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.

A new measure of Ability Peer Effects: National Exams score vs. School GPA, evidence from Portugal

David Vieira Marques da Costa

Work project carried out under the supervision of:

Ana Balcão Reis, Luís Catela Nunes & Maria do Carmo Seabra

4-01-2021

Abstract

We study ability peer effects at the class-level for 9th grade students in Portuguese state-owned schools. We resort to school-by-year and school-by-year-by-teacher fixed effects to lessen identification threats stemming from unobserved sorting of students to peer groups. We measure students and peer ability with two alternative forms of achievement: teacher assessment along a broader set of subjects and grades and scores in 6th grade external exams. Our findings suggest positive peer effects when using lagged National Exams and negative peer effects when measured using lagged teacher assessment of students. We discuss our results and provide suggestions for future research.

Keywords: peer effects, teacher assessment, ability, achievement

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). We thank DGEEC - Ministry of Education of Portugal for providing the data used in this project. We would also like to thank João Firmino, Pedro Freitas and Rui Costa for their valuable comments. The Master's degree student would like to thank and dedicate his work to his family.

1. Introduction

The topic of peer effects has consolidated its position among the relevant research themes of Economics of Education. Studying peer effects in Education can be at the same time challenging and valuable, given the broad range of peers' characteristics and educational contexts out of which students can be affected by their peers. On the one hand, it is challenging in the sense that proper identification of peer influences implies tackling different threats such as achievement reflection and unobserved selection of peer groups. On the other hand, assessing the existence and magnitude of peer effects can enhance educational achievement trough optimal education groups' composition conditional on the existing peer effects. Past studies on peer effects look to peer influences emanating from peers' ability, gender, ethnicity, income-level composition, among other features. Our work, however, focus only on ability peer effects, although it accounts for other types of peer group composition.

Our research uses a large micro dataset covering a broad range of academic and socio-demographic features of students enrolled in state-owned schools of the Portuguese Educational System along 6 cohorts. We aim at identifying ability peer effects of 9th grade peers on 9th grade National Exams student achievement. The main contribution of our research is the introduction of a students' ability measurement which measures ability across different subjects through a two-grade period using teacher assessment, alongside a more standard measure of achievement using lagged National Exams' achievement. We aim at finding whether measuring peers' ability under a more comprehensive students' assessment, in terms of curriculum and time, differs and, if so, in what ways from measuring ability of peers using Exam achievement. We address the potential identifications threats resulting from unobserved selection by restricting our analyses to samples using within school-by-year and school-by-year-by-teacher variation.

The rest of our work is structured in the following way. Section 2 provides a scope of the existing literature in ability peer effects. Section 3 describes the Portuguese Education setting underlying our research and a detailed description of the data used. Section 4 provides the econometric framework used for our estimation of ability peer effects and Section 5 provides our estimation results. Section 6 presents some robustness tests. The discussion of our results appears in Section 7. Section 8 summarizes our conclusions.

2. Literature Review

Along recent decades, we have witnessed the flourishment of peer effects related literature within the context of Economics of Education. Departing from the assumption that students' achievement does not occur by pure chance or random process, economists have studied and empirically quantified the channels through which peers can influence students' achievement. However, identifying the size and/or the mechanisms of peer effects on achievement is unanimously considered a hard task.

There are three main threats to the identification of peer effects, as pointed out in Manski (1993) and Moffit (2001). First, there is what is commonly mentioned as the Reflection Problem, emerging from the simultaneous effects that students' and their peers have on each other achievements. This is, student *i* achievement in a specific class or grade, will be influenced by her peers' (-*i*) achievement which is simultaneously influenced by student *i* achievement. Second, within the peers' context, there are two distinct effects that are very difficult to identify separately: the "endogenous" effects (coming from the peers' achievement) and the "exogenous" or "contextual" effects (emerging from peer background characteristics that also affect achievement: family traits, ethnicity, etc.). To be able to control for all exogenous effects correlated with ability is beyond reasonable aspirations of most researchers even with large micro datasets as the one used in this study.

Finally, non-random peer group formation can also threaten the identification of peer effects. If students are sorted into classes and schools based on observable characteristics and the researcher can measure those observables, controlling for such would hamper the selection problem. **Firmino et al. (2020)** provides evidence in favour of sorting on observable characteristics within Portuguese public schools' context, reporting mild levels of segregation across and within schools at three dimensions: academic performance, income-level and foreign background. However, if there are unobserved traits determining the selection of students to schools or classes, the estimated coefficients of peer effects would still be biased. It is reasonable to assume that students are not placed in schools and classes randomly due to the beliefs that school boards and/or parents may have regarding what is the best environment for students. Not accounting for such unobserved endogenous peer group formation could produce very misleading estimates of peer effects since these estimates would be capturing the effect of belonging to a group of peers with similar unobserved traits that affect achievement.

Some studies such as Duflo et al. (2011) and Sacerdote (2001) were able to circumvent the selection threats above mentioned by studying random allocation of students to educational groups. However, the scarce availability of natural experiments or randomized programs generating random placement of students impelled economists to rely on other methods to avoid unobserved selection. The use of fixed effects estimations (removing common variation between schools, classes and even students along different periods) is a popular approach to limit the identification obstacles to peer effects.

Throughout the last two decades, the literature has presented different estimates of linear-in-means ability peer effects based on of different fixed effects specifications. Hanushek et al. (2003) use data on Texas' public schools along three cohorts and estimates peer coefficients on 6^{th} graders using two period lagged state test scores. The paper finds positive and significant ability peer effects estimates in Math at the school level. Within the European context, Lavy et al. (2012) and Gibbons & Telhaj (2008) look to national exam scores of 14 years old students in England and use lagged national exam scores (11 years old) as their peer ability measure finding no significant peer effects in linear models at the grade level. In Sweden, Sund (2009) finds positive and significant peer effects for high school classmates, measuring peer ability with last compulsory school grade (age 16) achievement in subjects common to both schools' stages (Swedish, English, Math and Physical Education). Ammermueller & Pischke (2006) found positive and significant peer effects from increases in average peer ability in reading achievement in six different European countries. In Portugal, Firmino et al. (2018) found negative and non-significant peer effects, from increases in class shares of higher achieving peers, on 6th and 9th grade National Exams achievement, respectively. Their study uses lagged National Exam scores (at the 4th and 6th grade levels) to measure peer ability.

Our Proposal differs from previous approaches by including, alongside the usual measurement of ability using external assessment, a measure of peer ability that accounts for teacher assessment of students in schools along two consecutive grades and a broad set of subjects (2nd Cycle GPA for the 5th and 6th grades). Although, Sund (2009) also measures achievement using school internal evaluation of students, we could not find any other study of ability peer effects using both types of assessment (teacher and external) to measure peer ability simultaneously. Cerdeira et al. (2018) and Silva, PL. et al. (2020) compare the predictive power that Secondary School GPA and National External scores have on students' Higher Education GPA, in Portugal. These studies find that both measures significantly explain College GPA and that teacher assessment (Secondary School GPA) have a higher explanatory power than National Exams. Such results motivate the use of teacher assessment of student's achievement as a measure of peer ability, alongside external assessment of students.

3. Data & Institutional Context

The Portuguese Education System (PES) is composed of three stages: pre-school, basic-school and secondary-school. Basic schooling is divided into three different cycles: the 1^{st} (from 1^{st} to 4^{th} grade), the 2^{nd} (from 5^{th} to 6^{th} grade) and 3^{rd} (from 7^{th} to 9^{th} grade). Secondary schooling constitutes a single academic cycle composed of the 10^{th}, 11^{th} and 12^{th} grades. The expected entering age of students in the 1^{st} Cycle is between 5 to 6 years old, depending on whether the student was born in the first or second semester of the civil year. Analogously, for a student with no retentions, students' expected age at the beginning of each cycle is between 9 and 10 for the 2^{nd} Cycle, 11 and 12 for the 3^{rd} Cycle and 14 and 15 for Secondary schooling.

In each school year of the 2^{nd} and 3^{rd} Cycles, students are enrolled in a single class with peers of the same grade.[1] Students in each grade study different subjects that are taught separately to each class by an assigned teacher which can, however, teach more than one class per school and grade but in different schedules. School years are divided in three terms with each student receiving a score, ranging between 1 to 5 points, per term in each subject. The last term scores correspond to the final teacher assessment of students' achievement along the school year in each subject, hence, we use these scores to calculate students Grade Point Average (GPA). We decided to restrict our analyses to five subjects (Portuguese, Mathematics, English, History & Geography and Natural Sciences), although students' curriculum include more subjects. These subjects comprehend a broad range of topics taught in the 2^{nd} Cycle, including those evaluated through external national exams (Portuguese and Mathematics). Therefore, we believe that these subjects capture the core academic curriculum and achievement of students in the 2^{nd} Cycle. At the end of 2^{nd} and 3^{rd} Cycles (6^{th} and 9^{th} grades respectively) students take mandatory National Exams in Portuguese and Math.

[1] It is very unlikely to find classes with students belonging to different grades within the 2^{nd} and 3^{rd} cycles.

