

A Work Project, presented as part of the requirements for the Award of a Master's degree in Economics from the Nova School of Business and Economics.

THE SOCIOECONOMIC GAP BETWEEN TEACHER AND TEST-BASED
ASSESSMENTS

LUÍS MANUEL GOMES DIOGO LOPES RIBEIRO

Work project carried out under the supervision of:

Professor Ana Balcão Reis Professor Luís Catela Nunes Professor Maria do Carmo Seabra

28-05-2021

Abstract

The goal of this Master's thesis is to analyze whether teacher and test-based assessments differ systematically between low-income and high-income students throughout the Portuguese education system. Nationwide data on students and their assessments in 14 subjects in the 6th, 9th, 11th and 12th grades from 2008 to 2018 was analyzed following linear specifications to test whether there are significant socioeconomic gaps between both grading schemes. Results show all students are typically given worse results in exams than by their teachers, and low-income students tend to have a smaller gap between these two types of assessments.

Keywords: Economics of Education; Assessment; Socioeconomic Gap; Grading Schemes

This work used infrastructure and resources funded by Fundação para a Ciência e a Tecnologia (UID/ECO/00124/2013, UID/ECO/00124/2019 and Social Sciences DataLab, Project 22209), POR Lisboa (LISBOA-01-0145-FEDER-007722 and Social Sciences DataLab, Project 22209) and POR Norte (Social Sciences DataLab, Project 22209). I thank DGEEC (Direção Geral de Estatísticas da Educação e Ciência) for allowing access to the anonymized administrative education data analyzed in this research work project.

1. Introduction

The purpose of this thesis is to analyze whether different grading schemes generate a gap in student assessment by socioeconomic characteristics, using data on Portuguese public-school students and their teacher-given and exam scores, from 2008 to 2018. It follows Lindahl (2007), Marcenaro-Gutiérrez and Vignoles (2015), and Ângelo and Reis (2021) in the methodology and approach to find a significant assessment gap, however focusing on socioeconomic status instead of gender.

The findings of this study suggest that low-income students, identified as those who receive social support, effectively have lower scores overall. Moreover, in some subjects these students report smaller differences between the two types of assessments – teacher and exam score. The different grading schemes evaluate students differently, thus a careful consideration of findings such as these is needed when debating the criteria to access higher education for example.

The content is organized as follows: section 2 includes a review of some of the literature on education and socioeconomic status, and the impact of different grading schemes in student assessment; section 3 introduces the institutional setting of the analysis; section 4 dwells on the used data for the empirical study, as on its descriptive statistics and first observations; section 5 introduces the model to be estimated and its subsequent results; and section 6 presents some concluding remarks.

2. Literature Review

The relation between socioeconomic status and educational attainment has long been subject of research. Both as to whether educational achievement improves socioeconomic status and the other way around – whether socioeconomic status has a causal effect, and thus may be a predictor of, educational attainment.

The existence of a correlation between academic achievement and socioeconomic status had been widely accepted, yet its magnitude varied significantly in the existing literature by the time White (1982) conducted a meta-analysis of the research so far. This study concluded the correlation between SES and academic achievement was weak when analyzed at the individual level, but strong at the aggregate level of the school or neighborhood, suggesting people of similar SES tend to cluster. White (1982) also expressed concern for the measures of SES used in existing research, which included variables related to household environment that would have a direct impact on academic achievement without being immediately related to SES.

Johnson, McGue, and Iacono (2007) took advantage of a sample of adoptive and biological Minnesota families - the SIBS (Sibling Interaction and Behavior Study) – to quantify the role of environmental and parenting factors on educational achievement. This study proved SES had a small yet significant nonshared environmental impact on academic achievement measured as grades reported by the teachers, while controlling for gender, parenting practices, parental expectations for academic achievement, IQ, engagement in school, and genetic and environmental influences shared among siblings. According to Aitkens and Barbarin (2008), school and neighborhood conditions contribute even more than family background to disparities in reading learning rates by SES. Nevertheless, there is clear evidence that socioeconomic background, as well as school and teacher quality, greatly influences reading skills for example, especially at an early stage of education (Aitkens and Barbarin 2008; Buckingham, Wheldall, and Beaman-Wheldall 2013; van Bergen *et al.* 2017). Poor literacy environment at home, number of books at home, as well as parental occupation are all factors that contribute to this. In fact, poorer households have also been found to have less access to learning materials and experiences – books, computers, lessons or even tutoring for their children (Bradley *et al.* 2001; Orr 2003). This evidence builds a direct bridge between family income limitations and socioeconomic disparities in children's learning (reading in particular). Besides literacy and

cognitive differences, there is evidence that students of lower SES may have twice the propensity towards bad learning-related behavior (Morgan *et al.* 2009).

In fact, the question of the socioeconomic status' influence on academic achievement also poses other concerns, such as the psychological factors inherent to that relationship, both from the student's and the teacher's point of view. Lloyd and Barenblatt (1984) found intrinsic intellectual motivation had a positive relation with scholastic achievement, yet it had no relation with SES when controlling for IQ. On behalf of the teacher, there are many possible biases that may influence a teacher's assessment of their student, and thus the student's academic success, depending on the educational system in question. Evidence on self-fulfilling prophecies exist since the 1960s, when Rosenthal and Jacobson (1968) conducted an experiment attempting to bias teachers' assessments in San Francisco through their expectations and succeeded. Parental expectations have a substantial impact on educational achievement as well, namely when they coincide with students' expectations (Marcenaro-Gutiérrez and Lopez 2016).

A key point of this study relies on the difference between subjective teacher assessments and more objective test assessments. Whether socioeconomic heritage has a large causal effect on a student's educational outcomes or not depends itself on how to assess one's educational outcome. Teacher and test assessments differ largely on the skills they evaluate, so the scores given are of course vulnerable to biases. Teacher expectations are one form of bias in teachers' assessments. Dee (2005) presents evidence that «racial, ethnic and gender dynamics between students and teachers have consistently large effects on teacher perceptions of student performance». Lavy (2008) found a consistent gender bias harming boys, and Auwarter and Aruguete (2008) presented evidence of gender and socioeconomic biases through teachers' expectations, which harmed low-SES boys above all. On the other hand, Baird (1998), Krkovic *et al.* (2014) and Lindahl (2016) found no evidence of such biases. Brackett *et al.* (2013), Schutz and Zembylas (2009) and Sutton and Wheatley (2003) constitute mere examples of a growing

body of research on teacher emotions, and how these may affect teacher performance and assessments.

Test assessments are not unanimous either; some literature indicates these may be subject to systematic biases as well. Sackett, Broneman, and Connelly (2008) present an overall appraisal of the evidence found throughout large-scale studies in existing literature which support the validity of tests of developed ability and their predictive power, by dissecting several assertions about testing. While recognizing there are group differences between minorities and majorities in test assessments (especially regarding race and ethnicity), Sackett, Broneman, and Connelly (2008) argue that these differences correspond to posterior differences in performance, thus not constituting proof of systematic bias in this type of assessment. Additionally, the authors considered that differences in groups are not simply caused by different degrees of motivation, and that SES is not a main source of validity in these assessments, countering literature that suggested tests measured SES above all. The relationship between test scores and academic performance is only slightly affected when controlling for the effects of SES on both test scores and grades, proving tests are not artifacts of SES even though SES is associated with the development of characteristics that affect academic performance (Sackett, Broneman, and Connelly 2008).

Given the discrepancies in both types of grading schemes, comparisons of the two are crucial for an understanding of their impact in students' assessment. Delap (1995) found that results differ substantially across subjects in British education, even though girls tend to have higher teacher-given scores than boys with the same final grades. Plewis (1997) used data on English education as well and concluded there was some mismatch by ethnic group or minority, perhaps explained by low teacher expectations for boys, ethnic minorities, and students from less advantaged backgrounds. Still using data for England, Reeves, Boyle, and Christie (2001), found teacher assessment was more likely to be lower than national exam scores for students

with special education needs than for others. Gibbons and Chevalier (2008) found smaller differences between groups of students, but clear evidence that higher-achieving pupils were being under-assessed in teacher grading relative to exam grading, and lower-achieving pupils were more likely to be under-assessed in the exams relative to the teachers' grading. An explanation for this is the smaller variance in teacher grading, which benefits low-achieving students, and for that reason can hardly be explained by statistical discrimination or teacher stereotypes (Gibbons and Chevalier 2008). Findings of this sort support the idea that none of these grading schemes should be considered alone, especially not in the context of higher education access, even though the authors found little evidence of consequences of this divergence on post-16 pupil decisions or higher education participation rates (Gibbons and Chevalier 2008). Using American data, Martínez, Stecher, and Borko (2009) found evidence that teachers evaluate students not absolutely, but on a relative scale, and thus may adjust their grading based on perceived differences among the students.

Lindahl (2007) compared teacher and test assessment scores in Swedish education to conclude girls are more generously rewarded by teacher assessments relative to test scores in the three subjects analyzed, as well as non-native students in two out of the three subjects studied. Marcenaro-Gutiérrez and Vignoles (2015) conclude in a similar analysis applied to Andalusian data that girls are favored in their internal scores relative to exam scores in mathematics, while the same happens for immigrants in reading, and the opposite for students with more absences in mathematics (better in exam scores relative to teacher assessment). More recently, Ângelo and Reis (2021) estimated a linear specification model similar to the previous two, to conclude gender disparities exist in assessment gaps in Portugal. The following sections will follow the specifications in these last three papers to examine whether there are significant socioeconomic disparities in the gap between teacher assessment and exam scores in the Portuguese education system.

3. Institutional setting

3.1. Portuguese education system

The Portuguese education system is regulated by the Ministry of Education. It includes public state-owned schools, as well as semi-private and private schools. For students in any type of school, 12 years of education are compulsory, starting in the year they turn 6 and expectedly ending around the year they turn 18. After the 9th grade of basic education students enter secondary education and may follow the academic or professional track, whether they wish to eventually pursue higher education or not, respectively. The general programmes in the academic track are Sciences and Technologies, Socioeconomic Sciences, Languages and Humanities, and Visual Arts.

At the end of every term, for each discipline, students are given scores by their teachers, with the last term's score being the final score of a given discipline in that school year. It is also at this time, after the publication of teacher scores, that students in the 9th, 11th and 12th grades in the academic track take their national exams, as well as any student in the 6th grade from 2012 to 2015.

During basic education, the only disciplines evaluated through national exams are Portuguese and Mathematics. Both the teachers' and the exams' grades are in a scale of 1 to 5, with 3 being the passing grade. In secondary education, there are overall 21 disciplines subject to national exams. Teacher scores are in a scale of 0 to 20, with 10 as the passing grade. However, exam scores are in a scale of 0 to 200. For the purpose of this analysis these results were divided by 10. At the end of secondary education, every student must take the Portuguese national exam and three others, specific to their academic path. Two exams at the end of the 11th grade on the biennial disciplines they have taken in the 10th and 11th grades, and one more at the end of the 12th grade on the discipline mandatory for their field of studies.

Students may apply for the national exams as internal or external students. Internal students must have taken the classes on the discipline of the exam and completed it with an internal (teacher-given) grade of at least 10 (from 1 to 20). Their final grade in that discipline will be weighed using the internal grade (70%) and the exam grade (30%). For external students, the final grade is the national exam grade. For the purpose of this thesis, which compares teacher-given scores with exam grades, only internal students will be accounted for.

There are two phases of exams. Students enroll in the second phase to improve their grade, or if they have been forced to skip the first phase exam. Only first phase exams will be considered in this analysis.

