)	(4)	(5)	(6)	(7)
Teacher						
Male	-0.05***	0.08***	-0.00	-0.31***	-0.07***	-0.24***
	(0.003)	(0.004)	(0.003)	(0.026)	(0.008)	(0.034)
Age	-0.01***	-0.01***	-0.01	-0.01***	-0.01***	-0.01***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.001)	(0.000)
Education						
Master	0.07***	0.09***	0.06***	0.09***	0.09**	0.10***
	(0.010)	(0.015)	(0.014)	(0.021)	(0.032)	(0.028)
Phd	0.11***	0.14***	0.10***	-0.02	0.00	-0.03
	(0.014)	(0.022)	(0.019)	(0.029)	(0.045)	(0.038)
Student						
University Application	0.26***	0.22***	0.26***	0.51***	0.45***	0.50***
	(0.012)	(0.018)	(0.016)	(0.026)	(0.039)	(0.034)
Boy with Male Teacher	0.14***	-	-	0.04***	-	-
	(0.004)			(0.009)		
Constant	12.97***	12.42***	13.33***	8.06***	7.54***	8.33***
	(0.024)	(0.035)	(0.032)	(0.051)	(0.076)	(0.067)

Note. The model includes fixed effects for student, subject and year. Standard errors are robust. Significance levels are *** p<0.001, ** p<0.01, * p<0.05.

The allocation of students to male or female teachers could depend on unobservable characteristics and bias the results in **Tables 7** and **8**. Note that here our dependent variable is not a difference between scores as in the previous analysis, and so we are not in any way controlling for students unobserved characteristics. Including students' fixed effects as we do in **Table 9** increases our confidence in the results obtained.

Table 9 presents results from the estimation of the specifications from **Tables 7** and **8** with student covariates replaced by student fixed effects, and considering all subjects simultaneously. Subject fixed effects are also considered. The models were estimated including all students (columns 2 and 5), and also separately for boys and girls (columns 3 and 4, and 6 and 7 respectively). The results obtained confirm those in **Table 8**. Students assigned to male teachers obtain lower scores on exams than students assigned to female teachers, and this negative effect is greater for girls. With respect to teacher scores, overall results in **Table 9** show a negative (although small) effect from having a male teacher (column 2). Being assigned to a male instead of a female teacher increases boys' assigned grades while having no effect on those of girls'.

6 Concluding remarks

This paper looks at how grading practices affect boys' and girls' scores differently depending on teacher gender. We use scores obtained from two different types of assessment, the one that is carried out by the students' teachers and national exams, to analyze how the difference between these two types of assessment depends on gender of the student, of the teacher, and of the match between them.

We contribute to the literature on grading systems and gender gaps by testing how the gender difference in the difference between teacher grading and national exam scores is affected by teacher gender. We use Portuguese data on 12 subjects across humanities and sciences for all public schools' students taking exams at the end of secondary education from 2009 to 2017.

Our results show that the assessment gap is on average positive and higher for girls, meaning that on average students lower their scores on national exams with respect to their teacher-assigned grades, and this effect is greater for girls. Also, these findings hold regardless of teacher gender. Thus, exams are relatively more favorable for boys and teacher grading is more favorable for girls. We also obtain that students assigned to male teachers have a greater assessment gap than students who are assigned to female teachers, meaning that their exam scores fall more with respect to teacher grades when they are assigned to a male teacher. Our analysis suggests that this is mainly due to the fact that students assigned to male teachers obtain lower scores on national exams than students assigned to female teachers and not to the existence of systematic differences between male and female teachers in grading their students. The results show that the assessment gap is lower for students who are assigned to a female teacher, and the difference is larger for girls.

Teacher gender and its matching with student gender could be an explanation for the gender difference in the assessment gap. Moreover, the fact that the majority of teachers are female could contribute to a lower teacher assessment for boys. However, what we obtain is that the assessment gap is always greater for girls, regardless of teachers' gender and gender matching. Even when we compare boys with male teachers to girls with female teachers the result holds, although the difference is less. Thus, exams are relatively more favorable to boys whatever the teacher gender or the gender matching.

For educational systems in which teacher grading plays an extremely important role in students' academic success our findings show that it may be growing harder for boys to level their scores with those of girls', thereby constraining their choices regarding academic path and their performance. The choice of the right combination of weights attributed to exams and teacher assessment is important for gender balance in both higher education and the labor market. Also, as in the Swedish case (Karlsson and Wikström (2022)), allowing different students to make different choices regarding evaluation methods could be an interesting solution and one that deserves to be explored.