The data used in this project comes from two different databases. The first is *MISI* database which is an administrative database collected and provided by the Portuguese Ministry of Education containing academic records of all students enrolled in the Portuguese Public Educational System between 2006/07 and 2017/18 school years. *MISI* also provides relevant socio-demographic information about students such as gender, household income status, parents' education and geographic origin. The second is *Júri Nacional de Exames* (*JNE*) from which we retrieved data on National Exams taken by students (score obtained, the school year and the exam phase). The information in both databases can be matched by a single student identifier variable.

We analyse 9th grade students enrolled in Continental Portugal state-owned schools along six cohorts (from 2012/13 to 2017/18) for which we had viable data on 5th, 6th and 9th grades' achievement and socio-demographic characteristics.[2] Although students can choose different curriculums in the 3rd Cycle, we restricted our sample to those in the Regular Academic Curriculum (RAC)[3] which represents the biggest share of enrolments in Continental Portugal state-owned schools with 92% in 2017/18 for the 2nd and 3rd Cycles (see DGEEC (2019)).

Regarding the *MISI* data, there were several issues that had to be tackled in order to have a cleaned data profile for each student. These issues included enrolments in multiple schools in the same school year, enrolments in more than one class within the same school and school year and duplicated enrolment observations for the same individual within the same school, class and school year. Since there was no feasible way to ensure which observations corresponded to the main school, class, or individual for each of the three cases, these observations were cut-out

[2] The first two school years (2006/07 and 2007/08) available in MISI have a different codification of subjects relative to the remaining school years which makes it impossible to calculate 5th and 6th grade GPA.
[3] RAC offers a standard curriculum and evaluation of all students is similar. .

from our analysis.[4] Nevertheless, the dropped observations constituted marginal shares of the database (4.72% of the original number of observations).

To measure 2nd Cycle GPA, we had to restrict our analyses to those students with complete last term academic information on the five subjects of interest for their last 5th and 6th grades. Our 2nd Cycle GPA is computed in two steps: first, we compute students' GPA across the five subjects for the 5th and 6th grades separately and then we take the mean of the two GPA as our teacher assessment measure of ability in the 2nd Cycle. To calculate our external achievement measure of ability, we required students to have a complete record of the 6th grade National Exams scores. Likewise, to generate our outcome variable, we required complete record of students' 9th grade National Exams scores. The last two measures of achievement are constructed on the average score obtained by students on Portuguese and Math external exams taken at the end of each respective academic cycle. Since we look to the end of the 6th grade as our baseline moment, we want to capture each student's National Exams performance when leaving the 2nd Cycle justifying the use of last 6th grade examinations. Conversely, since we use 9th grade National Exams as our output, we want to grasp students' performance at their first encounter with such examination, hence, we use the first 9th grade National Exams only to prevent any bias emerging from students already acquainted with such examination.

Additionally, to use teacher fixed effects, we only included students for which we had 9th grade Portuguese and Mathematics teacher information. Furthermore, we restrain our analysis to a minimum reasonable class size of ten students, less restrictive than Firmino et al (2018)**,** and only used classes for which we kept at least 25% of the original class size after

[4] Furthermore, there were two additional issues at the individual identifier variable: for some individuals there were more than one birthdate and or chronological miscoding of grades. Concerning the latter case, some ids had "jump-backs" in grades, i.e., the same student would have completed grade 6th in school year *t* and grade 4th in school year *t+1* which doesn´t make sense. For the first case we kept the observations corresponding to the most frequent birthdate recorded for the same id. In the second case, we kept the correctly ordered sequence of observations, this is, all the observations until the one where the jump-back in grades would appear.

imposing our data restrictions mentioned in the above paragraph. Using Imbens t-test for means' differences[5] we found no statistically significant difference between our final sample and a sample of students with no class-size restriction (see Table A.1 in the Appendix).

Using *MISI* socio-demographic information we constructed several dummy variables. The socio-demographic characteristics include gender, internet and computer access at home, low-income status, parents' education level, parents' employment-status and country of birth.

For the first three characteristics, were generated three different dummies taking value one if the student is female, if the student has internet at home and if the student has a computer at home. Low-income household status was proxied by each student's eligibility to receive school social support, where a dummy was created giving students in such conditions value one.[6] Regarding Parents' Education we look to the highest education degree attained by each parent. There are two dummies for each parent, one that takes value one if the parent completed higher education and another that takes value one if the parent completed secondary education. We also distinguish employment-status across parents, this is, we have a dummy for whether the father is unemployed and another for whether the mother is unemployed. For the country of birth two dummies were generated for each household member (father, mother and student separately). The two dummies distinguish those born in a Foreign Portuguese Speaking Country (FPSC) from those born in a Foreign Non-Portuguese Speaking Country (FNPSC).[7] Under this approach we can identify those students with any foreign background within their families while distinguishing between foreign origin countries' level of familiarity with the Portuguese language, a trait that can facilitate learning and comprehension within class.

[5] We use this test for considering it the more appropriate test given the large sample sizes used (see Imbens (2015))
[6] The PES offers means-tested school benefits for students belonging to a household below a certain income level. There are two levels of social support that a students' household can receive and we define a student as a low-income if eligible to any of the two.
[7] FPSCs include Portugal's former colonies, where Portuguese is an official language: Angola, Brasil, Cabo Verde, Guiné-Bissau, Mozambique, São-Tomé e Príncipe and East-Timor.

Finally, given the Panel properties of *MISI* we generated dummies to account for whether the student reports any retention along the 2nd Cycle and to identify those individuals that remain with a substantial share of their initial peers along the 3rd cycle. This last dummy attributes value one if a given individual goes through consecutive 7th, 8th and 9th grades with at least 11 peers form her 7th grade class.

We found some missing/ambiguous records regarding some characteristics. Discarding all individuals with any kind of missing/ambiguous information would imply cutting a large portion of the available data. Instead, whenever the records on a given individual characteristic were fully missing or with even shares of contradictory values (e.g.: individual student's panels with equal number of male and female records), we opted for including them in the largest group of each characteristic and created "missing dummy variables" taking value one if the individual's characteristic information was missing/unconclusive.

Table 1 provides some descriptive statistics for Individual-Level variables. Our final sample has 300,638 students along six consecutive 9th grades cohorts (an average of 50,000 students per cohort) out of which 40% belong to low-income households and 73% have a computer at home. Regarding parents' education, 19% and 12% of mothers and fathers, respectively, have Higher Education. The percentage of unemployed mothers is higher than that of unemployed fathers and the percentage of foreign-born parents is close to 7%. We have a nearly balanced gender distribution (52% of female students) and 4% of students report at least one retention in the 2nd Cycle.

Table 2 reports Class-level descriptive statistics of the total 17,895 classes in our final sample (an average of 3,000 classes per cohort).[8] The average class size in our sample if 23 students and we find that class-size is significantly correlated with some social and academic

[8] Class-level variables were computed as the class share of individuals holding each respective academic and socio-demographic characteristic.

traits which leads to suspect that class-dimension may be used by schools as a students' sorting mechanism. Like Firmino et al. (2018), we find that class-size is significantly and negatively correlated with Class-Level variables signalling less favourable student's traits: 2nd Cycle retention and low-income status (-0.24 and -0.10 respectively). On the other hand, we find that students tend to find better achieving peers (independently of the measure type) in classes with higher size: correlation of 0.19 for teacher assessment measure and 0.18 for external assessment measure of ability.[9]

Table 1. Descriptive Statistics Individual-Level Variables

Individual Level Variable	N	mean	sd	Min	Max
Average 9th grade Exam Score	300,638	53.73	18.49	0	100
Peers' 2nd Cycle GPA	300638	3.68	0.29	2.58	4.79
Peers' Average 6th grade Exam Score	300638	3.16	0.38	1.75	4.52
Student's 2nd Cycle GPA	300,638	3.68	0.67	2	5
Student's Average 6th grade Exam Score	300,638	3.16	0.75	1	5
Same 3rd Cycle Peers	300,638	0.75	0.43	0	1
Internet	300,638	0.61	0.49	0	1
Computer	300,638	0.73	0.45	0	1
Mother Higher Education	300,638	0.19	0.39	0	1
Mother Secondary Education	300,638	0.22	0.41	0	1
Father Higher Education	300,638	0.12	0.32	0	1
Father Secondary Education	300,638	0.18	0.39	0	1
Mother Unemployment	300,638	0.10	0.30	0	1
Father Unemployment	300,638	0.05	0.22	0	1
Low Income	300,638	0.40	0.49	0	1
Female	300,638	0.52	0.50	0	1
Mother FPSC	300,638	0.05	0.21	0	1
Mother FNPSC	300,638	0.03	0.18	0	1
Father FPSC	300,638	0.04	0.21	0	1
Father FNPSC	300,638	0.03	0.16	0	1
Student FPSC	300,638	0.01	0.11	0	1
Student FNPSC	300,638	0.02	0.13	0	1
2nd Cycle Retention	300,638	0.04	0.18	0	1

Notes: FPSC refers to Foreign Portuguese Speaking Country while FNPSC refers to Foreign Non-Portuguese Speaking Country

[9] The presented correlations are all significant at 1% level.