National exams expectedly assess the same content planned by the Ministry to be taught and evaluated by the teachers in that specific discipline. The exams take between 90 and 150 minutes depending on the discipline and include both multiple choice and open-ended questions. The criteria for the grading of the exams are defined by the Ministry, and the grading itself is done by anonymous teachers, in a different school than that of the student who took the exam. The student is also not identifiable, only by an ID given for each exam.

By contrast, student evaluation by the teacher is computed through series of tests, and other more continuous assessments. Oral presentations, group works, attendance, participation, behavior, as well as lack (or not) of material, are all typically considered in the teacher's score of the student. Thus, the two types of evaluation assess different abilities, and students may do better in one or the other depending on their specific characteristics.

3.2. SASE

The variable through which socioeconomic status is measured in this study corresponds to students' families receiving social support through SASE (Serviço de Ação Social Escolar).

This service provides economic support to families with low incomes, seeking a more leveled playing field in pre-school, basic and secondary education. Support is provided in feeding, transportation, insurance, study material, and study trips. Even though it is only representative of a financial component of the complex concept of socioeconomic status, it is a clear flag for students with economic needs in their education. For the purpose of this thesis, we will separate students in two categories, SASE and non-SASE receiving students, and estimate the differences in differences of their assessments in section 5.

4. Data and Descriptive statistics

This study is based on data mostly extracted from the MISI system of information, provided by the DGEEC (Direção Geral de Estatísticas da Educação e Ciência), supervised by the Ministry of Education and Science. The MISI system aggregates data from pre-school, basic, and secondary education in Portugal. For the purpose of this thesis only data from public schools were considered, as socioeconomic information from private schools' students is not available. The data regarding scores in the national exams are from the ENEB (Exames Nacionais de Ensino Básico) and ENES (Exames Nacionais de Ensino Secundário) datasets, which aggregate anonymized information on the students' results in the national exams of Basic and Secondary education, respectively. Both are provided by the DGEEC as well.

The data used include anonymized information on students' characteristics such as age, gender, teacher-given score, exam score, school, class, and teacher id (anonymous as well). Socioeconomic characteristics are included, of which the student benefitting of SASE is the main variable considered. The focus of the following analysis is the difference in the gap between assessments for students receiving SASE and students not receiving the social support.

Given the results obtained by Ângelo and Reis (2021), which showed the existence of an assessment gap by gender, this analysis is separated by gender. Table 1 represents descriptive statistics of the data regarding scores in Portuguese and Mathematics of the 6th, 9th, and 12th grades. Boys and girls are represented separately. The table includes number of observations, percentage of students receiving SASE, means of teacher grades and exam scores, mean differences in assessments, and a T-test of those differences in square brackets below. These last three are shown for SASE and non-SASE-receiving students. Standard errors are represented in brackets and adjusted for clustering at the school level. Due to the period of analysis being much longer, the number of observations is considerably larger for the 9th grade scores than in the 6th grade scores. As mentioned above, only national exams are considered, and in Portugal in the 6th grade there have only been national exams from 2012 until 2015. The period of analysis for the 9th and 12th grades is from 2008 until 2018.

As mentioned in section 3.1, grades in basic education in Portugal are given on a scale of 1 to 5, while in secondary the scale is between 0 and 20. The weight of SASE recipients in the subsamples varies as well: it decreases from being in the range of 43.5-44.2% in the 6th grade, to 36-39.3% in 9th grade, and to 17.6-27.3% in the 12th grade.

In columns (4) and (5) the mean teacher-given scores and standard errors are presented for students not receiving SASE and students with the subsidy, respectively. Students receiving SASE tend to perform worse on average than students not receiving the support, which goes along with the existing literature on different academic performances by socioeconomic status.

Table 1: Means and standard errors of teachers' and exams' scores for grades 6 (2012-15), 9 (2008-18) and 12 (2008-18) – Mean comparison tests

Subject	Gender	Number of observations	% SASE	Mean Teacher Score		Mean Exam Score		Mean Gap (Teacher-Exam)		Diff. in mean gaps (9) – (8) T-stat
				Non SASE	SASE	Non SASE	SASE	Non SASE	SASE	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
6th grade										
Mathematics	Boys	166905	43.5	3.421 (0.01)	2.942 (0.02)	3 (0.02)	2.43 (0.02)	0.418*** (0.01)	0.512*** (0.01)	0.094*** [5.165]
	Girls	160975	44.2	3.49 (0.02)	3 (0.02)	2.987 (0.02)	2.433 (0.02)	0.504*** (0.01)	0.566*** (0.01)	0.061*** [3.396]
Portuguese	Boys	163747	43.5	3.4 (0.01)	3.06 (0.01)	3.07 (0.01)	2.73 (0.01)	0.335*** (0.01)	0.324*** (0.01)	-0.012 [-0.9302]
	Girls	157921	44.2	3.663 (0.01)	3.291 (0.01)	3.28 (0.01)	2.924 (0.01)	0.383*** (0.01)	0.367*** (0.01)	-0.016* [-1.286]
9th grade										
Mathematics	Boys	336566	36	3.149 (0.01)	2.813 (0.01)	2.847 (0.02)	2.374 (0.02)	0.302*** (0.01)	0.439*** (0.01)	0.137*** [8.235]
	Girls	361060	39.3	3.253 (0.01)	2.875 (0.01)	2.909 (0.02)	2.404 (0.02)	0.344*** (0.01)	0.471*** (0.01)	0.1272 [7.832]
Portuguese	Boys	332761	36	3.207 (0.01)	3.002 (0.01)	2.93 (0.01)	2.67 (0.01)	0.275*** (0.01)	0.329*** (0.01)	0.054*** [4.747]
	Girls	357151	39.2	3.49 (0.01)	3.235 (0.01)	3.184 (0.01)	2.888 (0.01)	0.305*** (0.01)	0.347*** (0.01)	0.041*** [3.718]
12th grade										
Mathematics	Boys	89706	17.6	13.71 (0.04)	13.26 (0.08)	11.58 (0.08)	10.67 (0.17)	2.137*** (0.07)	2.591*** (0.15)	0.455*** [2.703]
	Girls	106980	22.7	14.11 (0.05)	13.66 (0.08)	11.57 (0.08)	10.71 (0.15)	2.54*** (0.07)	2.957*** (0.13)	0.418*** [2.714]
Portuguese	Boys	150195	21.1	13.31 (0.05)	12.87 (0.08)	10.63 (0.04)	10.06 (0.07)	2.687*** (0.05)	2.806*** (0.09)	0.119 [1.19]
	Girls	206596	27.3	14.23 (0.05)	13.74 (0.08)	11.62 (0.05)	11 (0.07)	2.616*** (0.05)	2.741*** (0.08)	0.126* [1.382]

Note: Grading scale in 6th and 9th grades is 1 to 5. In the 12th grade it is 0 to 20. Standard errors in parenthesis below are corrected for clustering at the school level. The mean difference in columns (8) and (9) corresponds to mean teacher score minus mean exam score. In column (10) the difference tested is column (9) minus column (8). The T statistic in square brackets presented below corresponds to the T test for the difference in mean differences. In columns (8)-(10), *, **, *** denote statistical significance at 10%, 5% and 1%, respectively.

There is a gap in internal grade by gender as well – girls tend to perform better on average than boys in this aspect, whether we consider SASE recipients or non-recipients. Columns (6) and (7) show means and standard errors of exam scores. In exams, as in teacher grading, girls mostly outperform boys and non-SASE students seem to outperform students who benefit from SASE,

on average. Only in 12th grade Mathematics for non-SASE-receiving girls is this not true. This gender difference, both in internal grades and in the exam scores, is visibly larger in Portuguese than in Mathematics. This is in line with existing literature that refers to girls outperforming boys in reading (Ângelo and Reis, 2021), more so than in mathematics.

Columns (8) and (9) of Table 1 show the gap between assessments by SASE and gender. This gap, which is the difference between mean teacher scores and mean exam scores, is always positive in basic education. Students of both genders, either receiving SASE or not, perform worse in the exam on average. Column (10) includes the difference in mean assessment gaps between SASE and non-SASE recipients, which is positive and statistically significant in most subjects, with the exceptions being 9th grade Mathematics for girls, and Portuguese in the 6th and 12th grades for both genders. These statistics would suggest SASE students tend to be better assessed internally than externally in comparison with non-SASE recipients. This would go against some of the existing literature that found students of lower SES to be more likely to be under-assessed by their teachers, whatever the cause (Auwarter and Aruguete 2008; Dee 1995; Morgan *et al.* 2009; Plewis 1997).

A look at how estimated means and differences change along the distribution of students will prove useful to understand the real patterns behind these numbers. In Table 2 below, a similar description of the data is presented for each level of exam scores in Portuguese and Mathematics of the 12th grade.

Secondary education exams' scores are originally given in a scale of 0 to 200. Here, for comparison and simplification purposes, these are merged in a scale of 1 to 5 following the Ministry of Education's criteria (Ângelo and Reis 2021). Level 1 includes scores in the [0,40) interval; scores inside [40,100) fit in level 2; level 3 corresponds to the [100,140) interval; 4 to the [140,180) interval; and 5 to the [180,200] interval.

Table 2: Means and standard errors of teachers' and exams' scores in Portuguese and Mathematics of the 12th grade (2008-18) - Mean comparison tests by exam score level

Subject/Score	Gender	Number of observations	% SASE	Mean Teacher Score		Mean Exam Score		Mean Gap (Teacher-Exam)		Diff. in mean gaps (9) – (8) T-stat
				Non SASE	SASE	Non SASE	SASE	Non SASE	SASE	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Mathematics										
1	Boys	2433	24.5	10.13	10.16	2.35	2.32	7.782***	7.835***	0.054
				(0.06)	(0.11)	(0.02)	(0.03)	(0.07)	(0.12)	[0.399]
	Girls	2568	29.3	10.37	10.35	2.41	2.38	7.96***	7.97***	0.01
				(0.06)	(0.1)	(0.02)	(0.03)	(0.07)	(0.11)	[0.078]
2	Boys	26816	20.5	11.23	11.24	6.84	6.71	4.389***	4.527***	0.138
				(0.04)	(0.07)	(0.02)	(0.04)	(0.05)	(0.1)	[1.249]
	Girls	33446	26	11.6	11.61	6.85	6.75	4.743***	4.862***	0.119
				(0.04)	(0.07)	(0.02)	(0.03)	(0.06)	(0.09)	[1.103]
3	Boys	28823	17.6	13.06	13.03	11.19	11.09	1.868***	1.947***	0.079
				(0.05)	(0.1)	(0.01)	(0.02)	(0.05)	(0.11)	[0.652]
	Girls	33787	23.1	13.53	13.56	11.18	11.16	2.35***	2.4***	0.051
				(0.05)	(0.1)	(0.01)	(0.01)	(0.06)	(0.1)	[0.439]
4	Boys	21751	15.6	15.67	15.62	15.21	15.12	0.459***	0.505***	0.046
				(0.05)	(0.12)	(0.01)	(0.02)	(0.05)	(0.12)	[0.352]
	Girls	25418	20.6	16.12	16.1	15.21	15.17	0.901***	0.926***	0.025
				(0.06)	(0.1)	(0.01)	(0.02)	(0.05)	(0.1)	[0.216]
5	Boys	9883	12.3	18.23	18.24	18.54	18.45	-0.312***	-0.215***	0.097
				(0.04)	(0.1)	(0.01)	(0.02)	(0.04)	(0.1)	[0.881]
	Girls	11761	15.1	18.46	18.41	18.56	18.48	-0.099**	-0.072	0.028
				(0.04)	(0.1)	(0.01)	(0.02)	(0.04)	(0.09)	[0.271]
Portuguese										
1	Boys	774	27.3	10.87	10.77	2.69	2.64	8.18***	8.13***	-0.048
				(0.07)	(0.12)	(0.03)	(0.04)	(0.08)	(0.13)	[-0.315]
	Girls	595	35.8	11.2	11.1	2.67	2.74	8.52***	8.36***	-0.158
				(0.11)	(0.14)	(0.03)	(0.04)	(0.11)	(0.15)	[-0.846]
2	Boys	50229	24.3	11.84	11.72	7.27	7.19	4.578***	4.533***	-0.045
				(0.03)	(0.06)	(0.01)	(0.02)	(0.04)	(0.07)	[-0.568]
	Girls	48121	32	12.26	12.16	7.38	7.3	4.877***	4.863***	-0.014
				(0.04)	(0.06)	(0.01)	(0.02)	(0.05)	(0.06)	[-0.183]
3	Boys	70405	20.7	13.29	13.05	11.08	10.99	2.213***	2.064***	-0.149*
				(0.04)	(0.08)	(0.01)	(0.02)	(0.04)	(0.08)	[-1.62]
	Girls	97164	28	13.88	13.64	11.23	11.16	2.647***	2.479***	-0.168**
				(0.05)	(0.07)	(0.01)	(0.01)	(0.05)	(0.07)	[-2]
4	Boys	26885	16.7	15.54	15.18	14.79	14.69	0.764***	0.486***	-0.278**
				(0.05)	(0.11)	(0.01)	(0.02)	(0.05)	(0.11)	[-2.342]
	Girls	55188	22.9	15.97	15.63	14.9	14.78	1.068***	0.847***	-0.221**
				(0.05)	(0.09)	(0.01)	(0.01)	(0.05)	(0.08)	[-2.28]
5	Boys	2262	14.3	17.85	17.52	18	17.95	-0.169***	-0.431***	-0.2624**
				(0.06)	(0.14)	(0.01)	(0.03)	(0.06)	(0.14)	[-1.699]
	Girls	5528	16.5	17.9	17.63	18.01	17.95	-0.116**	-0.323***	-0.207**
				(0.05)	(0.11)	(0.01)	(0.02)	(0.05)	(0.11)	[-1.698]