References

Ambady, N., Shih, M., Kim, A., and Pittinsky, T. L. (2001). Stereotype susceptibility in children: Effects of identity activation on quantitative performance. *Psychological Science*, 12(5):385–390. PMID: 11554671.

Ammermueller, A. and Dolton, P. (2006). Pupil-Teacher Gender Interaction Effects on Scholastic Outcomes in England and the USA. *ZEW - Centre for European Economic Research Discussion Paper No. 06-060*.

Angelo, C. and Reis, A. B. (2020). Gender gaps in different grading systems. *Education Economics*, 0(0):1–15.

Azmat, G., Calsamiglia, C., and Iriberri, N. (2016). Gender differences in response to big stakes. *Journal of the European Economic Association*, 14(6):1372–1400.

Cai, X., Lu, Y., Pan, J., and Zhong, S. (2018). Gender gap under pressure: Evidence from China's national college entrance examination. *The Review of Economics and Statistics*, 101.

Carlana, M. (2019). Implicit stereotypes: Evidence from teachers' gender bias. *The Quarterly Journal of Economics*, 134(3):1163–1224.

Carrington, B., Tymms, P., and Merrell, C. (2008). Role models, school improvement and the 'gender gap' - do men bring out the best in boys and women the best in girls? *British Educational Research Journal*, 34:315–327.

Cho, I. (2012). The effect of teacher–student gender matching: Evidence from OECD countries. *Economics of Education Review*, 31(3):54–67.

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. *Journal of Human Resources*, Volume: 48(1):236–264.

Czibor, E., Onderstal, S., Sloof, R., and Praag, M. (2014). Does relative grading help male students? evidence from a field experiment in the classroom. Tinbergen Institute discussion paper 14-116/v. *SSRN Electronic Journal*.

Dee, T. S. (2007). Teachers and the Gender Gaps in Student Achievement. *Journal of Human Resources*, (September 2006).

Ehrenberg, R. G., Goldhaber, D. D., and Brewer, D. J. (1994). Do Teachers' Race, Gender, and Ethnicity Matter?: Evidence from NELS88. NBER Working Papers 4669, National Bureau of Economic Research, Inc.

Falch, T. and Naper, L. R. (2013). Educational evaluation schemes and gender gaps in student achievement. *Economics of Education Review*, 36:12–25.

Francis, B., Skelton, C., Carrington, B., Hutchings, M., Read, B., and Hall, I. (2008). A perfect match? pupils' and teachers' views of the impact of matching educators and learners by gender. *Research Papers in Education*, 23:21–36.

Hinnerich, B. T., Höglin, E., and Johannesson, M. (2011). Are boys discriminated in Swedish high schools? *Economics of Education Review*, 30(4):pp.682–690.

Hyde, J. S., Fennema, E., Ryan, M., Frost, L. A., and Hopp, C. (1990). Gender comparisons of mathematics attitudes and affect: A meta-analysis. *Psychology of Women Quarterly*, 14(3):299–324.

Jurajda, S. and Münich, D. (2011). Gender gap in performance under competitive pressure: Admissions to Czech universities. *American Economic Review*, 101(3):514–18.

Karlsson, L. and Wikström, M. (2022). Admission groups and academic performance: A study of marginal entrants in the selection to higher education. *The B.E. Journal of Economic Analysis & Policy*.

Krieg, J. (2005). Student gender and teacher gender: What is the impact on high stakes test scores? *Current Issues in Education*, 8.

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

Lavy, V. and Sand, E. (2018). On the origins of gender gaps in human capital: Short- and long-term consequences of teachers' biases. *Journal of Public Economics*, 167:263–279.

Lindahl, E. (2016). Are teacher assessments biased? Evidence from Sweden. *Education Economics*, 24(3):224–238.

Machin, S. and McNally, S. (2005). Gender and Student Achievement in English Schools. *Oxford Review of Economic Policy*, 21(3):357–372.

Marcenaro-Gutierrez, O. and Vignoles, A. (2015). A comparison of teacher and test-based assessment for Spanish primary and secondary students. *Educational Research*, 57(1):1–21.

Martin, C. L. and Ruble, D. (2004). Children's search for gender cues: Cognitive perspectives on gender development. *Current Directions in Psychological Science*, 13(2):67–70.

Nixon, L. A. and Robinson, M. D. (1999). The educational attainment of young women: Role model effects of female high school faculty. *Demography*, 36(2):185–194.

Ors, E., Palomino, F., and Peyrache, E. (2013). Performance gender gap: Does competition matter? *Journal of Labor Economics*, 31(3):443 – 499.