Table 2. Descriptive Statistics Class-Level Variables

Class Level Variable	N	mean	sd	Min	Max
Class Size	17,895	22.52	4.12	10	33
Class Internet	17,895	0.61	0.27	0	1.00
Class Computer	17,895	0.72	0.25	0	1.00
Class Mother Higher Education	17,895	0.18	0.17	0	1.00
Class Mother Secondary Education	17,895	0.22	0.14	0	0.85
Class Father Higher Education	17,895	0.11	0.13	0	0.92
Class Father Secondary Education	17,895	0.18	0.13	0	0.86
Class Mother Unemployed	17,895	0.10	0.09	0	0.70
Class Father Unemployed	17,895	0.05	0.06	0	0.55
Class Female	17,895	0.52	0.14	0	1.00
Class Low Income	17,895	0.41	0.21	0	1.00
Class Mother FPSC	17,895	0.05	0.08	0	0.80
Class Mother FNPSC	17,895	0.03	0.05	0	0.57
Class Father FPSC	17,895	0.05	0.08	0	0.90
Class Father FNPSC	17,895	0.03	0.05	0	0.61
Class Student FPSC	17,895	0.01	0.04	0	0.50
Class Student FNPSC	17,895	0.02	0.04	0	0.50
Class 2nd Cycle Retention	17,895	0.04	0.06	0	0.73

Notes: Class-level variables were computed as the class share of individuals holding each respective characteristic. FPSC refers to Foreign Portuguese Speaking Country while FNPSC refers to Foreign Non-Portuguese Speaking Country.

4. Methodology

4.1 Econometric Specification

Our base model of students' achievement in the 9th grade is given by:

$$Y_{ifcst}^{9^{th}\ grade} = \beta_0 + \beta_1 \overline{GPA}_{i,(-i)cst}^{2^{nd}Cycle} + \beta_2 \overline{Y}_{i,(-i)cst}^{6^{th}\ grade} + \gamma X_{icst}^{6^{th} grade} + \alpha P_{cst}^{6^{th} grade} + \lambda CS_{cst} + \varepsilon_{ifcst} \qquad (1)$$

Y_{ifcst} is the standardized, within each school year, average score obtained in the 9th grade National Exams (Portuguese and Math) by student *i*, with teacher *f*, in class *c*, school *s* and school year *t*. 9th grade National Exams are graded on a scale from 0-100 points.[10]

The error term ε_{ifcst} can be decomposed as:

[10] Since we do not estimate separate models for Portuguese and Math achievement, the teacher subscript *f* should be read as the Portuguese and Math teachers' combination allocated to each class in a school and school year.

$$\varepsilon_{ifcst} = \pi_t + \omega_{st} + \eta_{stf} + v_{ifcst} \qquad (2)$$

Where π_t captures time specific effects such as differences in external assessment features faced by different cohorts of students, ω_{st} is a school-by-year specific component capturing unobserved shocks across schools within each school year and η_{stf} is a school-by-year-by-teacher specific component capturing unobserved shocks across teachers within each school and school year. Lastly, v_{ifcst} is a random component capturing the remaining unobserved variation in achievement of students within each class, teacher, school and school year.

Our peer ability variables of interest are $\overline{GPA}^{2^{nd}Cycle}_{i,(-i)cst}$ and $\overline{Y}^{6^{th}\ grade}_{i,(-i)cst}$.[11] The first measures ability using students' 2nd Cycle grade point average resulting from teacher assessment in five different subjects throughout 5th and 6th grades. The second uses achievement in the 6th grade National Exams of Portuguese and Math computed as the average of the two scores. Both measures use lagged achievement of students and are calculated as the *leave-out* mean of 9th grade peers' achievement, i.e., the achievement of student *i* is excluded from the computation of her 9th grade class mean achievement. Following Hanushek et al. (2003), Sund (2009), Lavy et al. (2012) and Firmino et al. (2018), the use of lagged achievement while excluding the individual score from mean peer achievement is important to avoid reflexivity bias as pointed out in Section 2. On the one hand, not including student *i*'s achievement enables us to assess peer mean ability without the simultaneous impact of student *i*'s performance. On the other hand, lagged achievement provides us with a *predetermined* measure of ability, hence avoiding reverse causality bias with relation to current achievement (at the 9th grade).

The model controls for class-level characteristics by including vector $P^{6^{th}grade}_{cst}$. Although we are not interested in estimating other peer effects besides ability, we consider important to

[11] $\overline{GPA}^{2^{nd}Cycle}_{i,(-i)cst}$ and $\overline{Y}^{6^{th}\ grade}_{i,(-i)cst}$ are computed on a scale of 1-5 points.

account for the observed composition of classes given the tendency to find segregation within schools, as mentioned in Section 2. These controls include the academic and socio-demographic characteristics of students' 9th grade peers measured at the base-line moment (6th grade).[12] Following Firmino et al. (2018), we also control for class-size of 9th grade classes (CS_{cst}) for two reasons: because it provides a more accurate interpretation of class composition effects such as peer ability and because schools may allocate students in a non-random fashion using class-size as a tool to accommodate their beliefs regarding optimal allocation of children.

The model also controls for *predetermined* (6th grade) baseline achievement and demographic traits of individuals with the vector $X_i^{6^{th} grade}$.[13] The inclusion of these individual-level characteristics allows us to partially control for within-school sorting of students since $X_i^{6^{th} grade}$ includes some of the most used school criteria for student allocation to classes. At the individual level we also control for whether students keep at least 11 of their peers throughout the 7th, 8th and 9th grades in consecutive school years and for the school year in which the student took the 6th Grade National Exams.

4.2 Addressing Non-Random Class formation.

The econometric specification presented in (1) overlooks important questions regarding selection of students into classes which, as explained in Section 2, constitute the main threat to the identification of peer effects. In other words, there are aspects of peer group formation which are not accounted for in (1) and, by remaining in the error term, are prone to bias considerably our estimates of ability peer effects.

---

[12] We control for the class share of students holding the following features: internet at home, computer at home, low-income, female gender, foreign born, foreign born parents, unemployed parents, higher educated parents, secondary educated parents and retention in the 2nd Cycle. Furthermore, we include missing dummies to account for the share of students with missing information in any category within each class.

[13] Individual-level characteristics include students': baseline achievement (both 2nd Cycle GPA and Average 6th grade Exam Score), gender, family background (internet and computer availability at home, parents' education, income-level, employment-status, country of birth), retention records in the 2nd Cycle and Same Peers composition along consecutive 7th, 8th and 9th grades.

As mentioned in Section 3, we aim to identify and estimate ability peer effects at 9th grade classes. Under the reasonable assumption that assignment of students to schools takes place within each school year, we introduce a school-by-year specific component (ω_{st}) in our estimation model to account for unobserved students' sorting across schools within the same school year. If indeed, parents attempt to enrol their children into their preferred schools and/or schools screen their admitted students following unobserved criteria, the exclusion of the common variation within schools in the same school year should account for the bias coming from such group effects. To estimate our model using school-by-year fixed effects we had to restrict our sample to those students enrolled in schools for which there were more than one class in each school year (291,235 students).

An additional source of endogenous group formation which (1) ignores is teacher allocation to classes within each school and school year. It is reasonable to assume that teachers are not allocated to classes within a school without their own interests, those of school principals and those of parents being considered. It could be the case that a given teacher with a higher bargaining power within a school pressures the school board to allocate him to classes with higher ability students or that more concerned parents pressure schools to allocate the better teachers to their children classes. To deal with teacher unobserved sorting to classes, we introduce a school-by-year-by-teacher specific component (η_{stf}) in our model. Using school-by-year-by-teacher fixed effects implied restricting our sample to students whose teachers (of Portuguese and Maths) teach at least two classes within each school and school year (122,497 students).[14] It is a common feature of our sample to find several classes within each school-year cell being taught by the same Portuguese and Math teacher pair. Therefore, controlling for teacher allocation for such classes implies controlling for the allocation of the teacher pair. For

[14] Both fixed effects' restricted samples are representative of the original sample revealing no statistical differences in means using Imbens t-test for all variables included in the model (see Table A.2 and Table A.3 in the Appendix).

the remaining classes that do not share the same pair of teachers but whose teachers comply with the restriction of teaching more than one class per school and school year, the allocation of teachers is controlled separately for Portuguese and Math teachers.

5. Results

Table 3 presents estimates of ability peer effects for our two measures of peer ability, under different specifications. In Column 1 we present our estimates of β_1 and β_2 using our basic specification (1) for which we find two significant effects with opposite signs. We find that increasing average peer 2nd Cycle GPA by one-point decreases 9th Grade individual achievement by 0.29 standard deviations, while, increasing average peer 6th Grade National Exams score by one-point increases 9th Grade achievement by 0.35 standard deviations. Column 2 provides estimates of ability peer effects controlling for school year effects. The pattern of ability peer effects remains the same in terms of statistical significance and sign.

When we consider our most restrictive specifications in Columns 3 and 4 using within school-by-year and school-by-year-by-teacher variation, respectively, our estimates of ability peer effects suffer sizable changes in magnitude although the significance levels and direction of the effects are unchanged. The estimates of β_2 (using lagged National Exams as our ability measure) decrease to 0.13 (Column 3) and 0.11 (Column 4) which is less than half of the estimated effects under our most basic specification. Regarding our alternative measure of ability (lagged teacher assessment) our estimates (first row Columns 3 and 4) also report a bias in our early estimates, although this time with a negative sign. The ability peer effects remain negative when measured through peers' 2nd Cycle GPA, but once we start accounting for non-random peer group formation, our estimate of β_1 reduces its magnitude to less than a quarter of our early estimates.