Note: Grading scale is 0 to 20. Standard errors in parenthesis below are corrected for clustering at the school level. The mean difference in columns (8) and (9) corresponds to mean teacher score minus mean exam score. In column (10) the difference tested is column (9) minus column (8). The T statistic in square brackets presented below corresponds to the T test for the difference in mean differences. In columns (8)-(10), *, **, *** denote statistical significance at 10%, 5% and 1%, respectively.

The choice of the national exam scores as a reference for the segregation of the data is based on the higher objectivity of this type of assessment, which as mentioned in section 3.1 is

anonymous and not affected by most of the non-cognitive factors that influence teacher's scores. Therefore, it is the closest to a standardized type of assessment across the dataset, in each given year and subject, allowing us to better compare SASE and non-SASE beneficiaries.

As we can see in Table 2, results by level of exam score provide a different picture of the real differences. Generally, girls still outperform boys, and students receiving SASE are outperformed by students without the subsidy. However, the gaps in assessments change considerably along the distribution of scores in exams, as does the percentage of students with SASE. These changes are not independent; in fact, the concentration of SASE beneficiaries in the lower part of the distribution of scores, and the fact that students with the worst scores in the exams have much larger and positive assessment gaps (explainable by the smaller range of teacher assessment which by construction does not reach below 10 since we are only including internal students) are what caused differences in means in Table 1 to be misleading. In higher exam score levels, the assessment gap becomes increasingly small, and even negative (Table 2). This is expectable since the gap is the teacher's score minus the score in the exam. Students with an already high internal score are less likely to improve their grade in the exam, which explains a negative mean gap.

Differences in assessment gaps by SASE presented in column (10) of Table 2 are substantially different than those presented in Table 1. The strong and significant difference observed in 12th grade Mathematics in Table 1 is no longer seen in Table 2 – differences are residual and no longer statistically significant between SASE and non-SASE beneficiaries. Regarding 12th grade Portuguese, the difference in assessment gaps is much more negative and significant when we examine it by exam score level in column (10) of Table 2, especially for the subsamples with higher exam scores. Given that in this table we are bounding variation in exam scores up to a certain level, smaller internal grades for SASE beneficiaries may be the major reason for this significant difference.

Table A1 of the appendix provides the same descriptive statistics by level of exam score for Portuguese and Mathematics of the 6th and 9th grades. A negative gap in differences in assessments by SASE is clear in all levels of exam scores, and statistically significant for all but the students with 1 in the exam. As basic education scores are already in a scale of 1 to 5, segregation of exam scores in 5 levels restricts variation in this component of the gap completely. We may be certain that the gap in assessments here is thus caused by lower internal grades on behalf of SASE beneficiaries, which is in line with the literature. Basic education students of lower economic status are indeed given worse grades by their teacher relatively to those without SASE with a same score in the exam, whether because of worse behavior, higher absenteeism, teacher bias or simply lower ability in qualities evaluated by the teacher and not assessed in the exams. Table A2 in the appendix includes the same statistics as Tables 1 and 2 for the other 8 subjects in secondary education analyzed in this study, which show similar patterns to the ones of Portuguese and Mathematics of the 12th grade, depending on the subject.

5. Regression analysis

5.1. Specification

In this previous section, an analysis of descriptive statistics of the data available suggested that the SASE-receiving students benefit more from teacher grading than the non-SASE-receiving students when looking at the whole distribution of grades. However, a closer look at the assessment gaps by level of exam score proved this to be deceitful due to the unbalanced distribution of SASE students along the exam score levels. In this section 5, a regression analysis will be carried out to better understand the SASE gap in assessments, when controlling for certain characteristics of the students and the assessment context. This specification follows

the linear models used by Lindahl (2007), Marcenaro-Gutiérrez and Vignoles (2015), and Ângelo and Reis (2021) in studying the gender gap in assessments. The model goes as follows:

$$G_{it} = \beta_0 + \beta_1 S_{it} + \beta_2 A_{it} + \mu_t + \epsilon_{it}$$

The dependent variable G_{it} is the individual assessment gap analyzed before – the difference between teacher grading and exam score for a given student i in a certain year t . The goal of this study is to conclude whether G_{it} may be a function of the socioeconomic status of a given student, S_{it} , represented by a dummy equal to one if the student receives SASE, while controlling for other variables. The variable A_{it} controls for student age at the time t of the assessments. Both exams and teacher grading are subject to time variation, so year fixed effects are always included in all models (μ_t).

Given the discrete scale of the dependent variable in the case of basic education students, an ordered probit or logit model could be used. Nevertheless, the size of the sample and the focus on the marginal effect of SASE justify the linear approach (Lindahl 2007; Ângelo and Reis 2021). Four specifications of this model will be estimated: the regression model as specified above, and three additional ones including fixed effects at the school, class, and teacher level, respectively.

Additionally, since the assessment gap changes drastically along the distribution of scores, the four specifications of the equation above will also be estimated for each of the five levels of the national exam scores mentioned in the previous section.

5.2. Results

5.2.1. Mean estimations

Table 3 below presents the estimations specified in the previous subsection for the whole distribution of scores in Portuguese and Mathematics of the 6th, 9th and 12th grades.

Table 3: Socioeconomic gap in assessments: Portuguese and Mathematics – 6th, 9th and 12th grades

Subject	Gender	Number of observations	(1) OLS	(2) School FE	(3) Teacher FE	(4) Class FE
Mathematics						
6 th grade	Boys	166,905	0.0760*** (0.00624)	0.0452*** (0.00385)	0.0340*** (0.00354)	0.0696*** (0.00580)
	Girls	160,975	0.0529*** (0.00634)	0.0155*** (0.00388)	0.00364 (0.00372)	0.0473*** (0.00598)
9 th grade	Boys	336,566	0.119*** (0.00616)	0.0669*** (0.00275)	0.0665*** (0.00269)	0.108*** (0.00557)
	Girls	361,060	0.108*** (0.00592)	0.0560*** (0.00270)	0.0560*** (0.00258)	0.0983*** (0.00536)
12 th grade	Boys	89,706	0.485*** (0.0517)	0.207*** (0.0248)	0.184*** (0.0232)	0.403*** (0.0461)
	Girls	106,980	0.446*** (0.0466)	0.213*** (0.0212)	0.181*** (0.0200)	0.380*** (0.0406)
Portuguese						
6 th grade	Boys	163,747	-0.0130*** (0.00495)	-0.0359*** (0.00375)	-0.0401*** (0.00381)	-0.0164*** (0.00477)
	Girls	157,921	-0.0176*** (0.00478)	-0.0428*** (0.00373)	-0.0455*** (0.00378)	-0.0207*** (0.00467)
9 th grade	Boys	332,761	0.0412*** (0.00422)	0.00964*** (0.00273)	0.00735*** (0.00264)	0.0350*** (0.00388)
	Girls	357,151	0.0275*** (0.00425)	-0.00580** (0.00246)	-0.00683*** (0.00243)	0.0221*** (0.00386)
12 th grade	Boys	150,195	0.120*** (0.0302)	0.0246 (0.0183)	0.0153 (0.0165)	0.101*** (0.0273)
	Girls	206,596	0.114*** (0.0281)	0.0312** (0.0141)	0.0204 (0.0126)	0.101*** (0.0234)
Age			Yes	Yes	Yes	Yes
Year FE			Yes	Yes	Yes	Yes
School FE			No	Yes	No	No
Teacher FE			No	No	Yes	No
Class FE			No	No	No	Yes

Note: Dependent variable is the assessment gap, *TeacherScore* - *ExamScore*. Only the coefficient of interest β_1 is presented. Standard errors in parenthesis are robust and corrected for clustering at the school level in all models. *, **, *** denote statistical significance at 10%, 5% and 1%, respectively.

The coefficient presented in the table is our coefficient of interest β_1 , associated with the dummy variable equal to one if the student receives SASE support. As we can see, the four models present estimations which are consistent with the descriptive statistics in Table 1 of section 4. Receiving SASE is associated with an increase in the gap between the teacher's score and the exam's score, on average, all else constant, and for both genders. This is observable even when we include school, class, or teacher fixed effects. The exceptions are girls'

assessments in 9th grade Portuguese when adding school or teacher fixed effects, and all the scores in 6th grade Portuguese.

Regarding the other analyzed subjects of secondary education, estimates are mostly similar. Table A3 in the appendix presents these estimations. Positive coefficients β_1 suggest SASE-receiving students have on average a larger gap between teacher-given and exam scores for every subject analyzed. This is again consistent with the analysis of descriptive statistics done in section 4. However, we have already seen that by aggregating observations from sections of the data with clearly different percentages of SASE students and gaps in assessments, we are probably inducing a large bias in our coefficient of interest. Estimations by exam score level are necessary for a credible analysis of the socioeconomic gap in grading schemes.

5.2.2. By exam score level

Table 4 below presents the estimated coefficients of interest for the respective exam score level in Portuguese and Mathematics of the 6th, 9th, and 12th grades. The coefficients reported were estimated through specification (2) of Table 3, that is, including school fixed effects. Below, standard errors adjusted to clustering at the school level are presented in parenthesis, with the percentage of SASE students in each subsample below that. As we would expect from the contrast in the descriptive statistics when analyzed by exam level, the estimations differ significantly from the ones in Table 3. Our coefficient of interest is no longer consistently positive; on the contrary, estimates are mostly negative.

