RECent
Center for Economic Research

WORKING PAPER SERIES

Choose the school, choose the performance. New evidence on the determinants of student performance in eight European countries

Luca Bonacini, Irene Brunetti, Giovanni Gallo

Working Paper 150

August 2021

www.recent.unimore.it

RECent: c/o Dipartimento di Economia Politica, Viale Berengario 51, I-41100 Modena, ITALY
Phone +39 059 2056856, Fax +39 059 2056947, E-mail: RECent@unimore.it

Choose the school, choose the performance. New evidence on the determinants of student performance in eight European countries

Luca Bonacini [a, b], Irene Brunetti [c], Giovanni Gallo [d, a, b*]

[a] *University of Modena and Reggio Emilia, Italy*

[b] *Global Labor Organization (GLO), Essen, Germany*

[c] *National Institute for Public Policies Analysis (INAPP), Italy*

[d] *Sapienza University of Rome, Italy*

August 2021

Abstract

This study aims to identify the main determinants of student performance in reading and maths across eight European Union countries (Austria, Croatia, Germany, Hungary, Italy, Portugal, Slovakia, and Slovenia). Based on student-level data from the OECD's PISA 2018 survey and by means of the application of efficient algorithms, we highlight that the number of books at home and a variable combining the type and location of their school represent the most important predictors of student performance in all of the analysed countries, while other school characteristics are rarely relevant. Econometric results show that students attending vocational schools perform significantly worse than those in general schools, except in Portugal. Considering only general school students, the differences between big and small cities are not statistically significant, while among students in vocational schools, those in a small city tend to perform better than those in a big city. Through the Gelbach decomposition method, which allows measuring the relative importance of observable characteristics in explaining a gap, we show that the differences in test scores between big and small cities depend on school characteristics, while the differences between general and vocational schools are mainly explained by family social status.

Keywords: Gelbach decomposition, Education inequalities, Machine learning, PISA, Schooling tracking, Student performance.

JEL Classification: I21, I24, J24.

---

[*] Corresponding author. Sapienza University of Rome, Italy. E-mail: luca.bonacini@unimore.it (L. Bonacini), i.brunetti@inapp.org (I. Brunetti), giovanni.gallo@uniroma1.it (G. Gallo). The views expressed in this paper are those of the authors and do not necessarily reflect those of INAPP.

1. Introduction

Student academic performance is one of the key features in education, determining the success or the failure of an academic institution, and has a direct impact on the socioeconomic development of a country (Farooq et al. 2011). Low academic performance may result in student disengagement and consequent dropping out from the education system. Young people without an upper-secondary qualification tend to face severe difficulties entering and remaining in the labour market, and this has economic and social consequences for individuals and society alike. Improving educational outcomes is a prominent method of enhancing micro- and macroeconomic outcomes for many OECD countries, as well of as addressing inequalities (OECD 2019b).

Researchers have long been interested in exploring variables affecting the quality of performance of students. These variables are both internal and external to the school and can be categorised as student factors, family factors, school factors, and peer factors (Crosnoe et al. 2004). Student factors may include immigration status and age/grade, but also intelligence, attitude, motivation, interests, and study habits (Nisar et al. 2017). Family factors relate to socioeconomic status and include parents' qualifications and occupation, family size, income, social standing in society, and the home environment. Moreover, these factors affect student performance through the community context. If a school is located in a city, students may enjoy additional resources nearby (such as public libraries and museums) that support learning and may be less accessible to students attending a rural school. Finally, school factors include ownership, resources, the student–teacher ratio or the size of the class, and the type of school, while peer factors refer to interactions between individuals of a similar age that are fairly close friends and share the same activities.

This study aims to identify the main determinants of the reading and maths performances of students living in the European Union. To do this, we use data from the OECD's 2018 Programme for International Student Assessment (PISA) survey, which reports internationally comparable information on students, their households, and the characteristics of the schools they attend. Specifically, we focus on Austria, Croatia, Germany, Hungary, Italy, Portugal, Slovakia, and Slovenia, as these countries begin school tracking when students are 15 years old or earlier.

The novelty of this paper in the context of the existing literature on the topic is twofold. First, we identify the main determinants of student performance for each country through the application of the efficient computational procedures proposed by Furnival and Wilson (1974). Second, in a comparative perspective across the eight European Union countries, we deeply explore the role played by two specific determinants: the type of school (general vs vocational) and its location (big vs small city). In particular, we implement the innovative decomposition analysis proposed by Gelbach (2016) to assess the underlying mechanisms behind the average gaps in cognitive test scores among students attending schools that are different in terms of type or location.

Our main findings highlight that the number of books at home and a variable combining the type and location of the school represent the most important predictors of student performance in all analysed countries, while other school characteristics are rarely relevant. Econometric results show that students attending vocational schools perform significantly worse than those in general schools (except in Portugal) and that those attending a vocational school in a big city tend to perform worse than those in a small city. Furthermore, the decomposition analysis shows that the variables explaining the gap between types of schools are overall similar across countries: we find that family socioeconomic characteristics are the main underlying reason for the differences between general and vocational schools, while the quality of the school represents the main explanation for differences between big and small cities.

The remainder of the paper is structured as follows. Section 2 reviews the main empirical literature related to the determinants of student performance, and Section 3 introduces the data used in the analysis and some descriptive statistics. Section 4 presents the algorithms performed to identify the best model specifications. Section 5 shows and discusses the empirical results, and Section 6 offers some concluding remarks.

2. Literature review

The quality of education received plays a key role in students' opportunities and affects the economic growth of countries through its impact on productivity and social development (Hanushek and Woessmann 2007; Raitano and Vona 2013). The importance of education for social and economic policies highlights the need to monitor student performance (Giambona and Porcu 2015). The literature is very rich and explores different channels to develop policies with the aim of improving educational quality. For the estimation of the education production function, the following relevant factors are generally considered: student background; the ways in which teaching is organized and delivered in classes; the human and financial resources available to schools; and institutional structures of the educational system. Following Coleman (1966) and Hanushek (2008), education can be viewed as a process in which student output derives from inputs, and since outcomes cannot be changed policy attention has focused on inputs that can in part be controlled by policymakers (type of school, teachers, etc.) (Masci et al. 2018).

Using the PISA database, Fuchs and Woessmann (2007) show a strong relationship between family background and student achievement in standardized reading, mathematics, and science exams across a range of countries. These findings are confirmed by other analyses conducted both at the country level (see, for example, Lee and Barro 2001) and at the student level (Woessmann 2003). Factors such as socioeconomic status, immigration status, and age/grade are strongly related to student performance (Karakolidis et al. 2016; Pholphirul 2016). Gender is a special case, since its influence can favour male or female students depending on the competence under study (generally, boys outperform girls in maths and science and the opposite is true for reading), and with varying degrees of intensity (Gamazo et al. 2018). Regarding immigration background, Giannelli and Rapallini (2016) show that immigrant students have a significantly lower score gap in math and Tonello (2016) finds a weak negative impact of the share of immigrant students on the language scores of native peers. One factor that explains the lower performance of immigrant students with respect to natives is a less favourable family background (Ammermueller 2007; Schneeweis 2011). Concerning the age of students, if students were enrolled a year before the usual age they achieve better results than other students whereas, as discussed by Agasisti and Vittadini (2012), if students were held back a year during their past career their achievements worsen. The role of ICT availability, recently studied by the literature, is not clear (Banerjee et al. 2007), but some articles point out that there is difference between having ICT access at home or at school (Escueta et al. 2020; Murat and Bonacini 2020).