Table 3. Estimates of Ability Peer Effects on 9th Grade Average National Exam Score

	(1)	(2)	(3)	(4)
VARIABLES	OLS	Year Fixed Effects	School-Year Fixed effects	School-Year-Teacher Fixed effects
Peers' 2nd Cycle GPA	-0.292***	-0.308***	-0.065***	-0.066***
	(0.010)	(0.011)	(0.011)	(0.016)
Peers' 6th Grade National Exams	0.346***	0.373***	0.129***	0.110***
	(0.009)	(0.009)	(0.010)	(0.015)
Student's 2nd Cycle GPA	0.717***	0.722***	0.738***	0.734***
	(0.003)	(0.003)	(0.003)	(0.004)
Student's 6th Grade National Exams	0.405***	0.402***	0.386***	0.388***
	(0.002)	(0.002)	(0.003)	(0.004)
Additional Controls	YES	YES	YES	YES
Year FE		YES		
School x Year FE			YES	
School x Year x Teacher FE				YES
Observations	300,638	300,638	291,235	122,497
R-squared	0.661	0.678	0.711	0.722

Notes: Significance levels: *** p<0.01, ** p<0.05, * p<0.1. Robust standard errors clustered at the class level in parentheses. The outcome variable (Average 9th grade Exam Score) was standardized to have mean zero and standard deviation of one within each school year. All specifications have an intercept and use controls at the individual and class-level and dummies accounting for missing information for each individual and class-level control.

Assuming that the use of fixed effect brings us closer to the true peer effect parameter, by tackling the unobserved selection of students into peer groups, we can infer that our estimates of ability peer effects under the simplest specification (Column 1) are biased when compared with our most restrictive specifications (Columns 3 and 4). Furthermore, given the sizable change in size of our peer effects' estimates upon the introduction of school-by-year fixed effects and the similarity between estimates in Columns 3 and 4, we infer that unobserved sorting of students takes place, predominantly, across schools.

In line with Cerdeira et al. (2018) and Silva, PL. et al. (2020), our individual lagged achievement variables included in $X_i^{6^{th} grade}$ prove to be positively and significantly correlated with 9^{th} grade achievement (third and fourth rows of Table 3) across all specifications. The coefficients' estimates of individual achievement have higher magnitude than those of our ability peer measures and confirm the explanatory power of both ability measures at the individual level (2^{nd} Cycle GPA and Average 6^{th} grade Exam Score) regarding 9^{th} grade Average Exam achievement.

It is also worth mentioning that the estimated coefficients on Same 3^{rd} Cycle Peers control variable (controlling for whether the student kept at least eleven class peers along 7^{th}, 8^{th} and 9^{th} consecutive grades) are positive and significant (See Appendix Table A.4). Suggesting, therefore, that maintaining a stable peer composition along the 3^{rd} Cycle improves individual 9^{th} grade achievement.

Although our focus lays on ability measures, we also find, within our controls, that other individual and class dimensions produce significant estimates whose signs are in conformity with previous findings in peer literature (Table A.4). To provide some examples, having parents with a higher education degree increases 9^{th} grade achievement and, for our last two specifications, increasing the class share of parents holding such degree also produces positive and significant coefficients. In accordance with past findings (see Sacerdote (2011)) being in classes with higher shares of female students improves individual achievement. Conversely, belonging to a low-income household or attending classes with higher shares of low-income individuals significantly harms 9^{th} grade achievement. Likewise, recording at least one retention in the 2nd Cycle produces negative and significant estimates under all estimations, while being in classes with higher shares of students with 2nd Cycle retention produces negative and significant estimates except when we use teacher fixed effects.

6. Robustness Tests

We assess the robustness of our peer effects estimates under two alternative hypotheses. On one side, we estimate our peer models excluding all students that record any retention along the 3rd Cycle. On the other side, we check for the robustness of our estimates using school-by-year-by-teacher fixed effects if we restrict our sample to classes taught by the same pair of Portuguese and Math teachers, i.e., classes within each school and school year that share the same Portuguese teacher but do not share the same math teacher and vice-versa are excluded from our estimation.

Looking to the estimated coefficients when 3rd Cycle retention students are excluded (see Table A.5 provided in the Appendix Section), we find no changes in either statistical significance or sign of the estimated betas, across all specifications, when compared to the results in Table 3. [15] Ability peer effects remain positive when measured by lagged National Exams and negative when measured through 2nd Cycle teacher assessment. Differences in estimates' size are marginal for both ability measures in comparison with Table 3. Table A.5 also shows that, once fixed effects at the school and school-by-teacher level are introduced, the magnitude of our estimates change substantially, supporting the hypothesis that indeed our peer estimates under (1) are biased. It is also worth mentioning that the direction and dimension of such change is analogous to the one found in Table 3.

[15] Since students record no retention between their baseline year 9th grade exam years, all students within each cohort were assessed in the same 6th and 9th grade National Exams. Put together with the fact that all specifications control for the school years in which students took their 6th grade National Exams, using just year fixed effects adds nothing to our basic specification (1), therefore Table A.6 provides estimates for school-by-year and school-by-year-by-teacher fixed effects in addition to our basic model.

Our second test is useful to assess whether controlling for teacher allocation to classes as proposed in Section 4, i.e., controlling simultaneously for the allocation of a single teacher and/or teacher pairs, differs from controlling just for teacher pairs' allocation in terms of peer effects estimates. We are convinced that it does not since the peer estimates provided in the first and second row of Table A.6 (Appendix) show equal estimates, in terms of statistical significance and sign, to those found in school-by-year-by-teacher fixed effects specification of Table 3. Magnitudes of estimates does not seem to differ much, as well, from Table 3 ones (a change of -0.013 and 0.005 standard deviations for Peers' 2nd Cycle GPA and Peer's Average 6th grade Exam score, respectively).[16]

7. Discussion

Our results provide estimates of linear-in-means ability peer effects that suggest opposite effects depending on the type of lagged achievement used to measure peers' ability. On the one hand, when peer ability is measured by peers' achievement in external assessment (Average 6th grade Exam score) we find positive effects from increasing average achievement of peers, in line with several past studies (Hanushek et al. (2003), Sund (2009) and Ammermueller & Pischke (2006)).

On the other hand, when peer ability is measured with teacher assessment under a more comprehensive range of subjects (Portuguese, Mathematics, English, History & Geography and Natural Sciences) and grades (5th and 6th grades), we find that increasing average ability of peers in 9th grade classes decreases achievement in 9th grade National Exams. Although finding negative ability peer effects is not unprecedent in peer literature (Vigdor & Nechyba (2007)

[16] The restricted sample of used in this test does not show any significant difference, across all variables used, from the Final Sample (300,638 students) using Imbens t-test for differences in means (See Tables A.7 and A.8).

also find negative linear-in-means estimates for ability peer effects in reading and math), the conflicting nature of two alternative measures of ability is slightly more puzzling.

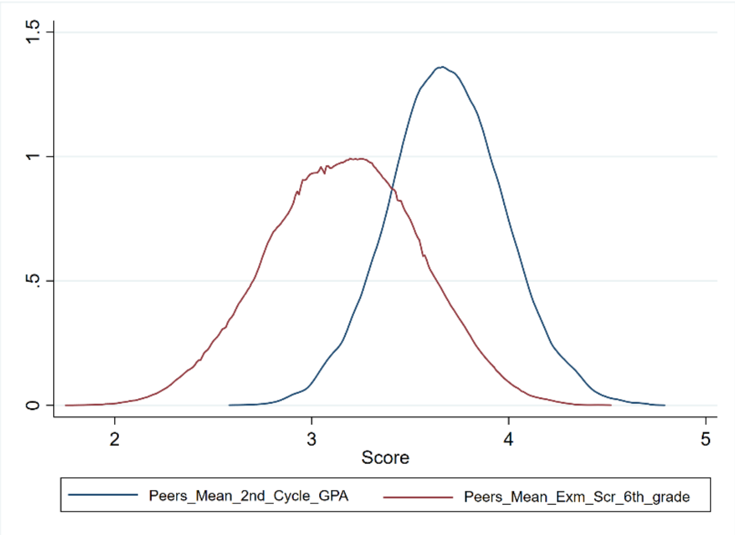
We propose two hypotheses for why signals differ depending on the way we measure ability. Firstly, we suspect that the high correlation (0.73) between our peer ability measures (2nd Cycle GPA and Average National Exams score) may be driving our results under simultaneous estimation of both types of ability peer effects. [17] In fact, when we estimate two separate models, each one using only one measure of ability (see Table A.9. and Table A.10), we see that our teacher assessment measure of ability produces very different estimates of ability peer effects than those found in Table 3. While peer ability measured through exams achievement (Table A.10) keeps the same pattern of peer effects estimates (all estimates positive and significant while smaller in magnitude), ability peer effects measured through teacher assessment (Table A.9) changes from negative and significant (OLS and Year FE specifications) to positive and significant (School-by-Year FE specification) and to non-significant (School-by-Year-by-Teacher FE).