Regarding basic education, estimated coefficients are negative and significant for all but level 1 of exam scores. This is true for both subjects, Portuguese and Mathematics. However, the amplitude of the coefficients in Table 4 is consistently larger in Portuguese assessments than in Mathematics assessments, especially in the regressions for higher-achieving students.

Table 4: Socioeconomic gap in assessments by exam score level: Portuguese and Mathematics – 6th, 9th and 12th grades.

Subject	Gender	Exam score				
		(1)	(2)	(3)	(4)	(5)
Mathematics 6 th grade	Boys	-0.0125 (0.00874) 66%	-0.0829*** (0.00479) 54%	-0.124*** (0.00605) 39%	-0.154*** (0.00902) 26%	-0.100*** (0.0172) 17%
		-0.0112 (0.00999) 67%	-0.122*** (0.00493) 56%	-0.138*** (0.00602) 39%	-0.160*** (0.00919) 27%	-0.0761*** (0.0175) 18%
	Girls	-0.00546 (0.00414) 50%	-0.0455*** (0.00339) 43%	-0.0464*** (0.00444) 33%	-0.102*** (0.00673) 23%	-0.0818*** (0.0124) 15%
		-0.00191 (0.00391) 54%	-0.0531*** (0.00332) 47%	-0.0812*** (0.00433) 36%	-0.107*** (0.00591) 26%	-0.0747*** (0.0102) 18%
9 th grade	Boys	0.000176 (0.0897) 25%	0.0312 (0.0308) 21%	-0.0264 (0.0340) 18%	-0.0531 (0.0409) 16%	-0.00439 (0.0447) 12%
		0.0470 (0.107) 29%	0.0188 (0.0261) 26%	-0.0641** (0.0268) 23%	-0.0712** (0.0295) 21%	-0.0589 (0.0369) 15%
	Girls	0.000176 (0.0897) 25%	0.0312 (0.0308) 21%	-0.0264 (0.0340) 18%	-0.0531 (0.0409) 16%	-0.00439 (0.0447) 12%
		0.0470 (0.107) 29%	0.0188 (0.0261) 26%	-0.0641** (0.0268) 23%	-0.0712** (0.0295) 21%	-0.0589 (0.0369) 15%
Portuguese 6 th grade	Boys	0.00325 (0.0488) 70%	-0.0664*** (0.00483) 56%	-0.162*** (0.00445) 43%	-0.266*** (0.00919) 28%	-0.218*** (0.0372) 16%
		-0.0945 (0.121) 73%	-0.0725*** (0.00531) 60%	-0.177*** (0.00488) 46%	-0.242*** (0.00768) 31%	-0.164*** (0.0236) 20%
	Girls	-0.00463 (0.0324) 52%	-0.0201*** (0.00290) 44%	-0.0829*** (0.00304) 35%	-0.191*** (0.00721) 23%	-0.160*** (0.0293) 14%
		0.00226 (0.0913) 60%	-0.0285*** (0.00301) 51%	-0.0940*** (0.00301) 40%	-0.197*** (0.00543) 28%	-0.148*** (0.0153) 18%
9 th grade	Boys	-0.155 (0.196) 27%	-0.0909*** (0.0227) 24%	-0.173*** (0.0219) 21%	-0.303*** (0.0370) 17%	-0.215* (0.129) 14%
		-0.0961 (0.321) 36%	-0.0571*** (0.0219) 32%	-0.207*** (0.0174) 28%	-0.279*** (0.0208) 23%	-0.300*** (0.0590) 17%
	Girls	-0.155 (0.196) 27%	-0.0909*** (0.0227) 24%	-0.173*** (0.0219) 21%	-0.303*** (0.0370) 17%	-0.215* (0.129) 14%
		-0.0961 (0.321) 36%	-0.0571*** (0.0219) 32%	-0.207*** (0.0174) 28%	-0.279*** (0.0208) 23%	-0.300*** (0.0590) 17%
12 th grade	Boys	-0.155 (0.196) 27%	-0.0909*** (0.0227) 24%	-0.173*** (0.0219) 21%	-0.303*** (0.0370) 17%	-0.215* (0.129) 14%
		-0.0961 (0.321) 36%	-0.0571*** (0.0219) 32%	-0.207*** (0.0174) 28%	-0.279*** (0.0208) 23%	-0.300*** (0.0590) 17%
	Girls	-0.155 (0.196) 27%	-0.0909*** (0.0227) 24%	-0.173*** (0.0219) 21%	-0.303*** (0.0370) 17%	-0.215* (0.129) 14%
		-0.0961 (0.321) 36%	-0.0571*** (0.0219) 32%	-0.207*** (0.0174) 28%	-0.279*** (0.0208) 23%	-0.300*** (0.0590) 17%

Note: Dependent variable is the assessment gap, *TeacherScore* - *ExamScore*. The coefficient of interest β_1 is presented for estimations at each exam score level, following the specification with school fixed effects. Standard errors presented in parenthesis are robust and corrected for clustering at the school level. Percentage below corresponds to the share of students receiving SASE in each subsample. *, **, *** denote statistical significance at 10%, 5% and 1%, respectively.

Given that we are fixing the exam score component of the assessment gap in each of the estimations, the reported coefficients in Table 4 for basic education reflect differences in teacher-given grades for SASE and non-SASE beneficiaries. Therefore, these estimates suggest that, on average and all else constant, students receiving SASE tend to have lower internal

grades, even when we control for students' age and for time and school-specific factors that could explain some of the variation in the assessment gaps. Furthermore, that this effect is larger in Portuguese than in Mathematics could be additional evidence that socioeconomic status translates into reading and writing skills at an early age (Aitkens and Barbarin 2008; Buckingham, Wheldall, and Beaman-Wheldall 2013; van Bergen *et al.* 2017).

Looking at the estimates for the 12th grade subjects in Table 4, these are consistent with the descriptive statistics as well. Most estimated coefficients for Mathematics are not statistically different from zero, except for girls with scores of 3 and 4, and even these are relatively smaller in their amplitude than estimates for Portuguese, which are all negative and significant from the score of 2 upwards (Table 4). The amplitude and significance of the gap in assessment differences in 12th grade Portuguese is particularly concerning given that it is the only secondary school subject that is mandatory for all who follow general programmes in Portuguese secondary education, as is its exam. Even though it does include a larger number of observations which might influence its significance, so does 12th grade Mathematics, and gaps or biases in assessments in these subjects may have larger consequences on educational attainment and even access to higher education.

Table 5 below reports coefficients of interest for each of the remaining 8 subjects in high school education approached in this study, by exam score level and following the specification with school fixed effects.

Once again, these estimates by exam level and their contrast with the estimates for the whole distribution of grades in each subject coincide with the pattern found in the descriptive statistics¹. Estimates presented in Table 5 are mostly negative once again, especially for students with an exam score of level 3 upwards.

¹ Table A2 of the appendix

Table 5: Socioeconomic gap in assessments by exam score level: remaining secondary education subjects.

Subject	Gender	Exam score				
		(1)	(2)	(3)	(4)	(5)
Biology & Geology	Boys	-0.312 (0.259) 26%	-0.0557** (0.0254) 26%	-0.102*** (0.0273) 22%	-0.148*** (0.0393) 18%	-0.327** (0.137) 13%
	Girls	-0.226 (0.279) 34%	-0.0345 (0.0279) 32%	-0.144*** (0.0291) 26%	-0.0812** (0.0347) 21%	-0.220** (0.101) 15%
Descriptive Geometry	Boys	0.0925 (0.202) 31%	-0.00559 (0.0886) 25%	0.0285 (0.107) 21%	-0.0746 (0.112) 17%	-0.125 (0.0891) 14%
	Girls	-0.327** (0.165) 33%	-0.181** (0.0839) 29%	-0.149 (0.132) 23%	-0.207 (0.132) 19%	-0.271** (0.127) 14%
Physics & Chemistry	Boys	-0.0992 (0.149) 28%	-0.0209 (0.0278) 25%	-0.0853** (0.0396) 20%	-0.0676 (0.0476) 16%	-0.135 (0.0839) 12%
	Girls	-0.200* (0.118) 34%	-0.0277 (0.0232) 29%	-0.139*** (0.0322) 25%	-0.0678 (0.0468) 19%	-0.235*** (0.0716) 14%
Economics A	Boys	0.301 (1.184) 25%	-0.0840 (0.0899) 21%	-0.0430 (0.0711) 19%	-0.112 (0.0893) 16%	-0.166 (0.149) 14%
	Girls	-1.056 (1.488) 27%	-0.0900 (0.0861) 28%	-0.162** (0.0712) 25%	-0.0313 (0.0838) 21%	-0.164 (0.124) 18%
History A	Boys	0.111 (0.589) 34%	0.0171 (0.0482) 31%	-0.00414 (0.0477) 29%	-0.225*** (0.0800) 26%	-0.456* (0.264) 21%
	Girls	-0.0660 (0.191) 39%	0.00638 (0.0355) 37%	-0.0667** (0.0305) 33%	-0.120*** (0.0416) 29%	-0.145 (0.106) 25%
Philosophy	Boys	-0.244 (0.475) 33%	0.00404 (0.0731) 31%	-0.0243 (0.0712) 28%	-0.0551 (0.113) 22%	-0.181 (0.385) 18%
	Girls	-0.272 (0.310) 44%	-0.102** (0.0502) 38%	-0.0476 (0.0409) 34%	-0.0847 (0.0608) 27%	-0.435** (0.171) 21%
Geography	Boys		-0.0295 (0.0438) 35%	-0.0246 (0.0308) 28%	-0.126** (0.0491) 23%	-0.190 (0.229) 18%
	Girls	-0.570 (1.853) 46%	0.0251 (0.0286) 39%	-0.0698*** (0.0243) 35%	-0.190*** (0.0379) 28%	0.0136 (0.218) 22%
Mathematics Applied To Social Sciences	Boys	-0.664 (0.532) 37%	-0.0545 (0.0708) 33%	-0.0385 (0.0644) 31%	-0.0496 (0.114) 28%	0.0293 (0.280) 24%
	Girls	-0.504 (0.311) 38%	-0.0305 (0.0481) 38%	-0.0983** (0.0480) 34%	-0.0558 (0.0529) 31%	-0.249** (0.109) 27%

Note: Dependent variable is the assessment gap, $TeacherScore - ExamScore$. The coefficient of interest β_1 is presented for estimations at each exam score level, following the specification with school fixed effects. Standard errors presented in parenthesis are robust and corrected for clustering at the school level. Percentage below corresponds to the share of students receiving SASE in each subsample. The estimated coefficient for boys with exam score of 1 in Geography is not reported due to lack of observations. *, **, *** denote statistical significance at 10%, 5% and 1%, respectively.

There is no clear difference between science and humanities related subjects in Table 5. Nor between genders. However, Biology and Geology is the subject with the most significant coefficients. All coefficients statistically different from zero are negative, suggesting that students receiving SASE tend to have smaller gaps in assessments, all else constant.

5.2.3. By quintiles of exam score distribution

A problem with estimation by exam scores is the segregation of the observations in fixed categories which might translate in unbalanced subsamples that generate inefficient estimates while not reflecting the actual ranking of the students in full. Due to the more continuous scale of exam scores in secondary education, and the competitive nature of attainment in high school (namely through its consequences in access to higher education), estimation by quintiles of the exam score distribution might provide us with more useful information.

Table 6 reports estimates on the same subjects as Table 5 plus Portuguese and Mathematics of the 12th grade. Specification (2) of Table 3, which includes school fixed effects, is the one presented in the table.

Most of the estimated coefficients shown in Table 6 still present a negative signal, and those who do not are once again small and not statistically different from zero. This is consistent with the findings so far that SASE recipients tend to have a smaller gap between teacher-given and exam scores.