Paredes, V. (2014). A teacher like me or a student like me? Role model versus teacher bias effect. *Economics of Education Review*, 39:38–49.

Pekkarinen, T. (2015). Gender differences in behaviour under competitive pressure: Evidence on omission patterns in university entrance examinations. *Journal of Economic Behavior and Organization*, 115:94–110.

Rosenthal, R. and Jacobson, L. (1968). Pygmalion in the classroom. *The Urban Review*, 3(1):16–20.

Sansone, D. (2017). Why does teacher gender matter? *Economics of Education Review*, 61(May):9–18.

Terrier, C. (2005). Boys Lag Behind: How Teachers' Gender Biases Affect Student Achievement, IZA Discussion Paper No. 10343.

Winters, M. A., Haight, R. C., Swaim, T. T., and Pickering, K. A. (2013). The effect of same-gender teacher assignment on student achievement in the elementary and secondary grades: Evidence from panel data. *Economics of Education Review*.

References

Ambady, N., Shih, M., Kim, A., and Pittinsky, T. L. (2001). Stereotype susceptibility in children: Effects of identity activation on quantitative performance. *Psychological Science*, 12(5):385–390. PMID: 11554671.

Ammermueller, A. and Dolton, P. (2006). Pupil-Teacher Gender Interaction Effects on Scholastic Outcomes in England and the USA. *ZEW - Centre for European Economic Research Discussion Paper No. 06-060*.

Angelo, C. and Reis, A. B. (2020). Gender gaps in different grading systems. *Education Economics*, 0(0):1–15.

Azmat, G., Calsamiglia, C., and Iriberri, N. (2016). Gender differences in response to big stakes. *Journal of the European Economic Association*, 14(6):1372–1400.

Cai, X., Lu, Y., Pan, J., and Zhong, S. (2018). Gender gap under pressure: Evidence from China's national college entrance examination. *The Review of Economics and Statistics*, 101.

Carlana, M. (2019). Implicit stereotypes: Evidence from teachers' gender bias. *The Quarterly Journal of Economics*, 134(3):1163–1224.

Carrington, B., Tymms, P., and Merrell, C. (2008). Role models, school improvement and the 'gender gap' - do men bring out the best in boys and women the best in girls? *British Educational Research Journal*, 34:315–327.

Cho, I. (2012). The effect of teacher–student gender matching: Evidence from OECD countries. *Economics of Education Review*, 31(3):54–67.

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. *Journal of Human Resources*, Volume: 48(1):236–264.

Czibor, E., Onderstal, S., Sloof, R., and Praag, M. (2014). Does relative grading help male students? evidence from a field experiment in the classroom. Tinbergen Institute discussion paper 14-116/v. *SSRN Electronic Journal*.

Dee, T. S. (2007). Teachers and the Gender Gaps in Student Achievement. *Journal of Human Resources*, (September 2006).

Ehrenberg, R. G., Goldhaber, D. D., and Brewer, D. J. (1994). Do Teachers' Race, Gender, and Ethnicity Matter?: Evidence from NELS88. NBER Working Papers 4669, National Bureau of Economic Research, Inc.

Falch, T. and Naper, L. R. (2013). Educational evaluation schemes and gender gaps in student achievement. *Economics of Education Review*, 36:12–25.

Francis, B., Skelton, C., Carrington, B., Hutchings, M., Read, B., and Hall, I. (2008). A perfect match? pupils' and teachers' views of the impact of matching educators and learners by gender. *Research Papers in Education*, 23:21–36.

Hinnerich, B. T., Höglin, E., and Johannesson, M. (2011). Are boys discriminated in Swedish high schools? *Economics of Education Review*, 30(4):pp.682–690.

Hyde, J. S., Fennema, E., Ryan, M., Frost, L. A., and Hopp, C. (1990). Gender comparisons of mathematics attitudes and affect: A meta-analysis. *Psychology of Women Quarterly*, 14(3):299–324.

Jurajda, S. and Münich, D. (2011). Gender gap in performance under competitive pressure: Admissions to Czech universities. *American Economic Review*, 101(3):514–18.

Karlsson, L. and Wikström, M. (2022). Admission groups and academic performance: A study of marginal entrants in the selection to higher education. *The B.E. Journal of Economic Analysis & Policy*.

Krieg, J. (2005). Student gender and teacher gender: What is the impact on high stakes test scores? *Current Issues in Education*, 8.

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

Lavy, V. and Sand, E. (2018). On the origins of gender gaps in human capital: Short- and long-term consequences of teachers' biases. *Journal of Public Economics*, 167:263–279.