Secondly, the difference in signs of ability peer effects could be caused by a mismatch between what teachers assess and what National Exams assess in terms of student ability. It could be the case that teachers' assessment contemplates characteristics of students (such as behaviour in class, compliance with assigned homework, among others) that National Exams assessment criteria fail to capture upon scores attribution. In turn, given the common assessment nature between 6th and 9th grade National Exams, it is less likely to find the same type of disparities found between students' 2nd Cycle GPA and 9th grade National Exams.

[17] The presented correlation concerns our Final Sample of 300,638 students. School-by-Year and School-by-Year-by-Teacher Fixed Effect's samples also show high correlation between the two measures (0.75). All correlations are significant at 1% level.

For a visual illustration of the divergence between teacher and external assessment, consider Figure. 1 that shows the distributions of our two peer ability measures. One can notice that the mean score of peers under teacher assessment differs from the one found in 6th grade National Exams. Furthermore, this relationship between teacher and external assessment can be found both in peer and individual measures of achievement (see Figure A.1 and A.2 in the Appendix).

Figure 1. Kernel density estimates of the two Peer Mean ability variables



8. Conclusion

In this paper we use comprehensive micro-data for the Portuguese Public Education System to study the existence of linear-in-means ability peer effects for 9th grade National Exams Achievement under two alternative measures of students' ability (teacher assessment vs external assessment). We find significant estimates of ability peer effects for both ability measures, even when controlling for unobserved peer group sorting at School-by-Year and School-by-Year-by-Teacher level. Additionally, we find that ability peer effects differ in sign depending on the measure of ability.

For ability measured with Average 6th grade Exams score, we find positive effects from a one-point increase in mean peers' average exam score which is in conformity to the general finds in peer literature. On the contrary, when we use 2nd Cycle GPA to measure peers' ability, we get negative peer effects from marginal increases in mean peers' 2nd Cycle GPA. We suggest that this disparity in direction of ability peer effects may be a consequence of two facts: high correlation between both measures and mismatch between teacher and National Exams ability assessment criteria. Hence, we consider that controlling for 2nd Cycle teacher specific effects would be an interesting starting point for future research. Future work could, also, try an alternative specification using 3rd Cycle teacher assessment (3rd Cycle GPA) as the outcome variable, instead of 9th grade National Exam score, to assess whether different measures of ability peer effects maintain, or not, the same direction pattern found in our work. Moreover, it would be interesting to further explore the ability distribution of both measures by estimating non-linear ability peer effects, which can differ substantially from simpler linear-in-means estimates (see Lavy et al. (2012)).

References

Ammermueller, A., & Pischke, J.S., (2006). "Peer effects in European primary schools: Evidence from PIRLS." *NBER Working Paper* No. *12180.*

Cerdeira, J. M., Nunes, L. C., Reis, AB. & Seabra, C. (2018). "Predictors of Student Success in Higher Education: Secondary School Internal Scores Versus National Exams." *Higher Education Quarterly* 72 (4): 304–313.

Direção-Geral de Estatísticas da Educação e Ciência (DGEEC) (2019). Estatísticas da Educação 2017/2018. Lisboa: Direção-Geral de Estatísticas da Educação e Ciência.

Duflo, E., Dupas, P, & Kremer, M., (2011). “Peer Effects, Teacher Incentives, and the Impact of Tracking: Evidence from a Randomized Evaluation in Kenya.” *American Economic Review*, 101(5): 1739–1774.

Firmino, J., Nunes, L. C., Almeida, S. D., & Batista, S. (2020). “Student Segregation Across and Within Schools. The Case of the Portuguese Public School System.” *FEUNL Working Paper Series, 633*.

Firmino, J. Nunes LC, Reis AB, & Seabra C (2018). “Class Composition and Student Achievement. Evidence from Portugal.” *FEUNL Working Paper Series, 624*.

Gibbons, S., Telhaj, S., (2008). “Peers and achievement in England’s secondary schools. *SERC Discussion Paper*, 1.

Hanushek, E. A., Kain, J. F., Markman, J. M., & Rivkin, S. G. (2003). “Does peer ability affect student achievement?” *Journal of Applied Econometrics*, 18(5), 527-544.

Imbens, G. (2015). “Matching Methods in Practice: Three Examples.” *Journal of Human Resources*, 50(2), 373–419.

Lavy, V., Silva, O., & Weinhardt, F. (2012). “The Good, the Bad and the Average: Evidence on the Scale and Nature of Ability Peer Effects in Schools.” *Journal of Labor Economics*, 30(2), 367–414.

Manski, C.F., (1993). “Identification of endogenous social effects: The reflection problem.” *Review of Economic Studies*, 60, 531–542.

Moffitt, R.A. 2001. “Policy interventions, low-level equilibria, and social interactions.” In *Social Dynamics*, edited by Durlauf, S. N., and Young, H. P., 45–82. Cambridge, MA: MIT Press.

Sacerdote, B., (2001). “Peer effects with random assignment: Results for Dartmouth roommates.” *The Quarterly Journal of Economics*, 116, 681–704.

Sacerdote, B. (2011). “Peer effects in education: How might they work, how big are they and how much do we know thus far.” In *Handbook of the Economics of Education*, edited by Hanushek, E. A., Machin, S., and Woessmann, L. 3(3), 249-277. Amsterdam: North Holand

Silva, P L., Nunes, L. C., Seabra, C., & Reis, AB. (2020). “Student selection and performance in higher education: admission exams vs. high school scores”, *Education Economics*, 28:5, 437-454

Sund, K., (2009). “Estimating peer effects in Swedish high school using school, teacher, and student fixed effects.” *Economics of Education Review*, 28(3), 329-336.

Vigdor, J.L., & Nechyba, T.S., (2007). “Peer effects in North Carolina public schools.” In *Schools and the Equal Opportunity Problem*, edited by Woessmann, L., and Peterson, P.E., 73–101. Cambridge, MA: The MIT Press.

Appendix.

Figure A.1 Kernel Density of the Differential between Students' 2nd Cycle GPA and Average 6th grade Exam Score.

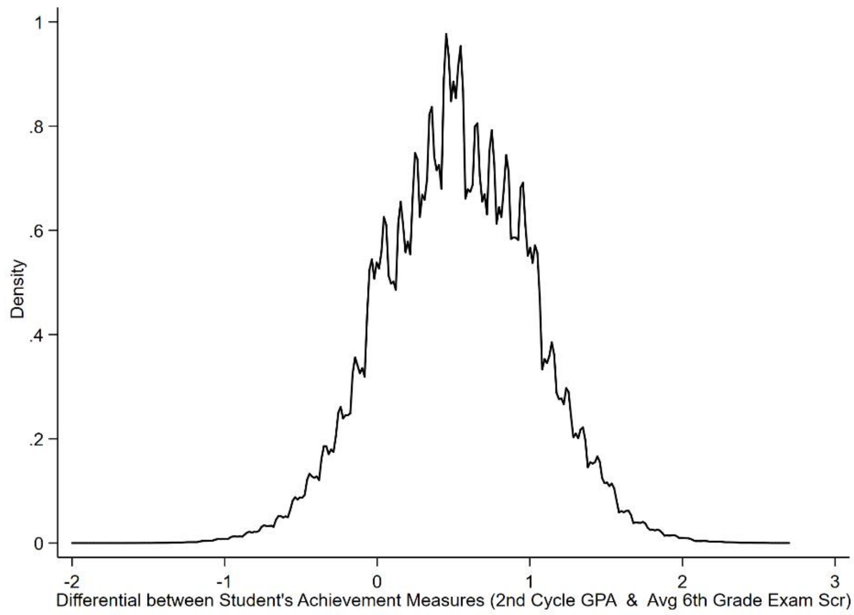


Figure A.2. Kernel Density of the Differential between Peers' 2nd Cycle GPA and Average 6th grade Exam Score.

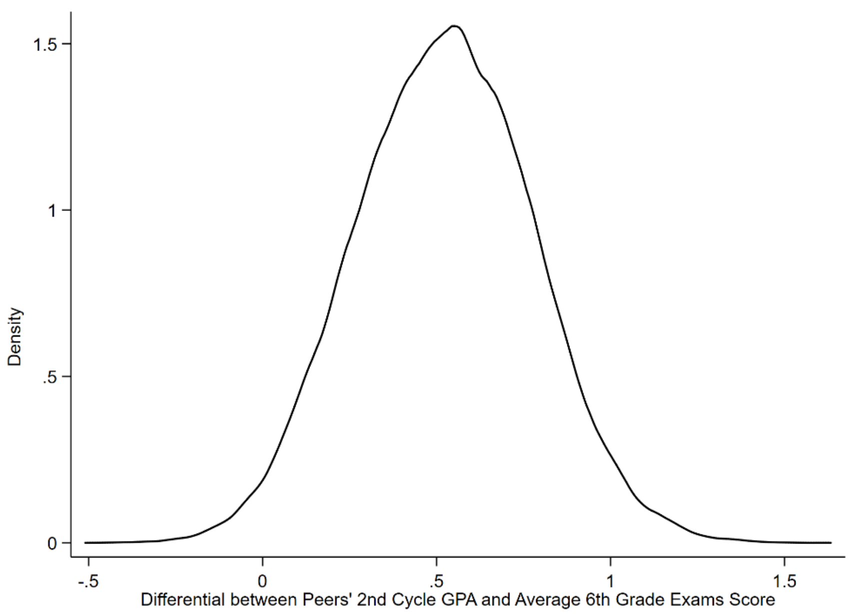


Table A.1. Imbens t-test results for differences in means between Unrestricted Sample and Final Sample.