Relatively to estimation by exam score level, estimation by quintiles of the exam score distribution presents statistically stronger estimates. We may now see with more certainty that Biology and Geology assessments do indeed transmit a statistically significant gap by SASE, for all quantiles except the first for girls.

Table 6: Socioeconomic gap in assessments by quintile of exam score distribution: all secondary education subjects.

Subject	Gender	Exam score quintile				
		(1)	(2)	(3)	(4)	(5)
Biology & Geology	Boys	-0.0743** (0.0293)	-0.110*** (0.0336)	-0.136*** (0.0379)	-0.0964** (0.0438)	-0.180*** (0.0486)
	Girls	-0.0248 (0.0321)	-0.117*** (0.0392)	-0.147*** (0.0379)	-0.202*** (0.0388)	-0.0950** (0.0372)
Descriptive Geometry	Boys	0.00853 (0.0929)	0.0493 (0.160)	-0.224 (0.176)	0.225 (0.148)	-0.0451 (0.0822)
	Girls	-0.176** (0.0813)	-0.264 (0.173)	-0.0403 (0.190)	-0.325* (0.184)	-0.158 (0.1000)
Physics & Chemistry	Boys	-0.0637* (0.0332)	-0.0749* (0.0414)	-0.0559 (0.0575)	-0.129** (0.0525)	-0.0494 (0.0428)
	Girls	-0.0523** (0.0249)	-0.104*** (0.0384)	-0.163*** (0.0463)	-0.152*** (0.0502)	-0.0449 (0.0443)
Mathematics	Boys	0.0606* (0.0355)	0.0501 (0.0401)	-0.0560 (0.0436)	-0.0864* (0.0463)	-0.0268 (0.0348)
	Girls	0.00996 (0.0318)	0.00108 (0.0390)	-0.0553 (0.0352)	-0.0964*** (0.0371)	-0.00672 (0.0282)
Portuguese	Boys	-0.107*** (0.0253)	-0.129*** (0.0249)	-0.184*** (0.0255)	-0.285*** (0.0360)	-0.311*** (0.0422)
	Girls	-0.0332 (0.0282)	-0.173*** (0.0234)	-0.213*** (0.0226)	-0.256*** (0.0228)	-0.283*** (0.0233)
Economics A	Boys	0.00756 (0.0990)	-0.282*** (0.0962)	-0.0514 (0.104)	-0.0865 (0.108)	-0.0649 (0.0846)
	Girls	-0.115 (0.116)	0.0411 (0.105)	-0.300*** (0.105)	-0.131 (0.104)	-0.0652 (0.0657)
History A	Boys	-0.0419 (0.0649)	0.0550 (0.0658)	0.0410 (0.0674)	-0.00997 (0.0728)	-0.262*** (0.0857)
	Girls	-0.0490 (0.0419)	-0.0493 (0.0391)	-0.0653 (0.0448)	-0.0865** (0.0404)	-0.114** (0.0455)
Philosophy	Boys	-0.0882 (0.0919)	-0.0644 (0.0849)	-0.0347 (0.102)	-0.0696 (0.115)	-0.0137 (0.132)
	Girls	-0.0569 (0.0594)	-0.165*** (0.0620)	-0.0294 (0.0590)	-0.0664 (0.0653)	-0.0925 (0.0687)
Geography	Boys	-0.0254 (0.0639)	-0.0150 (0.0413)	-0.0763* (0.0402)	-0.0406 (0.0454)	-0.187*** (0.0614)
	Girls	-0.00943 (0.0422)	-0.0640** (0.0308)	-0.0464 (0.0314)	-0.140*** (0.0341)	-0.167*** (0.0465)
Mathematics Applied To Social Sciences	Boys	-0.0180 (0.0940)	-0.0207 (0.0791)	0.0561 (0.0996)	-0.0454 (0.106)	-0.0953 (0.111)
	Girls	-0.139** (0.0579)	-0.0129 (0.0591)	-0.226*** (0.0652)	0.00581 (0.0636)	-0.0596 (0.0524)

Note: Dependent variable is the assessment gap, *TeacherScore* - *ExamScore*. The coefficient of interest β_1 is presented for estimations at each quintile of the exam score distribution, following the specification with school fixed effects. Standard errors presented in parenthesis are robust and corrected for clustering at the school level. *, **, *** denote statistical significance at 10%, 5% and 1%, respectively.

Other subjects where coefficients report significant gaps are Physics and Chemistry, especially for girls in all but the 5th quintile; Portuguese in all quintiles except for girls in the 1st; Economics in the 2nd and 3rd quintile for boys and girls, respectively; History A in the 4th and

5th quintiles for girls, and only the 5th for boys; Philosophy in the 2nd quintile for girls; Geography in the 2nd, 4th and 5th quintiles for girls, and in the 3rd and 5th for boys; and in Mathematics Applied to Social Sciences in the 1st and 3rd quintiles for girls.

Besides these peaks in significance, there are no clear patterns in the significance or amplitude of the coefficients along the quintiles. However, the subjects with the most consistent estimates, besides Portuguese, are Biology and Geology, and Physics and Chemistry, two biennial subjects assessed in the 11th grade with a heavy scientific program. That these subjects presented significantly negative estimates while Mathematics did not, may be a consequence of the nature of the criteria in teacher assessment in these courses, which does not benefit students of lower SES.

6. Concluding remarks

The goal of this thesis was to analyze whether differences in teacher assessment and test assessment in Portugal may be associated with students' socioeconomic characteristics or background, while controlling for other factors such as school effects, the level at which students of similar socioeconomic status tend to cluster.

Due to the unbalance of the weight of low-SES students along the distribution of scores, and the large degree of variability of the assessment gap along that score distribution, estimations of this effect must be carried out for different parts of that distribution, using exam scores as reference. Estimation of the linear model in all its specifications, specifically the one capturing school fixed effects, has provided consistent evidence of the existence of disparities between SASE beneficiaries and other students in their assessment gaps. Even though both groups of students tend to worsen their grades in exams, there is strong statistical evidence that, in several subjects, students receiving SASE tend to have a smaller gap between teacher-given scores and

national exam scores, or a more negative gap for the high-achievers which are more likely to worsen their grade in the exam. This gap is clear in Portuguese and Mathematics of 6th and 9th grades of basic education, as well as in Portuguese of the 12th grade, Biology and Geology (11th grade), and Physics and Chemistry (11th grade). To extrapolate on the mechanism behind this effect goes beyond the scope of this thesis – whether internal grades are pushed down due to bad behavior, absenteeism, teachers’ expectations, or other reasons is unknown for now. Even that this gap is caused by internal grades being pushed down is not clear and would be very much based on the more objective nature of test-assessments and their prediction of capabilities. All we may say is that these results confirm findings in existing literature that teacher and test-based assessments may differ systematically across students’ socioeconomic characteristics. Additionally, we may draw policy implications in the debate for fair access to higher education in the sense that both teacher and test assessments are vulnerable to biases and favor different characteristics of students. As such, the two grading schemes should be weighed cautiously and none of them should be discarded, as it would have consequences on the distribution of students accessing higher education and eventually the labor force.

7. References

- Aikens, Nikki L., and Oscar Barbarin. 2008. "Socioeconomic differences in reading trajectories: The contribution of family, neighborhood, and school contexts." *Journal of Educational Psychology*, 100(2): 235–251. <https://doi.org/10.1037/0022-0663.100.2.235>
- Ângelo, Catarina, and Ana B. Reis. 2021. "Gender gaps in different grading systems." *Education Economics*, 29(1): 105-119.
<https://doi.org/10.1080/09645292.2020.1853681>
- Auwarter, Amy E., and Mara S. Aruguete. 2008. "Effects of student gender and socioeconomic status on teacher perceptions." *The Journal of Educational Research*, 101(4): 242-246. <https://doi.org/10.3200/JOER.101.4.243-246>
- Baird, Jo-Anne. 1998. "What's in a name? Experiments with blind marking in A-Level examinations." *Educational research*, 40(2): 191-202.
<https://doi.org/10.1080/0013188980400207>
- Brackett, Mark A., James L. Floman, Claire Ashton-James, Lillia Cherkasskiy, and Peter Salovey. 2013. "The influence of teacher emotion on grading practices: A preliminary look at the evaluation of student writing." *Teachers and teaching*, 19(6): 634-646.
<https://doi.org/10.1080/13540602.2013.827453>
- Bradley, Robert H., Robert F. Corwyn, Harriette P. McAdoo, and Cynthia G. Coll. 2001. "The home environments of children in the United States part I: Variations by age, ethnicity, and poverty status." *Child development*, 72(6): 1844-1867.
<https://doi.org/10.1111/1467-8624.t01-1-00382>

- Buckingham, Jennifer, Kevin Wheldall, and Robin Beaman-Wheldall. 2013. "Why poor children are more likely to become poor readers: The school years." *Australian Journal of Education*, 57(3): 190-213. <https://doi.org/10.1177%2F0004944113495500>
- Dee, Thomas S. 2005. "A teacher like me: Does race, ethnicity, or gender matter?" *American Economic Review*, 95(2): 158-165. DOI: 10.1257/000282805774670446
- Delap, Martin R. 1995. "Teachers' estimates of candidates' performances in public examinations." *Assessment in Education: Principles, Policy & Practice*, 2(1): 75-92. <https://doi.org/10.1080/0969594950020106>
- Gibbons, Stephen, and Arnaud Chevalier. 2008. "Assessment and age 16+ education participation." *Research Papers in Education*, 23(2): 113-123. <https://doi.org/10.1080/02671520802048638>
- Johnson, Wendy, Matt McGue, and William G. Iacono. 2007. "Socioeconomic status and school grades: Placing their association in broader context in a sample of biological and adoptive families." *Intelligence*, 35(6): 526-541. <https://doi.org/10.1016/j.intell.2006.09.006>
- Krkovic, Katarina, Samuel Greiff, Sirkku Kupiainen, Mari-Pauliina Vainikainen, and Jarkko Hautamäki. 2014. "Teacher evaluation of student ability: What roles do teacher gender, student gender, and their interaction play?" *Educational Research*, 56(2): 244-257. <https://doi.org/10.1080/00131881.2014.898909>
- Lavy, Victor. 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. <https://doi.org/10.1016/j.jpubeco.2008.02.009>

- Lindahl, Erica. 2007. "Comparing teachers' assessments and national test results: evidence from Sweden." Working Paper, No. 2007:24. Institute for Labour Market Policy Evaluation (IFAU), Uppsala. <http://hdl.handle.net/10419/78672>
- Lindahl, Erica. 2016. "Are teacher assessments biased? – evidence from Sweden." *Education economics*, 24(2): 224-238. <https://doi.org/10.1080/09645292.2015.1014882>
- Lloyd, Jean, and Lloyd Barenblatt. 1984. "Intrinsic intellectuality: Its relations to social class, intelligence, and achievement." *Journal of Personality and Social Psychology*, 46(3): 646-654. <https://content.apa.org/doi/10.1037/0022-3514.46.3.646>
- Marcenaro-Gutiérrez, Oscar D., and Luis A. Lopez-Agudo. 2016. "Testing the impact on educational achievement of expectations." Paper presented at the XXIII Encuentro de Economía Pública, Ourense. <https://dialnet.unirioja.es/descarga/articulo/5696543>
- Marcenaro-Gutiérrez, O., and Anna Vignoles. 2015. "A comparison of teacher and test-based assessment for Spanish primary and secondary students." *Educational Research*, 57(1): 1–21. <https://doi.org/10.1080/00131881.2014.983720>
- Martínez, José F., Brian Stecher, and Hilda Borko. 2009. "Classroom assessment practices, teacher judgments, and student achievement in mathematics: Evidence from the ECLS." *Educational Assessment*, 14(2): 78-102. <https://doi.org/10.1080/10627190903039429>
- Morgan, Paul L., George Farkas, Marianne M. Hillemeier, and Steven Maczuga. 2009. "Risk factors for learning-related behavior problems at 24 months of age: Population-based estimates." *Journal of abnormal child psychology*, 37(3): 401-413. doi: 10.1007/s10802-008-9279-8