At the school level, factors like ownership, the student–teacher ratio, or size have produced diverse results. There are studies that find positive relationships and others no significant relationship or contradictory results (Kim and Law 2012; Acosta and Hsu 2014). For class size, there is no consensus on what the best ratio of students to teachers is, although the 'conventional wisdom' is that students need more time and interaction with teachers for quality education. Some studies find a small positive impact of reduced class size on long-term outcomes such as overall educational attainment (Bingley et al. 2005; Browning and Heinesen 2007). In contrast, Denny and Oppedisano (2011) analyse PISA data for the UK and USA and suggest that bigger classes lead to better results. The quality of teachers is recognized as a factor that leads to improved student achievement (Coe et al. 2014). Better teachers can mean better student lifetime outcomes such as a higher probability of attending higher education and higher initial earnings (Chetty et al. 2014). According to other research, three main institutional features influence the performance of students: how much autonomy individual schools have; how they are held accountable; and whether and how much competition there is between the publicly and privately operated schools. School autonomy in terms of process and personnel decisions is associated with better educational outcomes (Smidova 2019). Because of a principal-agent problem, according to which local decisionmakers can act opportunistically, autonomy works in an environment where schools are accountable for student achievement (Hanushek and Woessmann 2011). The mix of public and private funding seems important. The highest tests scores for the 2003 PISA math test were found in countries with both a high share of privately operated schools and high average share of government funding (Woessmann et al. 2013). Woessman (2016), focusing on the maths scores in the PISA from 2003, finds that resource inputs such as expenditure per student appear to have limited effects on student performance. The opposite result arises when taking into consideration the number of books at home as a proxy for the educational, social, and economic background of the students' families.

An additional determinant of student performance is the type of school attended (general vs vocational). Creating a homogeneous group of students might help teachers be more effective, and some students might benefit from more practical vocational training that prepares them for the labour market. However, other students in these tracks might lose more than they gain—from lower expectations from their teachers to more disengaged classmates. An OECD (2016) PISA report shows that the share of low performers is twice as large among students enrolled in a vocational track than among students enrolled in a general track. On average across OECD countries, 41% of students pursuing a vocational education were low performers in mathematics in 2012, whereas 21% of students in a general track were. Brunello and Rocco (2015) suggest that students with vocational education do not perform as well as those with a general education both in terms of labour market outcomes and the level of basic skills, including literacy and numeracy. Looking at the Italian context, Bratti et al. (2007) confirm these findings, highlighting that academic and technical schools perform better than vocational ones.

Finally, the literature offers evidence on the effects of the location of school. Rural areas have long been associated with inferior opportunities, one of which is the provision of education (Arnold et al. 2005). Evidence suggests that rural schools and students are educationally at a disadvantage in comparison to their urban counterparts in terms of academic achievement. On average across OECD countries, students who attend schools in big cities perform better in the PISA than students who attend schools in villages, rural areas, or small towns (OECD 2010). The size of the gap differs across countries, reflecting differences in the resources and learning opportunities available in rural and urban areas as well as differences in population density or the distribution of labour markets. For example, in Italy, the Slovak Republic, and Romania, the performance gap between students in city schools and those in rural schools is more than 45 score points after accounting for student socioeconomic background. This gap is double in Hungary and Bulgaria (OECD 2010). PISA data from 2015 (OECD 2019a) confirm these results but also shows that in Belgium, the United Kingdom, and the United States, students in rural schools outperform those in city schools. Despite the challenges facing rural schools, some studies argue that small and rural schools can be particularly beneficial to socioeconomically disadvantaged students (Bauch 2001; Semke and Sheridan 2012). Nonetheless, PISA 2015 data reveal that on average across OECD countries, the share of resilient students is somewhat higher in cities than in rural schools.

To the best of our knowledge, none of the studies mentioned in this section combine different factors, however. This paper tries to fill this gap by jointly analysing the effect of the type and location of schools on student performance.

3. Data and descriptive statistics

The analysis relies on data from the PISA survey for the year 2018. PISA is a standardized international project implemented every three years by the Organization for Economic Cooperation and Development (OECD) that measures the cognitive abilities in reading and mathematics and the science skills of all students between the age of 15 years and 3 months and 16 years and 2 months. Each PISA wave concentrates on and provides detailed information regarding one particular subject. PISA 2018 focuses on reading. In each test, the results are scaled to fit approximately normal distributions, with means for OECD countries of around 500 score points and standard deviations of around 100 score points.[1] Other than data on the skills and knowledge of students, PISA collects interesting information on students, their families, and on school characteristics. The sample follows a stratified two-stage sample design: in the first stage, schools with 15-year-old students are randomly selected in each country. In the second stage, students are randomly selected with equal probability within schools. In our analysis, we focus on 8 European countries—Austria, Croatia, Germany, Hungary, Italy,

[1] For more details, see: https://www.oecd.org/pisa/.

Portugal, Slovakia, and Slovenia—that begin school tracking when students are 15 years old or earlier.[2] As is common in empirical studies using PISA survey data, all descriptive statistics and estimates take into account individual sample weights provided.

The dependent variables of our analysis are student scores in mathematics and reading, and our main variable of interest combines the type of institution attended (i.e. general or vocational) and the municipality size in which the school is located. If a school is located in a municipality with more than 15,000 inhabitants, then we consider it to be in a big city, otherwise it is in a small city. Therefore, we create a categorical variable that distinguishes four types of schools: i) general – big city; ii) general – small city; iii) vocational – big city; iv) vocational – small city.

Overall, once observations with missing values for the variables used in this analysis are dropped, our sample counts 46,459 students enrolled in over 2,112 schools in the eight countries. Table 1 indicates the number of students observed and the average scores in reading and mathematics for each country considered. Italy represents the largest subsample, as it counts more than 10,600 students. Other than this case, the other countries range between 3,761 students in the case of Germany and 6,311 in the case of Croatia. Table 1 also points out that students in Germany, Austria, and Slovenia present the highest mean scores in our sample (which are also higher than the average of OECD countries) both in mathematics and in reading. The mean scores in Portugal are almost equal to 500 and are slightly higher than those for Hungary and Italy. Finally, the lowest averages are from Croatia and Slovakia. We also note that countries' ranking in reading and the one in mathematics are not equal to each other.