	Unrestricted Sample		Final Sample		
Variable	mean	s.d.	mean	s.d.	**Imbens**
Average 9th grade Exam Score	53.25	18.51	53.73	18.49	-0.03
Student's 2nd Cycle GPA	3.67	0.67	3.68	0.67	-0.02
Student's Average 6th grade Exam Score	3.15	0.75	3.16	0.75	-0.01
Same 3rd Cycle Peers	0.73	0.44	0.75	0.43	-0.05
Same 3rd Cycle Peers Missing	0.06	0.25	0.06	0.24	0.02
Internet	0.61	0.49	0.61	0.49	-0.01
Computer	0.72	0.45	0.73	0.45	-0.01
Mother Higher Education	0.18	0.39	0.19	0.39	-0.01
Mother Secondary Education	0.22	0.41	0.22	0.41	0.00
Mother Education Missing	0.12	0.33	0.12	0.33	0.00
Father Higher Education	0.12	0.32	0.12	0.32	-0.01
Father Secondary Education	0.18	0.39	0.18	0.39	0.00
Father Education Missing	0.16	0.36	0.15	0.36	0.01
Mother Unemployment	0.10	0.30	0.10	0.30	0.00
Father Unemployment	0.05	0.22	0.05	0.22	0.00
Low Income	0.41	0.49	0.40	0.49	0.01
Female	0.52	0.50	0.52	0.50	0.00
Female Missing	0.00	0.02	0.00	0.02	0.00
Mother FPSC	0.05	0.21	0.05	0.21	0.01
Mother FNPSC	0.03	0.18	0.03	0.18	0.00
Mother Origin Missing	0.00	0.06	0.00	0.06	0.00
Father FPSC	0.05	0.21	0.04	0.21	0.01
Father FNPSC	0.03	0.16	0.03	0.16	0.00
Father Origin Missing	0.01	0.10	0.01	0.10	0.00
Student FPSC	0.01	0.11	0.01	0.11	0.01
Student FNPSC	0.02	0.14	0.02	0.13	0.00
Student Origin Missing	0.00	0.02	0.00	0.02	0.00
2nd Cycle Retention	0.04	0.19	0.04	0.18	0.01

Notes: Sample sizes are 321,344 and 300,638 students for the Unrestricted and Final Sample, respectively. Imbens t-test critical value for statistical significance is 0.3. No variable shows any significant difference.

Table A.2. Imbens t-test results for differences in means between Final Sample and Fixed Effects Specifications Samples for Individual-Level Variables.

	Final Sample		*School-Year FE Sample*			*School-Year-Teacher FE Sample*		
Individual Level Variable	mean	s.d.	mean	s.d.	**Imbens**	mean	s.d.	**Imbens**
Average 9th grade Exam Score	53.73	18.49	53.81	18.50	0.00	54.13	18.45	-0.02
Peers' 2nd Cycle GPA	3.68	0.29	3.68	0.29	0.00	3.69	0.29	0.00
Peers' Average 6th grade Exam Score	3.16	0.38	3.16	0.38	0.00	3.17	0.37	-0.02
Student's 2nd Cycle GPA	3.68	0.67	3.68	0.67	0.00	3.69	0.67	0.00
Student's Average 6th grade Exam Score	3.16	0.75	3.16	0.75	0.00	3.17	0.75	-0.01
Same 3rd Cycle Peers	0.75	0.43	0.76	0.43	-0.01	0.76	0.43	-0.02
Same 3rd Cycle Peers Missing	0.06	0.24	0.06	0.24	0.00	0.06	0.24	0.00
Internet	0.61	0.49	0.62	0.49	0.00	0.62	0.49	-0.01
Computer	0.73	0.45	0.73	0.45	0.00	0.73	0.44	-0.01
Mother Higher Education	0.19	0.39	0.19	0.39	0.00	0.19	0.39	0.00
Mother Secondary Education	0.22	0.41	0.22	0.41	0.00	0.22	0.41	0.01
Mother Education Missing	0.12	0.33	0.12	0.33	0.00	0.12	0.33	0.01
Father Higher Education	0.12	0.32	0.12	0.33	0.00	0.12	0.33	0.00
Father Secondary Education	0.18	0.39	0.19	0.39	0.00	0.18	0.39	0.00
Father Education Missing	0.15	0.36	0.15	0.36	0.00	0.15	0.36	0.01
Mother Unemployment	0.10	0.30	0.10	0.30	0.00	0.10	0.30	0.00
Father Unemployment	0.05	0.22	0.05	0.22	0.00	0.05	0.22	0.00
Low Income	0.40	0.49	0.40	0.49	0.00	0.40	0.49	0.00
Female	0.52	0.50	0.52	0.50	0.00	0.52	0.50	0.00
Female Missing	0.00	0.02	0.00	0.02	0.00	0.00	0.02	0.00
Mother FPSC	0.05	0.21	0.05	0.21	0.00	0.04	0.20	0.01
Mother FNPSC	0.03	0.18	0.03	0.18	0.00	0.03	0.18	0.00
Mother Origin Missing	0.00	0.06	0.00	0.06	0.00	0.00	0.06	0.00
Father FPSC	0.04	0.21	0.04	0.21	0.00	0.04	0.20	0.01
Father FNPSC	0.03	0.16	0.03	0.16	0.00	0.03	0.16	0.00
Father Origin Missing	0.01	0.10	0.01	0.10	0.00	0.01	0.09	0.00
Student FPSC	0.01	0.11	0.01	0.11	0.00	0.01	0.10	0.01
Student FNPSC	0.02	0.13	0.02	0.13	0.00	0.02	0.13	0.00
Student Origin Missing	0.00	0.02	0.00	0.02	0.00	0.00	0.01	0.00
2nd Cycle Retention	0.04	0.18	0.03	0.18	0.00	0.03	0.18	0.01

Notes: All variables are Individual-Level Variables. The sample sizes are 300,638 students for the Final Sample, 291,235 for the School x Year Fixed Effect Sample and 122,497 for the School x Year x Teacher Fixed effect Sample. Imbens t-test critical value for statistical significance is 0.3. No variable shows any significant difference.

Table A.3. Imbens t-test results for differences in means between Final Sample and Fixed Effects Specifications Samples for Class-Level Variables.

	Final Sample		***School-Year FE Sample***			***School-Year-Teacher FE Sample***		
Class Level Variable	mean	s.d.	mean	s.d.	**Imbens**	mean	s.d.	**Imbens**
Class Size	22.52	4.12	22.62	4.07	-0.02	22.61	4.10	-0.02
Class Internet	0.61	0.27	0.61	0.26	-0.01	0.61	0.26	-0.02
Class Computer	0.72	0.25	0.72	0.25	0.00	0.73	0.25	-0.03
Class Mother Higher Education	0.18	0.17	0.18	0.17	-0.01	0.18	0.17	0.00
Class Mother Secondary Education	0.22	0.14	0.22	0.14	0.00	0.21	0.14	0.02
Class Mother Education Missing	0.12	0.15	0.12	0.15	0.00	0.12	0.14	0.02
Class Father Higher Education	0.11	0.13	0.11	0.13	-0.01	0.11	0.13	-0.01
Class Father Secondary Education	0.18	0.13	0.18	0.13	-0.01	0.18	0.13	0.01
Class Father Education Missing	0.16	0.15	0.16	0.15	0.00	0.15	0.15	0.02
Class Mother Unemployment	0.10	0.09	0.10	0.09	0.00	0.10	0.09	-0.01
Class Father Unemployment	0.05	0.06	0.05	0.06	0.00	0.05	0.06	0.01
Class Female	0.52	0.14	0.52	0.14	0.00	0.52	0.14	-0.01
Class Female Missing	0.00	0.01	0.00	0.01	0.01	0.00	0.01	0.00
Class Low Income	0.41	0.21	0.41	0.21	0.01	0.41	0.21	0.00
Class Mother FPSC	0.05	0.08	0.05	0.08	0.00	0.04	0.07	0.04
Class Mother FNPSC	0.03	0.05	0.03	0.05	0.00	0.03	0.05	-0.01
Class Mother Origin Missing	0.00	0.02	0.00	0.02	0.00	0.00	0.02	0.02
Class Father FPSC	0.05	0.08	0.05	0.08	0.00	0.04	0.07	0.04
Class Father FNPSC	0.03	0.05	0.03	0.05	0.00	0.03	0.05	-0.01
Class Father Origin Missing	0.01	0.03	0.01	0.03	0.01	0.01	0.03	0.02
Class Student FPSC	0.01	0.04	0.01	0.03	0.00	0.01	0.03	0.05
Class Student FNPSC	0.02	0.04	0.02	0.04	0.01	0.02	0.04	0.01
Class Student Origin Missing	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.02
Class 2^{nd} Cycle Retention	0.04	0.06	0.04	0.06	0.00	0.03	0.06	0.03

Notes: All variables are Class-Level Variables. The sample sizes are 17,895 classes for the Final Sample, 17,215 classes for the School x Year Fixed Effect Sample and 7,157 classes for the School x Year x Teacher Fixed Effect Sample. Imbens t-test critical value for statistical significance is 0.3. No variable shows any significant difference.