- Orr, Amy J. 2003. "Black-white differences in achievement: The importance of wealth." *Sociology of education*: 281-304. <https://doi.org/10.2307/1519867>
- Reeves, David J., William F. Boyle, and Thomas Christie. 2001. "The relationship between teacher assessments and pupil attainments in standard test tasks at Key Stage 2, 1996–98." *British Educational Research Journal*, 27(2): 141-160.
<https://doi.org/10.1080/01411920120037108>
- Rosenthal, Robert, and Lenore Jacobson. 1968. "Pygmalion in the classroom." *The urban review*, 3(1): 16-20. <https://doi.org/10.1007/BF02322211>
- Sackett, Paul R., Matthew J. Borneman, and Brian S. Connelly. 2008. "High stakes testing in higher education and employment: Appraising the evidence for validity and fairness." *American Psychologist*, 63(4): 215–227. <https://doi.org/10.1037/0003-066X.63.4.215>
- Schutz, Paul A., and Michalinos Zembylas. 2009. "Introduction to advances in teacher emotion research: The impact on teachers' lives." In *Advances in teacher emotion research*, edited by Paul A. Schutz, and Michalinos Zembylas: 3-11. Springer, Boston, MA. https://doi.org/10.1007/978-1-4419-0564-2_1
- Sutton, Rosemary E., and Kevin F. Wheatley. 2003. "Teachers' emotions and teaching: A review of the literature and directions for future research." *Educational psychology review*, 15(4): 327-358. <https://doi.org/10.1023/A:1026131715856>
- van Bergen, Elsje, Titia van Zuijen, Dorothy Bishop, and Peter F. de Jong. 2017. "Why are home literacy environment and children's reading skills associated? What parental skills reveal." *Reading Research Quarterly*, 52(2): 147-160. <https://doi.org/10.1002/rrq.160>

White, Karl R. 1982. "The relation between socioeconomic status and academic achievement." *Psychological bulletin*, 91(3): 461. <https://doi.org/10.1037/0033-2909.91.3.461>

8. Appendix

Table A1: Means and standard errors of teachers' and exams' scores in Portuguese and Mathematics of the 6th (2012-15) and 9th grades (2008-18) - Mean comparison tests by exam score level

Subject/score	Gender	Number of observations	% SASE	Mean Teacher Score		Mean Gap (Teacher-Exam)		Diff. in mean gaps (7) – (6) T-stat
				Non SASE	SASE	Non SASE	SASE	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
6th grade								
Mathematics								
1	Boys	13795	66.4	2.19 (0.01)	2.17 (0.01)	1.187*** (0.01)	1.173*** (0.01)	-0.014 [-0.943]
	Girls	12343	67.3	2.23 (0.01)	2.21 (0.01)	1.226*** (0.01)	1.211*** (0.01)	-0.015 [-0.986]
2	Boys	63792	54.4	2.75 (0.01)	2.65 (0.01)	0.748*** (0.01)	0.648*** (0.01)	-0.1*** [-7.412]
	Girls	62912	55.5	2.8 (0.01)	2.67 (0.01)	0.803*** (0.01)	0.674*** (0.01)	-0.129*** [-9.056]
3	Boys	47442	39	3.38 (0.01)	3.25 (0.01)	0.383*** (0.01)	0.249*** (0.01)	-0.133*** [-7.845]
	Girls	46827	38.7	3.48 (0.01)	3.35 (0.01)	0.484*** (0.01)	0.353*** (0.01)	-0.131*** [-7.579]
4	Boys	33484	26.5	4.12 (0.01)	3.98 (0.02)	0.117*** (0.01)	-0.022** (0.02)	-0.139*** [-6.647]
	Girls	31798	26.8	4.22 (0.01)	4.08 (0.01)	0.224*** (0.01)	0.077*** (0.02)	-0.147*** [-6.992]
5	Boys	8392	16.7	4.74 (0.01)	4.66 (0.02)	-0.257*** (0.01)	-0.338*** (0.02)	-0.08*** [-3.983]
	Girls	7095	18.3	4.79 (0.01)	4.74 (0.02)	-0.209*** (0.01)	-0.265*** (0.02)	-0.056*** [-3.202]
Portuguese								
1	Boys	1367	70.1	2.42 (0.03)	2.38 (0.02)	1.416*** (0.03)	1.383*** (0.02)	-0.033 [-0.914]
	Girls	539	73.1	2.58 (0.05)	2.55 (0.03)	1.579*** (0.05)	1.548*** (0.03)	-0.031 [-0.56]
2	Boys	48965	55.8	2.86 (0.01)	2.77 (0.01)	0.862*** (0.01)	0.769*** (0.01)	-0.093*** [-9.536]
	Girls	34254	60	3 (0.01)	2.9 (0.01)	0.999*** (0.01)	0.904*** (0.01)	-0.094*** [-10.636]
3	Boys	77001	42.9	3.29 (0.01)	3.11 (0.01)	0.294*** (0.01)	0.11*** (0.01)	-0.185*** [-14.557]
	Girls	73415	46.1	3.44 (0.01)	3.25 (0.01)	0.439*** (0.01)	0.25*** (0.01)	-0.188*** [-14.968]
4	Boys	33647	27.9	3.99 (0.01)	3.72 (0.02)	-0.01 (0.01)	-0.284*** (0.02)	-0.274*** [-13.363]
	Girls	44750	31.4	4.12 (0.01)	3.89 (0.01)	0.121*** (0.01)	-0.111*** (0.01)	-0.233*** [-12.844]
5	Boys	2767	16	4.62 (0.01)	4.42 (0.03)	-0.38*** (0.01)	-0.581*** (0.03)	-0.201*** [-5.687]
	Girls	4963	20.1	4.69 (0.01)	4.54 (0.02)	-0.308*** (0.01)	-0.464*** (0.02)	-0.156*** [-7.039]
9th grade								
Mathematics								
1	Boys	42979	49.8	2.03 (0.01)	2.03 (0.01)	1.027*** (0.01)	1.035*** (0.01)	0.008 [0.714]

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	Girls	45101	54.3	2.07	2.09	1.074***	1.088***	0.014
				(0.01)	(0.01)	(0.01)	(0.01)	[1.337]
2	Boys	123309	43.3	2.62	2.58	0.625***	0.58***	-0.045***
				(0.01)	(0.01)	(0.01)	(0.01)	[-4.239]
	Girls	131422	47.4	2.62	2.58	0.625***	0.577***	-0.048***
				(0.01)	(0.01)	(0.01)	(0.01)	[-4.361]
3	Boys	88928	32.8	3.19	3.15	0.193***	0.149***	-0.045***
				(0.01)	(0.01)	(0.01)	(0.01)	[-3.874]
	Girls	90482	35.9	3.28	3.22	0.285***	0.217***	-0.068***
				(0.01)	(0.01)	(0.01)	(0.01)	[-5.093]
4	Boys	62383	23.1	3.87	3.8	-0.13***	-0.203***	-0.074***
				(0.01)	(0.02)	(0.01)	(0.02)	[-3.976]
	Girls	70902	25.8	3.99	3.93	-0.006	-0.071***	-0.064***
				(0.01)	(0.02)	(0.01)	(0.02)	[-3.556]
5	Boys	18967	15.3	4.62	4.58	-0.384***	-0.424***	-0.04***
				(0.01)	(0.02)	(0.01)	(0.02)	[-2.195]
	Girls	23153	18.3	4.68	4.64	-0.324***	-0.364***	-0.04***
				(0.01)	(0.01)	(0.01)	(0.01)	[-2.543]
Portuguese								
1	Boys	2347	51.6	2.27	2.26	1.272***	1.264***	-0.008
				(0.02)	(0.02)	(0.02)	(0.02)	[-0.36]
	Girls	894	60.3	2.5	2.49	1.501***	1.49***	-0.012
				(0.03)	(0.03)	(0.03)	(0.03)	[-0.276]
2	Boys	115389	44.3	2.83	2.79	0.826***	0.793***	-0.033***
				(0.01)	(0.01)	(0.01)	(0.01)	[-4.245]
	Girls	86229	50.8	2.96	2.92	0.963***	0.925***	-0.038***
				(0.00)	(0.00)	(0.00)	(0.00)	[-5.914]
3	Boys	154215	34.8	3.15	3.06	0.152***	0.056***	-0.096***
				(0.01)	(0.01)	(0.01)	(0.01)	[-11.357]
	Girls	169480	40.4	3.3	3.2	0.302***	0.197***	-0.104***
				(0.01)	(0.01)	(0.01)	(0.01)	[-11.554]
4	Boys	55275	23.3	3.79	3.6	-0.207***	-0.403***	-0.196***
				(0.01)	(0.02)	(0.01)	(0.02)	[-10.84]
	Girls	88697	28.1	3.97	3.79	-0.029***	-0.21***	-0.181***
				(0.01)	(0.01)	(0.01)	(0.01)	[-11.246]
5	Boys	5535	14	4.51	4.37	-0.49***	-0.627***	-0.137***
				(0.01)	(0.03)	(0.01)	(0.03)	[-4.602]
	Girls	11851	18.3	4.62	4.49	-0.381***	-0.513***	-0.132***
				(0.01)	(0.02)	(0.01)	(0.02)	[-6.576]

Note: Grading scale is 1 to 5. Standard errors in parenthesis below are corrected for clustering at the school level. The mean difference in columns (6) and (7) corresponds to mean teacher score minus mean exam score. In column (8) the difference tested is column (7) minus column (6). The T statistic in square brackets presented below corresponds to the T test for the difference in mean differences. In columns (6)-(8), *, **, *** denote statistical significance at 10%, 5% and 1%, respectively.