Table 1. Number of sample observations and student performance scores by country

Country	Abbreviation	Observations	Reading score	Mathematics score
Germany	DE	3,761	510.5	510.8
Croatia	HR	6,311	481.7	466.6
Italy	IT	10,631	482.1	491.5
Hungary	HU	4,539	488.8	494.5
Austria	AT	5,519	505.2	518.0
Portugal	PT	5,316	495.9	496.4
Slovenia	SI	5,395	505.0	518.4
Slovakia	SK	4,987	464.1	491.3
Total		46,459	495.3	500.4

Source: Elaborations of the authors on PISA 2018 data.
Note: Student scores in both reading and mathematics are calculated as the average of the ten plausible values.

3.1. Some descriptive statistics

Table 2 shows the distribution of students, distinguishing for each country between the categories of the type of school explained above: general school in a big city (G-BC), general school in a small city (G-SC), vocational school in a big city (V-BC), and vocational school in a small city (V-SC). Across countries, we find heterogeneity in the distribution of students: more than two fifths of children are enrolled in a general school in a big city in Slovakia, Germany, and Italy, while the percentage is about 20–27 percent in Austria, Portugal, and Croatia. Students attending a general school in a small city constitute about a tenth of the sample in each country, with the exception of Hungary where they account for 3 percent. As we expected, the share of vocational students in big cities is always greater than in small cities, with the exception of Austria (respectively, 32 and 35 percent). The differences are lower in Germany (respectively, 21 and 24 percent), and

[2] Starting with the sample containing all European Union countries, we excluded countries where tracking begins after 15 years of age (source: https://eacea.ec.europa.eu/national-policies/eurydice/national-description_en) and those in which the maths or reading performance of students still at the lower-secondary level is statistically different to that of students in upper-secondary school.

Slovakia (respectively, 20 and 26 percent). In the other countries, about half of the students attend a vocational school in a big city.

Table 2. Share of students by type of school by country

Country	General - Big City (%) (G-BC)	General- Small City (%) (G-SC)	Vocational - Big City (%) (V-BC)	Vocational - Small City (%) (V-SC)
Germany	44.98	9.91	24.17	20.94
Croatia	26.76	6.28	52.67	14.29
Italy	42.61	9.48	40.32	7.59
Hungary	36.53	3.05	49.75	10.68
Austria	20.00	12.32	32.57	35.11
Portugal	23.74	13.99	48.05	14.22
Slovenia	30.99	9.64	45.55	13.82
Slovakia	45.02	8.50	20.72	25.77
Total	40.31	9.61	34.59	15.49

Source: Elaborations of the authors on PISA 2018 data.

4. Selection of good-fitting models

One aim of this paper is to identify, for each country analysed, the best set of predictors of student performance in reading and maths. In this situation—even considering the usual information limitations of a survey dataset—the number of potential predictors is large, but the inclusion of all of them in the model specification may affect the estimate efficiency. A solution generally adopted to avoid this issue consists of limiting the model specification to the relevant covariates only, but this procedure has an important issue as well: it may suffer from arbitrariness.

An alternative approach increasingly common in applied economic studies (Chernozhukov et al. 2020; Bloise and Tancioni 2021; Bonacini et al 2021; Denisova-Schmidt 2021) is the use of algorithms and machine learning techniques to select the best set of regression variables. Among the number of existing methodologies, in this analysis we adopt the efficient computational procedures proposed by Furnival and Wilson (1974). This methodology relies on a branch-and-bound algorithm and uses sequences of sweep operations to carry out the computations (Lawless and Singhal 1978). The aim of these efficient algorithms is to determine the best-fitting models for each possible number of covariates k using tests on residual sums of squares to compare the 2^k possible submodels. Once all of the best-fitting models for each quantity of predictors are selected, the best model specification for a specific country is identified among them as the one reporting the lowest Bayesian information criterion (BIC) value.[3] All estimates are based on linear regression models where standard errors are clustered by the school identification number.

We develop these procedures considering two different dependent variables: the logarithmic transformation of the students' reading and maths scores. The basket of predictors from which we select the best models is composed of the following nine variables: gender (i.e. female or male), language spoken at home (i.e. local or foreign), age (both years and months), highest parental occupational level (i.e. high, average, and low occupational level on the basis of the ISEI classification),[4] number of books at home (i.e. 11 books or fewer, 11–25 books, 26–100 books, 101–200 books, 201–500 books, 500 or more books),[5] a variable combining type

[3] To perform the best variable subset selection, we used the `gvselect` Stata command created by Charles Lindsey and Simon Sheather.

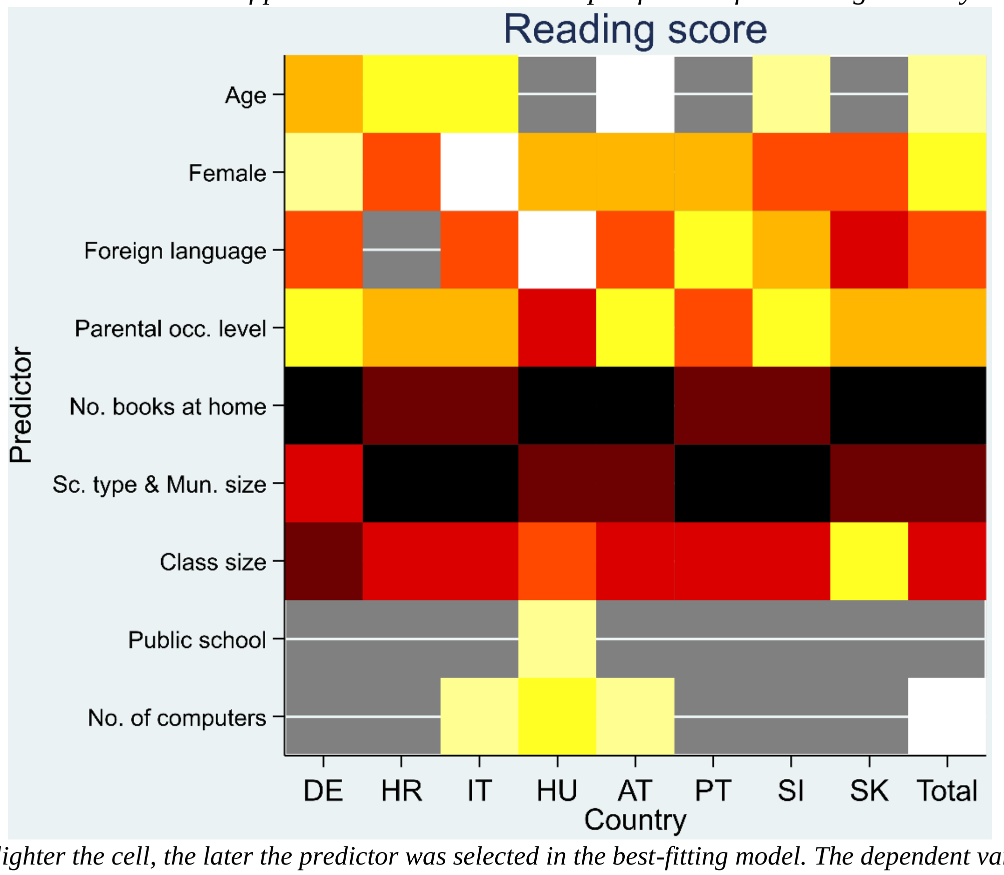
[4] The highest parental occupational level is measured by the HISEI variable. For a deeper explanation of its construction, see: https://www.oecd-ilibrary.org/sites/0a428b07-en/index.html?itemId=/content/component/0a428b07-en.