Table A.4. Estimates of Ability Peer Effects on 9th Grade Average National Exam Score

(Including Individual and Class-Level Controls).

	(1)	(2)	(3)	(4)
VARIABLES	OLS	Year Fixed Effects	School-Year Fixed effects	School-Year-Teacher Fixed effects
Peers' 2nd Cycle GPA	-0.292***	-0.308***	-0.065***	-0.066***
	(0.010)	(0.011)	(0.011)	(0.016)
Peers' Average 6th grade Exam Score	0.346***	0.373***	0.129***	0.110***
	(0.009)	(0.009)	(0.010)	(0.015)
(Individual-Level Controls)				
Student's 2nd Cycle GPA	0.717***	0.722***	0.738***	0.734***
	(0.003)	(0.003)	(0.003)	(0.004)
Student's Average 6th grade Exam Score	0.405***	0.402***	0.386***	0.388***
	(0.002)	(0.002)	(0.003)	(0.004)
Same 3rd Cycle Peers	0.106***	0.106***	0.111***	0.118***
	(0.004)	(0.004)	(0.003)	(0.005)
Same 3rd Cycle Peers Missing	-0.067***	-0.084***	-0.091***	-0.100***
	(0.005)	(0.006)	(0.005)	(0.008)
Internet	0.012***	0.013***	0.013***	0.010*
	(0.003)	(0.003)	(0.004)	(0.005)
Computer	0.009**	0.009**	0.010**	0.013**
	(0.004)	(0.004)	(0.004)	(0.006)
Mother Higher Education	0.070***	0.070***	0.067***	0.070***
	(0.004)	(0.004)	(0.004)	(0.006)
Mother Secondary Education	0.013***	0.013***	0.012***	0.013***
	(0.003)	(0.003)	(0.003)	(0.005)
Mother Education Missing	0.030***	0.030***	0.030***	0.034***
	(0.005)	(0.005)	(0.005)	(0.008)

Table A.4. (*Continued*)

Father Higher Education	0.097***	0.096***	0.098***	0.093***
	(0.004)	(0.004)	(0.004)	(0.007)
Father Secondary Education	0.034***	0.034***	0.034***	0.031***
	(0.003)	(0.003)	(0.003)	(0.005)
Father Education Missing	-0.005	-0.005	-0.005	-0.016**
	(0.005)	(0.005)	(0.005)	(0.008)
Mother Unemployment	-0.029***	-0.029***	-0.028***	-0.028***
	(0.004)	(0.004)	(0.004)	(0.006)
Father Unemployment	-0.027***	-0.026***	-0.026***	-0.027***
	(0.005)	(0.005)	(0.005)	(0.007)
Female	0.112***	0.112***	0.113***	0.111***
	(0.002)	(0.002)	(0.002)	(0.004)
Female Missing	0.091*	0.090*	0.099*	0.186***
	(0.052)	(0.052)	(0.055)	(0.068)
Low Income	-0.036***	-0.037***	-0.037***	-0.035***
	(0.002)	(0.002)	(0.002)	(0.004)
Mother FPSC	-0.007	-0.007	-0.007	-0.004
	(0.006)	(0.006)	(0.006)	(0.009)
Mother FNPSC	0.004	0.004	0.003	0.007
	(0.006)	(0.006)	(0.006)	(0.010)
Mother Origin Missing	-0.032*	-0.031*	-0.030*	-0.023
	(0.018)	(0.018)	(0.018)	(0.029)
Father FPSC	0.005	0.005	0.004	0.015*
	(0.006)	(0.006)	(0.006)	(0.009)
Father FNPSC	0.002	0.001	0.003	-0.002
	(0.007)	(0.007)	(0.007)	(0.011)
Father Origin Missing	-0.008	-0.008	-0.007	0.012
	(0.011)	(0.011)	(0.012)	(0.018)
Student FPSC	-0.023**	-0.024**	-0.024**	-0.064***
	(0.012)	(0.011)	(0.011)	(0.018)
Student FNPSC	-0.023***	-0.023***	-0.026***	-0.035***
	(0.008)	(0.008)	(0.008)	(0.013)

Table A.4. (*Continued*)				
Student Origin Missing	0.008	0.008	0.039	0.000
	(0.061)	(0.062)	(0.062)	(0.109)
2nd Cycle Retention	-0.151***	-0.146***	-0.145***	-0.142***
	(0.005)	(0.005)	(0.005)	(0.009)
(Class-Level Controls)				
Class Size	-0.005***	-0.005***	-0.001**	-0.001
	(0.000)	(0.000)	(0.000)	(0.001)
Class Internet	-0.020	-0.024	0.014	-0.030
	(0.018)	(0.018)	(0.018)	(0.025)
Class Computer	0.025	0.028	-0.010	-0.005
	(0.018)	(0.018)	(0.019)	(0.027)
Class Mother Higher Education	0.177***	0.172***	0.176***	0.141***
	(0.024)	(0.024)	(0.021)	(0.030)
Class Mother Secondary Education	-0.044**	-0.043**	0.044***	0.031
	(0.019)	(0.019)	(0.016)	(0.023)
Class Mother Education Missing	0.089***	0.085***	0.049*	0.120***
	(0.033)	(0.033)	(0.027)	(0.041)
Class Father Higher Education	0.018	0.011	0.046**	0.063*
	(0.026)	(0.026)	(0.023)	(0.033)
Class Father Secondary Education	-0.031	-0.029	0.045***	0.078***
	(0.021)	(0.020)	(0.017)	(0.025)
Class Father Education Missing	-0.122***	-0.123***	0.039	-0.016
	(0.031)	(0.030)	(0.026)	(0.039)
Class Mother Unemployment	-0.027	-0.029	0.008	-0.013
	(0.023)	(0.023)	(0.020)	(0.028)
Class Father Unemployment	-0.066**	-0.068**	0.060**	0.142***
	(0.032)	(0.032)	(0.026)	(0.038)
Class Female	0.057***	0.058***	0.043***	0.032**
	(0.013)	(0.013)	(0.011)	(0.015)
Class Female Missing	-0.429	-0.414	0.122	0.272
	(0.346)	(0.348)	(0.287)	(0.373)

Table A.4. (*Continued*)				
Class Low Income	-0.075***	-0.070***	-0.052***	-0.050***
	(0.014)	(0.014)	(0.013)	(0.018)
Class Mother FPSC	-0.002	0.004	0.069**	-0.006
	(0.037)	(0.036)	(0.032)	(0.046)
Class Mother FNPSC	0.143***	0.139***	-0.021	-0.092*
	(0.040)	(0.039)	(0.034)	(0.049)
Class Mother Origin Missing	-0.252**	-0.249**	-0.044	-0.088
	(0.117)	(0.117)	(0.100)	(0.133)
Class Father FPSC	-0.168***	-0.162***	-0.049	0.040
	(0.036)	(0.036)	(0.031)	(0.045)
Class Father FNPSC	0.106**	0.110**	-0.043	-0.068
	(0.045)	(0.045)	(0.038)	(0.053)
Class Father Origin Missing	-0.103	-0.099	0.019	0.291***
	(0.073)	(0.073)	(0.060)	(0.085)
Class Student FPSC	-0.121*	-0.111	-0.091	0.057
	(0.069)	(0.069)	(0.059)	(0.086)
Class Student FNPSC	-0.044	-0.037	0.063	0.073
	(0.050)	(0.050)	(0.042)	(0.059)
Class Student Origin Missing	-0.519	-0.533	0.068	-0.009
	(0.348)	(0.340)	(0.320)	(0.436)
Class 2nd Cycle Retention	-0.183***	-0.178***	-0.090***	0.005
	(0.034)	(0.034)	(0.030)	(0.043)
Dummies for 6th grade Exam Year	YES	YES	YES	YES
Year FE		YES		
School x Year FE			YES	
School x Year x Teacher FE				YES
Observations	300,638	300,638	291,235	122,497
R-squared	0.661	0.678	0.711	0.722

Notes: Significance levels: *** p<0.01, ** p<0.05, * p<0.1. Robust standard errors clustered at the class level in parentheses. The outcome variable (Average 9th grade Exam Score) was standardized to have mean zero and standard deviation of one within each school year. All specifications have an intercept and use controls at the individual and class-level and dummies accounting for missing information for each individual and class-level control.

Table A.5. Estimates of Ability Peer Effects on 9th Grade Average National Exam Score using students without 3rd Cycle Retention.