Table A2: Means and standard errors of teachers' and exams' scores in secondary education subjects (2008-18) - Mean comparison tests by exam score level

Subject/Score	Gender	Number of observations	% SASE	Mean Teacher Score		Mean Exam Score		Mean Gap (Teacher-Exam)		Diff. in mean gaps (9) – (8) T-stat
				Non SASE	SASE	Non SASE	SASE	Non SASE	SASE	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Geography										
1	Boys	34	35.3	11.27	11	2.32	2.95	8.95***	8.05***	-0.9
				(0.37)	(0.49)	(0.17)	(0.22)	(0.41)	(0.55)	[-1.306]
	Girls	123	45.5	11.84	11.34	2.74	2.7	9.099***	8.638***	-0.461
				(0.31)	(0.33)	(0.07)	(0.08)	(0.32)	(0.34)	[-0.975]
2	Boys	9511	32.3	11.52	11.44	7.78	7.75	3.736***	3.687***	-0.049
				(0.04)	(0.06)	(0.02)	(0.03)	(0.05)	(0.07)	[-0.558]
	Girls	20005	38.9	11.89	11.89	7.59	7.53	4.302***	4.364***	0.062
				(0.05)	(0.06)	(0.02)	(0.03)	(0.06)	(0.07)	[0.673]
3	Boys	25368	28.3	12.64	12.52	11.31	11.21	1.32***	1.31***	-0.016
				(0.04)	(0.07)	(0.01)	(0.02)	(0.04)	(0.07)	[-0.197]
	Girls	37108	34.7	13.24	13.11	11.22	11.13	2.019***	1.983***	-0.036
				(0.05)	(0.07)	(0.01)	(0.02)	(0.05)	(0.07)	[-0.417]
4	Boys	10871	22.9	14.91	14.68	14.83	14.74	0.078*	-0.059	-0.137*
				(0.05)	(0.09)	(0.01)	(0.03)	(0.05)	(0.09)	[-1.332]
	Girls	14213	27.9	15.6	15.39	14.8	14.71	0.801***	0.682***	-0.119
				(0.05)	(0.08)	(0.01)	(0.02)	(0.05)	(0.08)	[-1.195]
5	Boys	727	18.4	17.37	16.99	18.01	17.94	-0.633***	-0.945***	-0.312*
				(0.09)	(0.18)	(0.02)	(0.05)	(0.09)	(0.18)	[-1.552]
	Girls	703	21.8	17.68	17.6	18	17.95	-0.323***	-0.348***	-0.026
				(0.09)	(0.16)	(0.02)	(0.05)	(0.09)	(0.16)	[-0.141]
Total:	Boys	46511	27.7	13.07	12.72	11.63	11.13	1.446***	1.595***	0.149*
				(0.05)	(0.07)	(0.05)	(0.08)	(0.05)	(0.08)	[1.518]
	Girls	72152	34.4	13.45	13.12	11.13	10.59	2.327***	2.523***	0.195**
				(0.05)	(0.07)	(0.05)	(0.07)	(0.06)	(0.08)	[1.907]
History A										
1	Boys	311	34.1	10.85	11.34	2.53	2.6	8.322***	8.74***	0.418**
				(0.13)	(0.18)	(0.05)	(0.07)	(0.14)	(0.19)	[1.8]
	Girls	833	39.1	11.2	11.23	2.562	2.623	8.641***	8.604***	-0.037
				(0.11)	(0.14)	(0.03)	(0.04)	(0.12)	(0.15)	[-0.195]
2	Boys	8762	31.2	11.55	11.63	7.07	7.04	4.483***	4.597***	0.114
				(0.05)	(0.08)	(0.03)	(0.04)	(0.06)	(0.1)	[0.983]
	Girls	20585	36.7	11.91	11.96	7.02	6.93	4.898***	5.027***	0.129
				(0.05)	(0.07)	(0.02)	(0.03)	(0.07)	(0.09)	[1.182]
3	Boys	11520	29.4	12.75	12.83	11.14	11.1	1.61***	1.725***	0.114
				(0.06)	(0.08)	(0.02)	(0.02)	(0.06)	(0.09)	[1.09]
	Girls	25892	33.5	13.34	13.35	11.19	11.12	2.146***	2.223***	0.077
				(0.06)	(0.08)	(0.01)	(0.02)	(0.06)	(0.09)	[0.726]
4	Boys	5746	25.8	15.04	14.88	15.02	14.94	0.02	-0.057	-0.077
				(0.07)	(0.11)	(0.02)	(0.03)	(0.07)	(0.11)	[-0.596]

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
	Girls	13689	29.3	15.47	15.5	15.02	14.99	0.449***	0.507***	0.058
				(0.07)	(0.1)	(0.01)	(0.02)	(0.07)	(0.1)	[0.49]
5	Boys	869	20.9	17.34	16.95	18.27	18.14	-0.931***	-1.194***	-0.263*
				(0.09)	(0.17)	(0.03)	(0.05)	(0.08)	(0.16)	[-1.455]
	Girls	2127	24.7	17.45	17.35	18.26	18.16	-0.81***	-0.805***	0.006
				(0.07)	(0.12)	(0.02)	(0.03)	(0.07)	(0.12)	[0.041]
Total:	Boys	27208	29	13.03	12.87	10.89	10.46	2.136***	2.414***	0.278**
				(0.04)	(0.07)	(0.07)	(0.11)	(0.08)	(0.12)	[1.947]
	Girls	63126	33.4	13.52	13.33	10.94	10.4	2.574***	2.924***	0.351***
				(0.05)	(0.07)	(0.07)	(0.1)	(0.08)	(0.11)	[2.536]
Philosophy										
1	Boys	417	33.3	12.03	12.13	2.55	2.55	9.483***	9.584***	0.101
				(0.16)	(0.23)	(0.04)	(0.06)	(0.17)	(0.26)	[0.33]
	Girls	508	44.1	11.99	12.2	2.57	2.62	9.416***	9.576***	0.16
				(0.15)	(0.17)	(0.04)	(0.05)	(0.17)	(0.19)	[0.638]
2	Boys	6220	30.7	12.07	12.11	6.81	6.71	5.263***	5.394***	0.131
				(0.06)	(0.09)	(0.03)	(0.05)	(0.08)	(0.12)	[0.94]
	Girls	9267	38.4	12.51	12.47	6.9	6.84	5.611***	5.63***	0.02
				(0.06)	(0.08)	(0.03)	(0.04)	(0.08)	(0.1)	[0.155]
3	Boys	7140	27.7	13.37	13.23	11.13	11.03	2.243***	2.201***	-0.042
				(0.06)	(0.09)	(0.02)	(0.03)	(0.06)	(0.1)	[-0.373]
	Girls	11098	34.1	13.87	13.8	11.22	11.15	2.654***	2.646***	-0.008
				(0.06)	(0.08)	(0.02)	(0.02)	(0.06)	(0.08)	[-0.073]
4	Boys	3027	22.3	15.39	15.21	15.02	14.87	0.367***	0.337***	-0.03
				(0.07)	(0.13)	(0.02)	(0.04)	(0.08)	(0.13)	[-0.191]
	Girls	7311	27.3	15.84	15.72	15.1	14.99	0.748***	0.729***	-0.019
				(0.05)	(0.08)	(0.02)	(0.03)	(0.06)	(0.09)	[-0.178]
5	Boys	507	17.6	17.4	17.37	18.26	18.2	-0.863***	-0.834***	0.029
				(0.1)	(0.2)	(0.03)	(0.07)	(0.1)	(0.2)	[0.13]
	Girls	1302	21.2	17.44	17.05	18.27	18.13	-0.834***	-1.078***	-0.245**
				(0.07)	(0.13)	(0.02)	(0.04)	(0.07)	(0.13)	[-1.4]
Total:	Boys	16011	27.7	13.41	13.1	10.36	9.63	3.048***	3.464***	0.416**
				(0.05)	(0.08)	(0.1)	(0.16)	(0.1)	(0.16)	[2.205]
	Girls	29486	33.4	14.17	13.76	11.25	10.37	2.913***	3.39***	0.478***
				(0.05)	(0.07)	(0.09)	(0.13)	(0.1)	(0.13)	[2.942]
Biology & Geology										
1	Boys	576	25.9	10.9	10.87	2.65	2.67	8.25***	8.194***	-0.056
				(0.09)	(0.15)	(0.03)	(0.06)	(0.09)	(0.16)	[-0.3]
	Girls	571	34.3	10.98	10.8	2.67	2.72	8.31***	8.072***	-0.238*
				(0.09)	(0.12)	(0.03)	(0.04)	(0.09)	(0.12)	[-1.622]
2	Boys	28923	26	11.74	11.66	7.08	7.01	4.665***	4.646***	-0.018
				(0.03)	(0.05)	(0.02)	(0.03)	(0.04)	(0.06)	[-0.248]
	Girls	23130	31.7	12	11.92	7.01	6.9	4.989***	5.018***	0.029
				(0.04)	(0.05)	(0.02)	(0.03)	(0.04)	(0.06)	[0.377]
3	Boys	27401	22.3	13.67	13.52	11.13	11.03	2.539***	2.484***	-0.054
				(0.04)	(0.07)	(0.01)	(0.02)	(0.04)	(0.07)	[-0.642]
	Girls	23388	26.2	14.13	13.92	11.18	11.09	2.95***	2.835***	-0.115*
				(0.04)	(0.07)	(0.01)	(0.02)	(0.04)	(0.07)	[-1.371]

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
4	Boys	12549	18.2	16.18	15.93	14.96	14.84	1.217***	1.083***	-0.134*
				(0.04)	(0.08)	(0.01)	(0.03)	(0.04)	(0.08)	[-1.493]
	Girls	12861	21.5	16.43	16.22	15.05	14.91	1.378***	1.305***	-0.074
				(0.04)	(0.08)	(0.01)	(0.02)	(0.04)	(0.08)	[-0.818]
5	Boys	7020	12.5	18.15	17.99	18.08	18.1	0.068	-0.116	-0.184*
				(0.05)	(0.12)	(0.02)	(0.04)	(0.05)	(0.12)	[-1.403]
	Girls	1884	14.6	18.27	18.08	18.11	18.02	0.157***	0.057	-0.1
				(0.05)	(0.12)	(0.01)	(0.03)	(0.05)	(0.12)	[-0.778]
Total:	Boys	71115	22.9	13.48	13.03	10.38	9.72	3.1***	3.307***	0.207**
				(0.04)	(0.07)	(0.05)	(0.09)	(0.05)	(0.08)	[2.224]
	Girls	61834	27	14.02	13.45	10.76	9.9	3.257***	3.556***	0.299***
				(0.04)	(0.07)	(0.06)	(0.09)	(0.05)	(0.08)	[3.107]
Descriptive Geometry										
1	Boys	1054	30.8	11.67	11.7	2.34	2.26	9.326***	9.434***	0.108
				(0.15)	(0.24)	(0.03)	(0.05)	(0.16)	(0.26)	[0.358]
	Girls	1337	32.8	11.77	11.42	2.34	2.29	9.436***	9.126***	-0.31
				(0.13)	(0.19)	(0.03)	(0.04)	(0.15)	(0.21)	[-1.232]
2	Boys	4900	25.1	12.47	12.32	6.48	6.28	5.988***	6.039***	0.05
				(0.08)	(0.14)	(0.04)	(0.07)	(0.11)	(0.18)	[0.239]
	Girls	5145	28.5	12.59	12.46	6.27	6.13	6.325***	6.325***	-0.0003
				(0.09)	(0.15)	(0.04)	(0.07)	(0.12)	(0.19)	[-0.001]
3	Boys	3815	20.7	14.02	14.08	11.27	11.22	2.747***	2.861***	0.114
				(0.1)	(0.18)	(0.03)	(0.05)	(0.1)	(0.19)	[0.54]
	Girls	3125	22.5	14.3	14.13	11.19	11.07	3.112***	3.057***	-0.055
				(0.11)	(0.21)	(0.02)	(0.05)	(0.12)	(0.22)	[-0.222]
4	Boys	3586	17.1	15.88	15.84	15.37	15.31	0.514***	0.53***	0.016
				(0.1)	(0.2)	(0.02)	(0.05)	(0.1)	(0.2)	[0.07]
	Girls	2624	19.3	16.14	15.94	15.33	15.2	0.815***	0.736***	-0.079
				(0.11)	(0.21)	(0.03)	(0.05)	(0.11)	(0.22)	[-0.315]
5	Boys	4453	13.5	18.06	18.1	19.14	19.15	-1.078***	-1.056***	0.021
				(0.07)	(0.17)	(0.02)	(0.04)	(0.17)	(0.17)	[0.116]
	Girls	2744	14.2	18.22	17.88	19.11	18.98	-0.886***	-1.1***	-0.213
				(0.07)	(0.17)	(0.02)	(0.05)	(0.07)	(0.17)	[-1.149]
Total:	Boys	17808	20	14.98	14.24	12.56	10.74	2.419***	3.495***	1.076***
				(0.07)	(0.13)	(0.19)	(0.36)	(0.17)	(0.32)	[3.01]
	Girls	14975	23.4	14.7	13.77	11.31	9.38	3.393***	4.385***	0.992***
				(0.08)	(0.14)	(0.21)	(0.37)	(0.19)	(0.32)	[2.672]
Mathematics Applied to SS										
1	Boys	309	36.9	10.71	10.33	2.31	2.35	8.403***	7.988***	-0.415*
				(0.16)	(0.21)	(0.07)	(0.09)	(0.18)	(0.23)	[-1.428]
	Girls	522	37.9	10.87	10.57	2.41	2.41	8.456***	8.166***	-0.29*
				(0.12)	(0.16)	(0.05)	(0.06)	(0.13)	(0.16)	[-1.409]
2	Boys	5445	32.9	11.04	11.01	7.07	7.05	3.971***	3.959***	-0.012
				(0.05)	(0.07)	(0.03)	(0.05)	(0.07)	(0.1)	[-0.1]
	Girls	11148	37.8	11.53	11.45	7.11	7.05	4.414***	4.404***	-0.01
				(0.05)	(0.07)	(0.03)	(0.03)	(0.07)	(0.09)	[-0.091]