[5] The PISA survey dataset also provides a composite index that provides synthetic information on all items owned by a household. Among others, the number of books at home represents one of the items composing this index, thus the

of school and municipality size (see Section 3), class size (i.e. 15 students or fewer, 16–20 students, 21–25 students, 26 or more students), a public school dummy, the number of computers per student.[6] The main descriptive statistics for these variables for the total sample of students are provided in Table A.1.

Figure 1 shows the outcome of Furnival and Wilson's (1974) algorithms with respect to student performance in reading for each country and for the pooled sample of students (labelled as 'Total'). This figure provides three different types of information: i) how many predictors compose the best model specification; ii) which predictors are included in the best model; iii) the selection order of predictors in the best model. The latter is obtained looking at variable additions into the best-fitting models when the quantity of predictors increases. For instance, when looking at Germany the best one-predictor model is the one containing the variable on the number of books at home, while the best two-predictor model adds the class size variable to the former, the best three-predictor model adds the dummies for type of school and municipality size to the former, and so on. Somehow, we may then argue that the number of books at home is the most important predictor of students' reading scores in Germany, followed by the class size, the combination of the type of school and municipality size, and the other variables observed.[7]

Figure 1. Selection order of predictors and best model specification for reading score by country



Notes: The lighter the cell, the later the predictor was selected in the best-fitting model. The dependent variable is the average of the ten plausible values for reading (in logarithmic form). The 'Total' model refers to the pooled sample of students and also includes country dummies.
Source: Elaborations of the authors on PISA 2018 data.

---

correlation existing between these two variables is strong. Moreover, once this composite index is included in the basket of predictors, the results of efficient algorithms for predictor selection report it in the best-fitting models in one country only (Italy) while the number of books at home is preferred in all other cases. More details are available upon request to the authors.

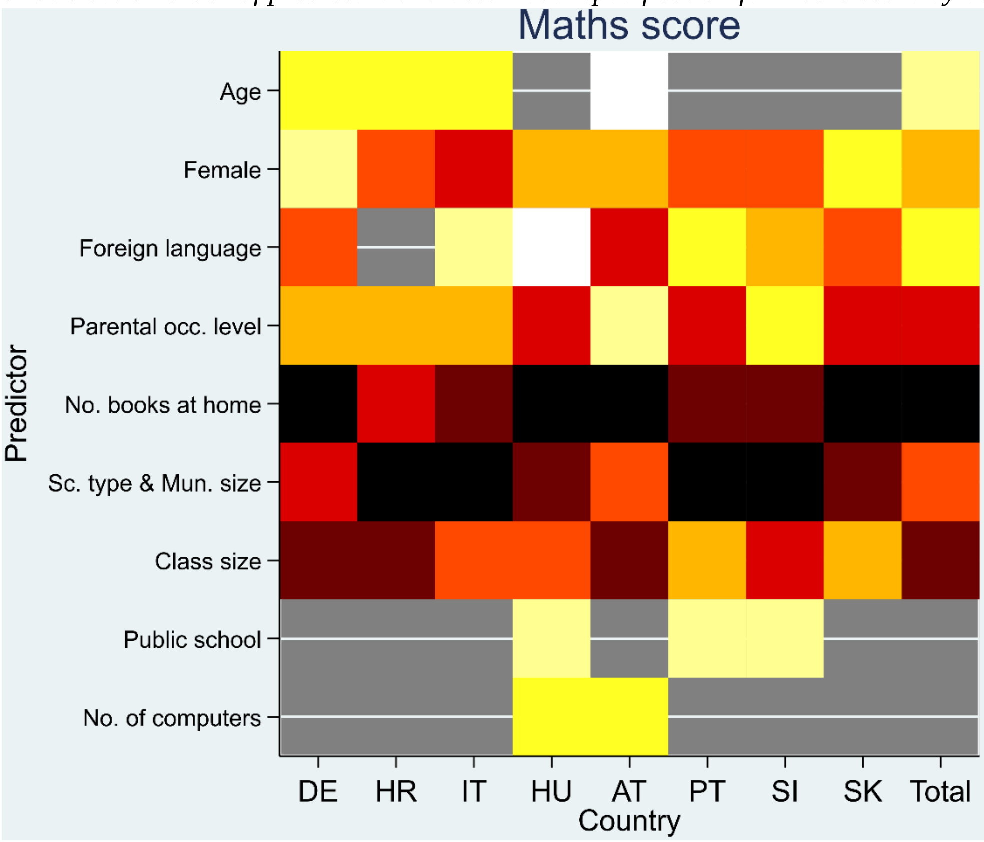
[6] From the variables provided by the PISA survey dataset that are somewhat related to student performance, we exclude here having repeated a school year, the percentage of full professors, the percentage of qualified professors, the percentage of government expenditure, and the percentage of student fees, due to the large number of missing values.

[7] To be noted, while in this application of the adopted methodology all best-fitting models appear nested when the quantity of predictors increases, the same situation may not necessarily stand in other applications.

As illustrated in Figure 1, only Italy, Hungary, and Austria (and the total sample) count eight predictors in their best model specifications, whereas Germany and Slovenia count seven predictors and the remaining countries (Croatia, Portugal, and Slovakia) have six predictors in their best models. The combination of predictors differs across countries as well. The public school dummy is included only in the best Hungarian model specification. The number of computers per student is selected in the best models for Italy, Hungary, and Austria only. Variables regarding the language students speak at home is included in all best models except for the Croatian one, while the age variable is missing in Hungary, Portugal, and Slovakia only. As regards the selection order of predictors, one of either the variable regarding the number of books at home or the set of dummies reflecting the type of school and the municipality size always represents the most important predictor of students' reading scores in the analysed countries. In terms of importance, these two variables are followed by class size, except for the Slovakian case where the class size is only sixth. When included, the public school dummy, the number of computers per student, and student age tend to be among the least important predictors.

When we replicate this analysis for the students' maths scores, some interesting differences occur with respect to the results presented above (Figure 2).

Figure 2. Selection order of predictors and best model specification for maths score by country



Notes: The lighter the cell, the later the predictor was selected in the best-fitting model. The dependent variable is the average of the ten plausible values for reading (in logarithmic form). The 'Total' model refers to the pooled sample of students and also includes country dummies.
Source: Elaborations of the authors on PISA 2018 data.

First, Portugal now reports as best a model specification with one additional predictor (the public school dummy), while the Italian and pooled samples have one variable less (the number of computers per student in both cases). Second, student gender is much more important here in the best Italian model, as it is third in this case, while the opposite occurs for speaking a foreign language. This difference represents an interesting corroboration regarding the size of the gender gap in maths scores in Italy. As many previous papers have found (Contini et al. 2017; Granato 2020), Italy is among the countries displaying the largest differential between boys and girls in STEM disciplines. The relevance of the gender variable to explain maths scores, in contrast to reading scores, confirms from a new perspective the challenge this country faces. Finally, the age variable is no longer present in the best Slovenian model, while the public school dummy is. The only country reporting no changes in either the best model specification or the selection order of predictors is Hungary.