Lindahl, E. (2016). Are teacher assessments biased? Evidence from Sweden. *Education Economics*, 24(3):224–238.

Machin, S. and McNally, S. (2005). Gender and Student Achievement in English Schools. *Oxford Review of Economic Policy*, 21(3):357–372.

Marcenaro-Gutierrez, O. and Vignoles, A. (2015). A comparison of teacher and test-based assessment for Spanish primary and secondary students. *Educational Research*, 57(1):1–21.

Martin, C. L. and Ruble, D. (2004). Children's search for gender cues: Cognitive perspectives on gender development. *Current Directions in Psychological Science*, 13(2):67–70.

Nixon, L. A. and Robinson, M. D. (1999). The educational attainment of young women: Role model effects of female high school faculty. *Demography*, 36(2):185–194.

Ors, E., Palomino, F., and Peyrache, E. (2013). Performance gender gap: Does competition matter? *Journal of Labor Economics*, 31(3):443 – 499.

Paredes, V. (2014). A teacher like me or a student like me? Role model versus teacher bias effect. *Economics of Education Review*, 39:38–49.

Pekkarinen, T. (2015). Gender differences in behaviour under competitive pressure: Evidence on omission patterns in university entrance examinations. *Journal of Economic Behavior and Organization*, 115:94–110.

Rosenthal, R. and Jacobson, L. (1968). Pygmalion in the classroom. *The Urban Review*, 3(1):16–20.

Sansone, D. (2017). Why does teacher gender matter? *Economics of Education Review*, 61(May):9–18.

Terrier, C. (2005). Boys Lag Behind: How Teachers' Gender Biases Affect Student Achievement, IZA Discussion Paper No. 10343.

Winters, M. A., Haight, R. C., Swaim, T. T., and Pickering, K. A. (2013). The effect of same-gender teacher assignment on student achievement in the elementary and secondary grades: Evidence from panel data. *Economics of Education Review*.

Apendix

Table A1: Heckman selection model - Student/Teacher Gender Effect on the Assessment Gap

Subject Type/Subject	Grade	Observations	Lambda	Boy	Male Teacher	Interaction	Total Effect
Humanities							
Geography	11	130003	1.60***	-0.81***	0.25**	-0.05	-0.61***
			(0.063)	(0.024)	(0.076)	(0.044)	(0.071)
History A	12	93820	0.98*	-0.60***	0.43***	-0.09	-0.26**
			(0.447)	(0.033)	(0.093)	(0.054)	(0.092)
History of Culture&Arts	11	20298	0.08	-0.11	0.41*	-0.21	0.08
			(0.237)	(0.076)	(0.176)	(0.122)	(0.175)
Philosophy	11	53268	0.49	0.10	0.18	-0.06	0.23
			(0.464)	(0.061)	(0.098)	(0.073)	(0.123)
Portuguese	12	402307	0.10	0.01	0.26***	-0.04	0.23***
			(0.084)	(0.014)	(0.058)	(0.027)	(0.058)
Sciences							
Biology&Geology	11	281793	-0.96***	-0.25***	0.23***	-0.08**	-0.10*
			(0.075)	(0.014)	(0.050)	(0.028)	(0.049)
Descriptive Geometry	11	50138	0.08	-0.59***	0.66***	-0.19	-0.13
			(0.094)	(0.081)	(0.167)	(0.111)	(0.175)
Mathematics A	12	293494	-0.24***	-0.05**	0.21**	-0.10**	0.06
			(0.040)	(0.019)	(0.070)	(0.037)	(0.075)
Mathematics Applied to SS	11	62165	0.58***	-0.35***	0.29**	-0.19*	-0.25
			(0.158)	(0.038)	(0.108)	(0.077)	(0.130)
Physics&Chemistry	11	304520	-1.54***	-0.26***	0.22***	-0.11***	-0.15**
			(0.078)	(0.014)	(0.050)	(0.027)	(0.051)
Other							
Drawing	12	29019	-0.41**	-0.49***	0.10	0.16	-0.23*
			(0.143)	(0.053)	(0.104)	(0.085)	(0.110)
Economics	11	44060	-0.17	-0.41***	0.54***	-0.15*	-0.02
			(0.088)	(0.036)	(0.133)	(0.069)	(0.127)

Note. The model includes year fixed effects and controls for student, family, and teacher characteristics. Standard errors in parentheses are clustered at the school level. Significance levels are *** p<0.001, ** p<0.01, * p<0.05.

Nova School of Business & Economics
Campus de Carcavelos
Rua da Holanda 1
2775-405 Carcavelos | Portugal
novasbe.pt