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
3	Boys	6274	30.6	12.06	12.06	11.16	11.15	0.894***	0.918***	0.024
				(0.06)	(0.09)	(0.02)	(0.03)	(0.06)	(0.09)	[0.223]
	Girls	13343	34	12.87	12.73	11.25	11.19	1.621***	1.536***	-0.085
				(0.06)	(0.09)	(0.01)	(0.02)	(0.07)	(0.09)	[-0.745]
4	Boys	3390	28.3	14.09	14.09	15.03	14.97	-0.936***	-0.883***	0.053
				(0.07)	(0.11)	(0.03)	(0.04)	(0.07)	(0.12)	[0.39]
	Girls	8864	30.9	14.97	14.9	15.12	15.08	-0.153***	-0.18**	-0.027
				(0.07)	(0.1)	(0.02)	(0.03)	(0.07)	(0.1)	[-0.212]
5	Boys	743	24	16.77	16.5	18.39	18.16	-1.623***	-1.665***	-0.042
				(0.12)	(0.2)	(0.03)	(0.06)	(0.11)	(0.18)	[-0.197]
	Girls	2322	27.2	17.15	16.85	18.36	18.3	-1.215***	-1.452***	-0.237**
				(0.07)	(0.12)	(0.02)	(0.03)	(0.07)	(0.11)	[-1.747]
Total:	Boys	16161	30.7	12.38	12.19	10.88	10.46	1.504***	1.739***	0.235*
				(0.04)	(0.06)	(0.09)	(0.13)	(0.09)	(0.14)	[1.447]
	Girls	36199	34	13.29	12.95	11.42	10.86	1.869***	2.089***	0.22*
				(0.05)	(0.07)	(0.09)	(0.12)	(0.09)	(0.13)	[1.407]
Physics & Chemistry										
1	Boys	1355	27.9	10.47	10.42	2.65	2.64	7.815***	7.78***	-0.035
				(0.06)	(0.1)	(0.02)	(0.03)	(0.06)	(0.1)	[-0.286]
	Girls	1614	34.2	10.81	10.65	2.68	2.65	8.12***	8***	-0.125
				(0.07)	(0.1)	(0.02)	(0.03)	(0.07)	(0.1)	[-0.981]
2	Boys	28281	24.6	11.51	11.42	6.8	6.61	4.718***	4.81***	0.092
				(0.03)	(0.05)	(0.02)	(0.04)	(0.04)	(0.08)	[1.033]
	Girls	32206	28.9	11.93	11.87	6.8	6.64	5.139***	5.233***	0.095
				(0.03)	(0.05)	(0.02)	(0.03)	(0.05)	(0.07)	[1.113]
3	Boys	19515	19.7	13.54	13.49	11.15	11.05	2.398***	2.435***	0.036
				(0.05)	(0.09)	(0.01)	(0.02)	(0.05)	(0.09)	[0.35]
	Girls	20949	24.7	14.08	13.98	11.1	11.03	2.988***	2.946***	-0.042
				(0.05)	(0.09)	(0.01)	(0.02)	(0.05)	(0.09)	[-0.409]
4	Boys	10498	16.2	16.16	16.06	15.17	15.05	0.989***	1.017***	0.028
				(0.05)	(0.1)	(0.01)	(0.03)	(0.05)	(0.1)	[0.248]
	Girls	9958	19.4	16.54	16.41	15.14	14.98	1.404***	1.427***	0.023
				(0.05)	(0.09)	(0.01)	(0.03)	(0.05)	(0.09)	[0.219]
5	Boys	3330	11.7	18.44	18.28	18.4	18.36	0.033	-0.083	-0.115
				(0.04)	(0.1)	(0.01)	(0.04)	(0.04)	(0.11)	[-1.022]
	Girls	3059	13.5	18.65	18.42	18.36	18.26	0.292***	0.166**	-0.126
				(0.04)	(0.09)	(0.01)	(0.03)	(0.04)	(0.1)	[-1.184]
Total:	Boys	62979	21.1	13.37	12.79	10.26	9.21	3.111***	3.577***	0.467***
				(0.04)	(0.07)	(0.07)	(0.13)	(0.06)	(0.12)	[3.502]
	Girls	67786	25.6	13.67	13.12	9.99	9.02	3.68***	4.1***	0.416***
				(0.04)	(0.07)	(0.07)	(0.12)	(0.06)	(0.11)	[3.361]
Economics A										
1	Boys	103	25.2	11.3	10.69	2.58	2.93	8.722***	7.758***	-0.964**
				(0.22)	(0.38)	(0.07)	(0.13)	(0.26)	(0.43)	[-1.926]
	Girls	92	27.2	11.84	11.92	2.76	2.62	9.07***	9.3***	0.232
				(0.29)	(0.5)	(0.08)	(0.12)	(0.31)	(0.54)	[0.372]
2	Boys	4302	20.8	11.64	11.53	7.08	7.13	4.558***	4.397***	-0.16
				(0.08)	(0.14)	(0.04)	(0.07)	(0.1)	(0.17)	[-0.815]

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
	Girls	3343	28.2	12.18	12.02	7.08	7.07	5.106***	4.956***	-0.15
				(0.09)	(0.14)	(0.04)	(0.06)	(0.11)	(0.17)	[-0.745]
3	Boys	5942	19.4	12.94	12.77	11.24	11.17	1.699***	1.597***	-0.102
				(0.09)	(0.17)	(0.02)	(0.04)	(0.09)	(0.17)	[-0.519]
	Girls	4798	24.9	13.72	13.36	11.32	11.22	2.397***	2.141***	-0.256*
				(0.1)	(0.15)	(0.02)	(0.04)	(0.1)	(0.16)	[-1.395]
4	Boys	4737	15.8	15.33	15.18	15.25	15.18	0.078	-0.005	-0.083
				(0.08)	(0.16)	(0.02)	(0.05)	(0.08)	(0.16)	[-0.464]
	Girls	4380	21.5	15.83	15.68	15.29	15.31	0.54***	0.376***	-0.165
				(0.08)	(0.14)	(0.02)	(0.04)	(0.09)	(0.15)	[-0.969]
5	Boys	1370	13.9	17.37	17.29	18.33	18.33	-0.96***	-1.038***	-0.078
				(0.07)	(0.17)	(0.02)	(0.05)	(0.07)	(0.16)	[-0.442]
	Girls	1464	18.2	17.69	17.53	18.33	18.27	-0.631***	-0.741***	-0.11
				(0.08)	(0.14)	(0.02)	(0.04)	(0.07)	(0.14)	[-0.704]
Total:	Boys	16451	18.3	13.7	13.27	11.95	11.35	1.75***	1.918***	0.168
				(0.07)	(0.13)	(0.11)	(0.2)	(0.11)	(0.21)	[0.718]
	Girls	14077	23.9	14.48	13.95	12.37	11.69	2.111***	2.261***	0.15
				(0.07)	(0.12)	(0.11)	(0.18)	(0.11)	(0.18)	[0.697]

Note: Grading scale is 0 to 20. Standard errors in parenthesis below are corrected for clustering at the school level. The mean difference in columns (8) and (9) corresponds to mean teacher score minus mean exam score. In column (10) the difference tested is column (9) minus column (8). The T statistic in square brackets presented below corresponds to the T test for the difference in mean differences. In columns (8)-(10), *, **, *** denote statistical significance at 10%, 5% and 1%, respectively.

Table A3: Socioeconomic gap in assessments: remaining secondary education subjects

Subjects	Gender	Number of observations	(1) OLS	(2) School FE	(3) Teacher FE	(4) Class FE
Biology & Geology	Boys	71,115	0.200*** (0.0320)	0.0702*** (0.0221)	0.0497** (0.0195)	0.163*** (0.0294)
	Girls	61,834	0.294*** (0.0323)	0.132*** (0.0218)	0.118*** (0.0213)	0.242*** (0.0295)
Descriptive Geometry	Boys	17,808	0.984*** (0.108)	0.492*** (0.0726)	0.497*** (0.0691)	0.723*** (0.0920)
	Girls	14,975	0.919*** (0.112)	0.459*** (0.0806)	0.415*** (0.0761)	0.673*** (0.101)
Physics & Chemistry	Boys	62,979	0.465*** (0.0446)	0.210*** (0.0260)	0.164*** (0.0241)	0.422*** (0.0407)
	Girls	67,786	0.447*** (0.0404)	0.188*** (0.0236)	0.154*** (0.0212)	0.384*** (0.0355)
Economics A	Boys	16,451	0.122* (0.0702)	0.0388 (0.0560)	0.0905* (0.0534)	0.0779 (0.0621)
	Girls	14,077	0.156** (0.0788)	0.0945* (0.0557)	0.0683 (0.0529)	0.0972 (0.0700)
History A	Boys	27,208	0.231*** (0.0521)	0.0158 (0.0407)	0.000974 (0.0398)	0.128*** (0.0460)
	Girls	63,126	0.319*** (0.0438)	0.0606** (0.0251)	0.0635*** (0.0218)	0.212*** (0.0365)
Philosophy	Boys	16,011	0.425*** (0.0805)	0.119* (0.0609)	0.143** (0.0555)	0.413*** (0.0827)
	Girls	29,486	0.466*** (0.0555)	0.217*** (0.0389)	0.189*** (0.0368)	0.429*** (0.0534)
Geography	Boys	46,511	0.157*** (0.0375)	0.0592** (0.0277)	0.0588** (0.0264)	0.107*** (0.0351)
	Girls	72,152	0.205*** (0.0304)	0.0729*** (0.0206)	0.0573*** (0.0188)	0.135*** (0.0276)
Mathematics Applied To Social Sciences	Boys	16,161	0.226*** (0.0630)	0.0893* (0.0527)	0.0966* (0.0496)	0.152** (0.0592)
	Girls	36,199	0.223*** (0.0513)	0.105*** (0.0346)	0.118*** (0.0328)	0.145*** (0.0446)
Age			Yes	Yes	Yes	Yes
Year FE			Yes	Yes	Yes	Yes
School FE			No	Yes	No	No
Teacher FE			No	No	Yes	No
Class FE			No	No	No	Yes

Note: Dependent variable is the assessment gap, *TeacherScore* - *ExamScore*. Only the coefficient of interest β_1 is presented. Standard errors in parenthesis are robust and corrected for clustering at the school level in all models. *, **, *** denote statistical significance at 10%, 5% and 1%, respectively.