5. Results

5.1. Estimation of effects on student performance

This section presents the results of OLS estimations to present the effects of selected determinants on the reading and maths performances of students.

In line with the evidence illustrated in Figure 1, the number of books at home (a proxy of parental education level and of the level of household well-being in general), the variable combining the type of school and municipality size, and class size have the greatest effects, *ceteris paribus*, on the reading scores of students (Table 3). Specifically, we observe an increasing effect according to the number of books owned (at least up to 500 books) and the number of students in the same classroom. It should be noted that as the number of books at home is self-reported by the students, it is possible that some of those who report owning more than 500 books overestimated the real number of books they have at home.

Table 3. Estimated effects on the reading scores of students

Variables	DE	HR	IT	HU	AT	PT	SI	SK	Total
Age	0.060***	0.021***	0.031***		0.039***		0.026***		0.041***
Female	0.033***	0.043***	0.011*	0.037***	0.036***	0.029***	0.044***	0.057***	0.032***
Foreign language	-0.098***		-0.044***	-0.062***	-0.073***	-0.067***	-0.051***	-0.111***	-0.073***
Average occup. level	0.026***	0.032***	0.036***	0.045***	0.027***	0.034***	0.005	0.054***	0.039***
High occupational level	0.052***	0.046***	0.039***	0.080***	0.037***	0.055***	0.030***	0.060***	0.061***
11–25 books	0.065***	0.033***	0.060***	0.078***	0.069***	0.031***	0.058***	0.097***	0.066***
26–100 books	0.127***	0.054***	0.106***	0.120***	0.106***	0.075***	0.083***	0.151***	0.121***
101–200 books	0.156***	0.077***	0.113***	0.159***	0.127***	0.100***	0.107***	0.201***	0.146***
201–500 books	0.187***	0.081***	0.126***	0.194***	0.167***	0.113***	0.113***	0.244***	0.172***
500 books or more	0.194***	0.022	0.125***	0.205***	0.154***	0.104***	0.104***	0.180***	0.173***
General - big city	-0.009	-0.011	-0.006	-0.016	0.006	0.034**	-0.015	-0.022	0.004
Vocational - big city	-0.098***	-0.142***	-0.134***	-0.103***	-0.087***	0.179***	-0.157***	-0.056***	-0.075***
Vocational - small city	-0.071***	-0.176***	-0.124***	-0.132***	-0.091***	0.139***	-0.167***	-0.120***	-0.067***
16–20 students	0.060	0.009	0.064	0.032	0.061***	0.027	-0.007	0.022	0.043**
21–25 students	0.126***	0.069***	0.131***	0.017	0.107***	0.067	0.019	0.024	0.101***
26 students or more	0.193***	0.101***	0.139***	0.064***	0.137***	0.105	0.068***	0.068***	0.149***
Public school				0.033*					
No. of computers			-0.015	0.043***	0.015***				-0.004
Constant	4.993***	5.792***	5.511***	5.956***	5.387***	5.885***	5.745***	5.946***	5.320***
Country fixed effects	No	No	No	No	No	No	No	No	Yes
Observations	3,761	6,311	10,631	4,539	5,519	5,316	5,395	4,987	46,459
R-squared	0.384	0.343	0.294	0.376	0.316	0.366	0.384	0.384	0.322

*Notes: Standard errors are clustered at the school level. *** p < 0.01, ** p < 0.05, * p < 0.1. Each country regression is estimated considering the best model specification identified in Section 4. The dependent variable is the average of the ten plausible values for reading (in logarithmic form).*
Source: Elaborations of the authors on PISA 2018 data.

As for the type of school and the size of the municipality where the school is located, results show that students attending vocational schools perform significantly worse than those in general schools. Among general school students, no performance differences related to municipality size arise, while those attending a vocational school in a big city tend to perform worse than those in a small city (differences between these two groups of students are statistically significant in Hungary and Slovakia). Portugal represents a clear exception across the European countries analysed here. In fact, class size does not engender any effect on the reading scores of Portuguese students, those attending a vocational school present greater scores than those in a general school, and students in big cities tend to perform better than those in small cities, independent of the type of school.

Looking at the demographic characteristics of students, female students always report higher reading scores than male ones, *ceteris paribus*, but this effect is very small and slightly significant in Italy. When included in the best models, the age of students appears positively relevant in explaining student performance (especially in Germany), while students speaking a foreign language at home have a significantly lower score (especially in Slovakia). In addition, parental occupational level plays an important and positive role in the reading scores of students, and its effects are fairly stable across European countries. Finally, attending a public school increases student performance in Hungary, whereas the number of computers at school engenders significantly greater scores in Hungary and Austria only.

As regards the effects on maths scores, the results displayed in Table 4 overall confirm what was seen for reading scores, with two important exceptions. First, female students now report significantly lower performances than males, *ceteris paribus*, especially in Italy. Second, students attending a general school in a big city do not show a better performance than those attending the same school in a small city in Portugal, but they present lower scores in Slovakia in this case.

Table 4. Estimated effects on the maths scores of students

Variables	DE	HR	IT	HU	AT	PT	SI	SK	Total
Age	0.067***	0.031***	0.033***		0.031***				0.042***
Female	-0.035***	-0.049***	-0.069***	-0.041***	-0.048***	-0.044***	-0.046***	-0.027***	-0.044***
Foreign language	-0.074***		-0.019**	-0.046***	-0.080***	-0.049**	-0.070***	-0.081***	-0.050***
Average occup. level	0.027***	0.033***	0.040***	0.040***	0.028***	0.036***	0.007	0.067***	0.042***
High occupational level	0.059***	0.053***	0.048***	0.073***	0.035***	0.055***	0.036***	0.081***	0.068***
11–25 books	0.059***	0.027***	0.038***	0.052***	0.056***	0.039***	0.043***	0.114***	0.055***
26–100 books	0.108***	0.049***	0.089***	0.104***	0.092***	0.095***	0.065***	0.165***	0.109***
101–200 books	0.140***	0.067***	0.111***	0.140***	0.108***	0.111***	0.091***	0.209***	0.139***
201–500 books	0.161***	0.073***	0.114***	0.175***	0.138***	0.126***	0.103***	0.243***	0.156***
500 books or more	0.168***	0.023*	0.107***	0.188***	0.145***	0.106***	0.082***	0.193***	0.156***
General - big city	-0.017	-0.010	-0.009	-0.011	-0.001	0.015	-0.017	-0.031*	-0.003
Vocational - big city	-0.093***	-0.147***	-0.116***	-0.088***	-0.071***	0.187***	-0.156***	-0.051***	-0.064***
Vocational - small city	-0.058**	-0.178***	-0.103***	-0.116***	-0.075***	0.154***	-0.156***	-0.100***	-0.050***
16–20 students	0.095*	0.002	0.075*	0.025	0.051**	0.021	0.011	0.007	0.052**
21–25 students	0.152***	0.064***	0.143***	0.008	0.097***	0.056	0.034	0.019	0.110***
26 students or more	0.205***	0.095***	0.153***	0.050**	0.124***	0.090	0.077***	0.052***	0.149***
Public school				0.025*		-0.020	-0.000		
No. of computers				0.031***	0.016***				
Constant	4.921***	5.668***	5.535***	6.046***	5.600***	5.940***	6.228***	6.037***	5.341***
Country fixed effects	No	No	No	No	No	No	No	No	Yes
Observations	3,761	6,311	10,631	4,539	5,519	5,316	5,395	4,987	46,459
R-squared	0.380	0.369	0.280	0.419	0.338	0.430	0.426	0.374	0.311