	(1)	(2)	(3)
VARIABLES	OLS	School-Year Fixed effects	School-Year-Teacher Fixed effects
Peers' 2nd Cycle GPA	-0.322***	-0.074***	-0.079***
	(0.011)	(0.012)	(0.017)
Peers' 6th Grade National Exams	0.383***	0.137***	0.115***
	(0.009)	(0.011)	(0.016)
Student's 2nd Cycle GPA	0.727***	0.744***	0.740***
	(0.003)	(0.003)	(0.005)
Student's 6th Grade National Exams	0.404***	0.388***	0.390***
	(0.003)	(0.003)	(0.004)
Additional Controls	YES	YES	YES
School x Year FE		YES	
School x Year x Teacher FE			YES
Observations	272,491	263,951	111,262
R-squared	0.664	0.699	0.711

Notes: Significance levels: *** p<0.01, ** p<0.05, * p<0.1. Robust standard errors clustered at the class level in parentheses. The outcome variable (Average 9th grade Exam Score) was standardized to have mean zero and standard deviation of one within each school year. All specifications have an intercept and use controls at the individual and class-level and dummies accounting for missing information for each individual and class-level control. (2) uses only schools in each school year with at least two 9th grade classes. (3) uses only schools in each school year where each Portuguese and Math teacher teaches at least two 9th grade classes.

Table A.6. Estimates of Ability Peer Effects on 9th Grade Average National Exam Score for the Teacher Pair Restricted Sample.

	(1)
VARIABLES	School-Year-Teacher Fixed effects
Peers' 2nd Cycle GPA	-0.0592***
	(0.018)
Peers' 6th Grade National Exams	0.104***
	(0.016)
Student's 2nd Cycle GPA	0.736***
	(0.005)
Student's 6th Grade National Exams	0.385***
	(0.004)
Additional Controls	YES
School x Year x Teacher FE	YES
Observations	102,240
R-squared	0.722

Notes: Significance levels: *** p<0.01, ** p<0.05, * p<0.1. Robust standard errors clustered at the class level in parentheses. The outcome variable (Average 9th grade Exam Score) was standardized to have mean zero and standard deviation of one within each school year. All specifications have an intercept and use controls at the individual and class-level and dummies accounting for missing information for each individual and class-level control. This specification uses a restricted sample of classes taught by the same pair of Portuguese and Math teachers, within each school and school year.

Table A.7. Imbens t-test results for differences in means between Final Sample and School-by-Year-Teacher (Robust Teacher Pair) Sample for Individual-Level Variables.

	Final Sample		***School-Year-Teacher FE Sample***		
Individual Level Variable	mean	s.d.	mean	s.d.	**Imbens**
Average 9th grade Exam Score	53.73	18.49	53.87	18.42	-0.01
Peers' 2nd Cycle GPA	3.68	0.29	3.68	0.29	0.02
Peers' Average 6th grade Exam Score	3.16	0.38	3.16	0.37	0.00
Student's 2nd Cycle GPA	3.68	0.67	3.68	0.67	0.01
Student's Average 6th grade Exam Score	3.16	0.75	3.16	0.75	0.00
Same 3rd Cycle Peers	0.75	0.43	0.75	0.43	0.01
Same 3rd Cycle Peers Missing	0.06	0.24	0.06	0.24	0.00
Internet	0.61	0.49	0.62	0.49	-0.01
Computer	0.73	0.45	0.73	0.44	-0.01
Mother Higher Education	0.19	0.39	0.18	0.38	0.01
Mother Secondary Education	0.22	0.41	0.21	0.41	0.01
Mother Education Missing	0.12	0.33	0.12	0.32	0.01
Father Higher Education	0.12	0.32	0.12	0.32	0.01
Father Secondary Education	0.18	0.39	0.18	0.38	0.01
Father Education Missing	0.15	0.36	0.15	0.36	0.01
Mother Unemployment	0.10	0.30	0.10	0.30	-0.01
Father Unemployment	0.05	0.22	0.05	0.22	0.00
Low Income	0.40	0.49	0.41	0.49	-0.01
Female	0.52	0.50	0.72	0.45	-0.01
Female Missing	0.00	0.02	0.78	0.41	-0.02
Mother FPSC	0.05	0.21	0.19	0.39	0.01
Mother FNPSC	0.03	0.18	0.23	0.42	0.01
Mother Origin Missing	0.00	0.06	0.10	0.30	0.01
Father FPSC	0.04	0.21	0.12	0.33	0.01
Father FNPSC	0.03	0.16	0.19	0.39	0.01
Father Origin Missing	0.01	0.10	0.13	0.34	0.01
Student FPSC	0.01	0.11	0.11	0.31	-0.01
Student FNPSC	0.02	0.13	0.07	0.25	-0.01
Student Origin Missing	0.00	0.02	0.52	0.50	0.00
2nd Cycle Retention	0.04	0.18	0.00	0.02	0.00

Notes: All variables are Individual-Level Variables. The sample sizes are 300,638 students for the Final Sample and 102,2470 for the School x Year x Teacher Fixed Effect Sample. Imbens t-test critical value for statistical significance is 0.3. No variable shows any significant difference.

Table A.8. Imbens t-test results for differences in means between Final Sample and School-by-Year-Teacher (Robust Teacher Pair) Sample for Class-Level Variables.

	Final Sample		***School-Year-Teacher FE Sample***		
Class Level Variable	mean	s.d.	mean	s.d.	**Imbens**
Class Size	22.52	4.12	22.46	4.13	0.02
Class Internet	0.61	0.27	0.61	0.26	-0.01
Class Computer	0.72	0.25	0.73	0.25	-0.02
Class Mother Higher Education	0.18	0.17	0.17	0.16	0.03
Class Mother Secondary Education	0.22	0.14	0.21	0.14	0.02
Class Mother Education Missing	0.12	0.15	0.12	0.14	0.03
Class Father Higher Education	0.11	0.13	0.11	0.13	0.03
Class Father Secondary Education	0.18	0.13	0.18	0.13	0.03
Class Father Education Missing	0.16	0.15	0.15	0.15	0.03
Class Mother Unemployment	0.10	0.09	0.10	0.09	-0.03
Class Father Unemployment	0.05	0.06	0.05	0.07	-0.01
Class Female	0.52	0.14	0.52	0.14	-0.01
Class Female Missing	0.00	0.01	0.00	0.01	0.00
Class Low Income	0.41	0.21	0.42	0.21	-0.03
Class Mother FPSC	0.05	0.08	0.04	0.07	0.04
Class Mother FNPSC	0.03	0.05	0.03	0.05	-0.02
Class Mother Origin Missing	0.00	0.02	0.00	0.02	0.02
Class Father FPSC	0.05	0.08	0.04	0.07	0.04
Class Father FNPSC	0.03	0.05	0.03	0.05	-0.01
Class Father Origin Missing	0.01	0.03	0.01	0.03	0.02
Class Student FPSC	0.01	0.04	0.01	0.03	0.04
Class Student FNPSC	0.02	0.04	0.02	0.04	0.00
Class Student Origin Missing	0.00	0.00	0.00	0.00	0.02
Class 2^{nd} Cycle Retention	0.04	0.06	0.04	0.06	0.03

Notes: All variables are Class-Level Variables. The sample sizes are 17,895 classes for the Final Sample and 6,031 classes for the School x Year x Teacher Fixed Effect Sample. Imbens t-test critical value for statistical significance is 0.3. No variable shows any significant difference.

Table A.9. Estimates of Ability Peer Effects (2nd Cycle GPA) on 9th Grade Average National Exam Score.

	(1)	(2)	(3)	(4)
VARIABLES	OLS	Year Fixed Effects	School-Year Fixed effects	School-Year-Teacher Fixed effects
Peers' 2nd Cycle GPA	-0.091***	-0.093***	0.037***	0.018
	(0.010)	(0.010)	(0.008)	(0.012)
Additional Controls	YES	YES	YES	YES
Year FE		YES		
School x Year FE			YES	
School x Year x Teacher FE				YES
Observations	300,638	300,638	291,235	122,497
R-squared	0.631	0.632	0.684	0.696

Notes: Significance levels: *** p<0.01, ** p<0.05, * p<0.1. Robust standard errors clustered at the class level in parentheses. The outcome variable (Average 9th grade Exam Score) was standardized to have mean zero and standard deviation of one within each school year. All specifications have an intercept and use controls at the individual and class-level and dummies accounting for missing information for each individual and class-level control.

Table A.10. Estimates of Ability Peer Effects (Average 6th grade Exams score) on 9th Grade Average National Exam Score.

	(1)	(2)	(3)	(4)
VARIABLES	OLS	Year Fixed Effects	School-Year Fixed effects	School-Year-Teacher Fixed effects
Peers' 6th Grade National Exams	0.099***	0.095***	0.083***	0.055***
	(0.007)	(0.008)	(0.008)	(0.012)
Additional Controls	YES	YES	YES	YES
Year FE		YES		
School x Year FE			YES	
School x Year x Teacher FE				YES
Observations	300,638	300,638	291,235	122,497
R-squared	0.600	0.601	0.635	0.648

Notes: Significance levels: *** p<0.01, ** p<0.05, * p<0.1. Robust standard errors clustered at the class level in parentheses. The outcome variable (Average 9th grade Exam Score) was standardized to have mean zero and standard deviation of one within each school year. All specifications have an intercept and use controls at the individual and class-level and dummies accounting for missing information for each individual and class-level control.