*Notes: Standard errors are clustered at the school level. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$. Each country regression is estimated considering the best model specification identified in Section 4. The dependent variable is the average of the ten plausible values for reading (in logarithmic form).*
Source: Elaborations of the authors on PISA 2018 data.

5.2. An application of the Gelbach decomposition method

In order to assess the underlying mechanisms behind the differences in cognitive test scores among the types of schools, we implement the well-known decomposition analysis proposed by Gelbach (2016). This method allows us to decompose the contribution of each covariate (selected through the procedure explained in Section 4) to the change in our coefficients of interest from the base to the full regressions. In other words, we identify the main dimensions explaining the cognitive gap between types of schools in each country pointed out in the baseline analysis. The main positive feature of this method is that the results are independent of the order in which the variables are included into the regression.

Tables 5 and 6 present the results using general schools in small cities as a reference category. In the 'Base OLS' and 'Full OLS' columns, we report the coefficients of interest obtained by, respectively, the baseline analysis and the complete analysis shown in the previous section. The 'Total explained' column displays the difference between the baseline and full estimates. The following columns display the contribution, in percentage terms, of each selected covariate in composing the 'Total explained' values. The set of characteristics explaining the gap between the different types of schools is heterogeneous across countries and across school types within countries.

Starting with the analysis of the reading score (Table 5), we point out that in Germany, Austria, and Slovenia, the 'Total explained' coefficient is negative and significant with respect to that of vocational schools in big cities, and this is due to social background (i.e. being a foreigner, the number of books at home, and parental occupation level). It is also interesting to note that the average size of classes in Germany and in Slovenia gives an advantage to vocational schools in big cities over general schools in small cities. In Austria, the gap between general and vocational schools in small cities is also significant and seems to mainly be due to social status, and in particular to the number of books at home.

In Croatia, Italy, and Slovenia, the difference between general schools in small and big cities is due in particular to class size, suggesting that the difference between small and big cities in these countries seems to be explained by the quality of schools, although in Italy the number of books and parental occupational level also matter.

The gap between general and vocational schools in small cities in Croatia and in Slovenia is explained by parental occupation, the number of books at home, and the gender variable, suggesting that females (who typically show a better performance in the reading tests) are grouped into general schools rather than vocational schools in these countries.

In Italy, the difference between general schools in small cities and vocational schools in big cities is explained only by social status: being a foreigner, parental occupational level, and above all the number of books at home. This stresses the preponderant role of the family in shaping children's educational careers in this country.

Hungary seems to be a country where individual, family, and school characteristics are all relevant. The gap between general schools in big cities and in small cities is due to social characteristics (i.e. the number of books at home and parental occupational level) and school characteristics (i.e. public school and class size), while the number of computers at school goes in the opposite direction and gives an advantage to general schools in small cities. As for the difference between average scores in general and vocational schools in small cities, females tend to be enrolled more commonly in general schools than in vocational schools, but parental occupational level and the number of books at home are also important factors explaining the gap. The variables 'class size' and 'public school' balance each other.

In Portugal, the 'total explained' coefficients are significant for both the type of vocational school and the relevant characteristics are the same, stressing that in this country the difference is particularly between general and vocational schools rather than big and small cities. The gap is due both to social (i.e. the number of books at home and parental occupational level) and school characteristics (i.e. class size). Finally, in Slovakia the general–vocational school gap within small cities is explained in particular by the number of books at home and class size.

Table 6 shows results of the Gelbach decomposition on maths scores. In Germany, none of the categories show a statistically significant 'total explained', and this is significant only between general schools in big and in small cities in Croatia, Italy, and Hungary. In the latter, the decomposition shows heterogeneous results: by making the 'total explained' equal to one hundred percent, the number of books at home explains 40 percent, parental occupational level accounts for 25 percent, and the variables ‘class size’ and ‘public school’ together compose 45 percent. In Croatia and in Italy, the most important variable is class size.

In Austria and Portugal, the ‘total explained’ is statistically significant for vocational schools in both big and small cities. The number of books at home is the most important characteristic for both types of vocational school in Austria and for vocational schools in small cities in Portugal. In this country, half of the gap between general and vocational schools in big cities is explained by the number of books at home and almost a third is explained by parental occupational level and class size.

In Slovenia, all three levels of the variable of interest present a statistically significant explained part. The gap between general schools in big and small cities is almost entirely explained by class size, while for the difference between general schools in small cities and vocational schools in big cities there are many relevant covariates: parental occupational level, foreigner status, and in particular, the number of books at home explain this gap. On the other hand, class size and gender present a negative sign, meaning that they provide an advantage for vocational schools in big cities. Finally, in Slovakia the number of books at home and parental occupational level each explain a third of the gap between general schools in big cities and the reference category, general schools in small cities, while the difference between general and vocational schools in small cities is half due to the number of books at home.

To sum up, our results point out that the gap in test scores for both reading and mathematics on the basis of city size seems to depend relatively more on the quality of the school, and in particular on class size, while the differences between general and vocational schools appears to be explained above all by family social status.

Table 5. Decomposition of the link between type of school and reading score by country

Country	Type of school	Base OLS	Full OLS	Total explained	Age	Female	Foreign language	Parental occ. level	No. of books at home	Class size	Public school	No. of computers
DE	G-BC	0.028	-0.009	0.036	-	-	-	-	-	-		
	V-BC	-0.145***	-0.098***	-0.047*	-6.38%	4.26%	44.68%	21.28%	53.19%	-14.89%		
	V-SC	-0.106***	-0.071***	-0.035	-	-	-	-	-	-		
HR	G-BC	0.043***	-0.011	0.054***	-1.85%	-3.70%		12.96%	9.26%	83.33%		
	V-BC	-0.148***	-0.142***	-0.006	-	-		-	-	-		
	V-SC	-0.217***	-0.176***	-0.041***	0.00%	21.95%		31.71%	39.02%	4.88%		
IT	G-BC	0.031	-0.006	0.037***	-2.70%	0.00%	0.00%	10.81%	16.22%	67.57%		2.70%
	V-BC	-0.163***	-0.134***	-0.029**	3.45%	10.34%	20.69%	27.59%	72.41%	-34.48%		3.45%
	V-SC	-0.140***	-0.124***	-0.016	-	-	-	-	-	-		-
HU	G-BC	0.024	-0.016	0.039**		-5.13%	0.00%	25.64%	43.59%	20.51%	35.90%	-20.51%
	V-BC	-0.127***	-0.103***	-0.024			-	-	-	-	-	-
	V-SC	-0.180***	-0.132***	-0.048*		18.75%	2.08%	25.00%	66.67%	37.50%	-39.58%	-8.33%
AT	G-BC	-0.004	0.006	-0.010	-	-	-	-	-	-		-
	V-BC	-0.137***	-0.087***	-0.050***	-4.00%	8.00%	20.00%	18.00%	60.00%	2.00%		-2.00%
	V-SC	-0.133***	-0.091***	-0.042***	-4.76%	7.14%	4.76%	23.81%	66.67%	19.05%		-11.90%
PT	G-BC	0.043**	0.034**	0.009			-	-	-	-		
	V-BC	0.251***	0.179***	0.071***		5.63%	0.00%	26.76%	40.85%	28.17%		
	V-SC	0.170***	0.139***	0.031***		9.68%	3.23%	25.81%	48.39%	12.90%		
SI	G-BC	0.005	-0.015	0.020**	5.00%	-5.00%	-5.00%	10.00%	0.00%	90.00%		
	V-BC	-0.189***	-0.157***	-0.032***	0.00%	25.00%	12.50%	25.00%	71.88%	-34.38%		
	V-SC	-0.219***	-0.167***	-0.052***	1.92%	17.31%	5.77%	17.31%	50.00%	11.54%		
SK	G-BC	0.004	-0.022	0.026*		3.85%	23.08%	23.08%	30.77%	23.08%		
	V-BC	-0.066**	-0.056***	-0.010			-	-	-	-		
	V-SC	-0.186***	-0.120***	-0.067***		8.96%	13.43%	11.94%	41.79%	23.88%		

*Notes: Standard errors are clustered at the school level. *** p < 0.01, ** p < 0.05, * p < 0.1. The dependent variable is the average of the ten plausible values for reading (in logarithmic form). All results are weighted. The decomposition method is based on Gelbach (2016).*
Source: Elaborations of the authors on PISA 2018 data

Table 6. Decomposition of the link between type of school and mathematics score by country

Country	Type of school	Base OLS	Full OLS	Total explained	Age	Female	Foreign language	Parental occ. level	No. of books at home	Class size	Public school	No. of computers
DE	G-BC	0.015	-0.017	0.033	-	-	-	-	-	-		
	V-BC	-0.129***	-0.093***	-0.036	-	-	-	-	-	-		
	V-SC	-0.086**	-0.058**	-0.028	-	-	-	-	-	-		
HR	G-BC	0.048***	-0.010	0.058***	-0.10%	3.45%		12.07%	8.62%	77.59%		
	V-BC	-0.132***	-0.147***	0.015	-	-		-	-	-		
	V-SC	-0.199***	-0.178***	-0.022**	-	-		-	-	-		
IT	G-BC	0.026	-0.009	0.035***	-0.001	-2.86%	0.00%	14.29%	17.14%	74.29%		
	V-BC	-0.124***	-0.116***	-0.008	-	-	-	-	-	-		
	V-SC	-0.103***	-0.103***	0.000	-	-	-	-	-	-		
HU	G-BC	0.028	-0.011	0.039**		5.13%	0.00%	23.08%	41.03%	17.95%	28.21%	-15.38%
	V-BC	-0.102***	-0.088***	-0.014		-	-	-	-	-	-	-
	V-SC	-0.144***	-0.116***	-0.028		-	-	-	-	-	-	-
AT	G-BC	-0.009	-0.001	-0.009		-	-	-	-	-		-
	V-BC	-0.108***	-0.071***	-0.037***	0.10%	-16.22%	27.03%	21.62%	70.27%	0.00%		-2.70%
	V-SC	-0.108***	-0.075***	-0.032***	0.20%	-12.50%	6.25%	28.13%	78.13%	21.88%		-15.63%
PT	G-BC	0.021	0.015	0.006		-	-	-	-	-	-	
	V-BC	0.247***	0.187***	0.061***		-9.84%	0.00%	31.15%	52.46%	29.51%	-3.28%	
	V-SC	0.176***	0.154***	0.022*		-18.18%	0.00%	36.36%	77.27%	18.18%	-13.64%	
SI	G-BC	0.002	-0.017	0.019**		5.26%	-5.26%	10.53%	5.26%	89.47%	0.00%	
	V-BC	-0.174***	-0.156***	-0.018**		-44.44%	33.33%	55.56%	116.67%	-55.56%	0.00%	
	V-SC	-0.191***	-0.156***	-0.035***		-25.71%	11.43%	28.57%	65.71%	17.14%	0.00%	
SK	G-BC	-0.008	-0.031*	0.023*		-4.35%	17.39%	34.78%	30.43%	21.74%		
	V-BC	-0.051**	-0.051***	0.000		-	-	-	-	-		
	V-SC	-0.157***	-0.100***	-0.057***		-5.26%	10.53%	19.30%	52.63%	22.81%		

*Notes: Standard errors are clustered at the school level. *** p < 0.01, ** p < 0.05, * p < 0.1. The dependent variable is the average of the ten plausible values for mathematics (in logarithmic form). All results are weighted. The decomposition method is based on Gelbach (2016).*
Source: Elaborations of the authors on PISA 2018 data

6. Conclusions

In this study, we used PISA 2018 data from eight European countries to identify the main determinants of students' mathematics and reading scores. In particular, we focus on the role played by two specific determinants: the type of secondary school and its location. Our econometric strategy starts with the computational procedure proposed by Furnival and Wilson (1974), through which we select the best set of predictors of student performance both in reading and in maths and then regress the selected variables on the reading and maths performances of students. Finally, we measure the relative importance of observable characteristics to explain the differences in scores between general and vocational schools and big and small cities.

The results in all steps are quite homogeneous across countries. First, the most important predictors resulting from the computational procedure are the number of books at home and the type and location of the school. These characteristics also have the greatest effect on student performance. More specifically, while the type of school presents results rather as expected, this is not true for the location of the school. Students attending general schools perform better than those in vocational schools. Among general school students, there are no differences between big and small cities, whereas when considering only students in vocational schools those in small cities tend to perform better than those in big cities. In this context, Portugal represents an outlier since students attending a vocational school show higher scores than those in general schools, on average, and students in big cities show better results than those in small cities. Finally, we implement a Gelbach decomposition to capture the main drivers of these differences. The results of this analysis are homogeneous across countries and are very interesting in terms of policy implications aimed at improving education systems. The analysis clearly points out that family socioeconomic characteristics are the main cause of the gap between general and vocational schools, and the quality of the school represents the main explanation for the score differences between big and small cities.

These results are relevant in terms of policy as it is necessary to understand where public aid and support should be directed to improve student performance. Lower educated families tend to underestimate a child's education career and tend to show less support for their children's education (Heckman 2006). Consequently, these students are pushed towards schools focused on providing job skills rather than pursuing tertiary education (Dustmann 2004; Schizzerotto and Barone 2006; Mocetti 2012). Schools in small cities and rural areas tend to receive fewer funds than those in big cities. Thus, they are not able to invest in both the structure and the necessary tools to improve students' reading and mathematical skills (Mathis 2003). These pieces of evidence could engender an increase in inequality. Azzolini and Vergolini (2014) argue that reforms aiming to reduce curricula differences between tracks could be pursued to mitigate this.

It is important to invest resources in such a way as to guarantee that all students develop a common set of core competences and are exposed to common learning content for at least the first few years of upper-secondary school. The urban–rural achievement gaps can best be addressed through initiatives or reforms at the school level to take into account differences in the composition of students.

References

Acosta ST, Hsu HY (2014) Negotiating diversity: an empirical investigation into family, school and student factors influencing New Zealand adolescents' science literacy. Educational Studies 40: 98–115

Ammermueller A (2007) Poor background or low returns? Why immigrant students in Germany perform so poorly in the Programme for International Student Assessment. Education Economics 15(2): 215–230

Arnold Ml, Newman JH, Gaddy BB, Dean CB (2005) A look at the condition of rural education research: Setting a direction for future research. Journal of Research in Rural Education 20(6): 1–25

Azzolini D, Vergolini L (2014) Tracking, inequality and education policy. Looking for a recipe for the Italian case. Scuola Democratica, Learning for Democracy, 2. doi: 10.12828/77685

Banerjee AV, Shawn C, Duflo E, Linden L (2007) Remedying Education: Evidence from Two Randomized Experiments in India. The Quarterly Journal of Economics 122(3): 1235–1264

Bauch P (2001) School–community partnerships in rural schools: Leadership, renewal, and a sense of place. Peabody Journal of Education 76(2): 204–221

Bingley P, Jensen VM, Walker I (2005) The Effect of School Class Size on Length of Post-Compulsory Education: Some Cost-Benefit Analysis. IZA Discussion Paper 1605

Bloise F, Tancioni M (2021) Predicting the spread of COVID-19 in Italy using machine learning: Do socio-economic factors matter? Structural Change and Economic Dynamics 56: 310–329

Bonacini L, Gallo G, Patriarca F (2021) Identifying policy challenges of COVID-19 in hardly reliable data and judging the success of lockdown measures. Journal of Population Economics 34(1): 275–301

Bratti M, Checchi D, Filippin A (2007) Geographical differences in Italian students' mathematical competencies: evidence from PISA 2003. Giornale degli Economisti e Annali di Economia 66 (3): 299–333

Browning M, Heinesen E (2007) Class Size, Teacher Hours and Educational Attainment. The Scandinavian Journal of Economics 109(2): 415–438

Brunello G, Rocco L (2015) The Effects of Vocational Education on Adult Skills and Wages. Social, Employment and Migration Working Papers 168

Chernozhukov V, Fernández-Val I, Melly B (2020) Fast algorithms for the quantile regression process. Empirical Economics, 1–27. https://doi.org/10.1007/s00181-020-01898-0

Chetty R, Friedman J, Rockoff J (2014) Measuring the Impacts of Teachers II: Teacher Value-Added and Student Outcomes in Adulthood. American Economic Review 104(9): 2633–2679

Coe R, Aloisi C, Higgins S, Major LE (2014) What makes great teaching? Review of the underpinning research. The Sutton Trust, London

Coleman JS (1966) Equality of Educational Opportunity. National Center for Educational Statistics. John Hopkins University, Washington DC

Crosnoe R, Johnson MK, Elder GH (2004) School size and the interpersonal side of education: An examination of race/ethnicity and organizational context. Social Science Quarterly 85(5): 1259–1274

Denisova-Schmidt E, Huber M, Leontyeva E, Solovyeva A (2021) Combining experimental evidence with machine learning to assess anti-corruption educational campaigns among Russian university students. Empirical Economics 60(4): 1661–1684

Denny K, Oppedisano V (2013) The surprising effect of larger class sizes. Labour Economics 23: 57–65

Dustmann C (2004) Parental background, secondary school track choice, and wages. Oxford Economic Papers 56(2): 209–230

Escueta, M, Nickow AJ, Oreopoulos P, Quan V (2020) Upgrading Education with Technology: Insights from Experimental Research. Journal of Economic Literature, 58(4): 897–996

Farooq MS, Chaudhry AH, Shafiq, M, Berhanu G (2011) Factors affecting students' quality of academic performance: A case of secondary school level. Journal of Quality and Technology Management 7: 1–14

Fuchs T, Woessmann L (2007) What accounts for international differences in student performance? A reexamination using PISA data. Empirical Economics 32(2–3): 433–462

Furnival GM, Wilson RW (1974) Regression by leaps and bounds. Technometrics 16(4): 499–511 https://doi.org/10.2307/1267601

Gamazo A, Martínez-Abad F, Olmos-Migueláñez S, Rodríguez-Conde MJ (2018) Assessment of factors related to school effectiveness in PISA 2015. A Multilevel Analysis. Revista de Education 379: 56–84. doi: 10.4438/1988-592X-RE-2017-379-369

Giambona F, Porcu M (2015) Student background determinants of reading achievement in Italy. A quantile regression analysis. International Journal of Educational Development 44: 95–107

Giannelli GC, Rapallini C (2016) Immigrant student performance in Math: Does it matter where you come from? Economics of Education Review 52: 291–304

Granato S (2020) Early Influences and the Gender Gap in STEM. Available at SSRN: https://ssrn.com/abstract=3845391

Hanushek E (2008) Education Production Functions. In: Durlauf SN, Blume LE (ed) The New Palgrave Dictionary of Economics, Basingstoke. Palgrave Macmillan, London, pp. 749–752

Hanushek EA, Woessmann L (2011) The Economics of International Differences in Educational Achievement. Handbook of the Economics of Education 3: 89–200.

Hanushek EA, Woessmann L (2007) The Role of Education Quality for Economic Growth. World Bank Policy Research Working Paper 4122. https://doi.org/10.1596/1813-9450-4122

Heckman JJ (2006) Skill formation and the economics of investing in disadvantaged children. Science 312(5782): 1900–1902. doi:10.1126/science.1128898

Karakolidis A, Pitsia V, Emvalotis A (2016) Examining students' achievement in mathematics: a multilevel analysis of the programme for international student assessment (PISA) 2012 data for Greece. Int. J. Educ. Res. 79: 106–115. doi: 10.1016/j.ijer.2016.05.013

Kim DH, Law H (2012) Gender gap in maths test scores in South Korea and Hong Kong: role of family background and single-sex schooling. International Journal of Educational Development 32: 92–103

Lawless JF, Singhal K (1978) Efficient Screening of Nonnormal Regression Models. Biometrics 34(2): 318–327. https://doi.org/10.2307/2530022

Lee JW, Barro RJ (2001) Schooling quality in a cross-section of countries. Economica 68(272): 465–488

Masci C, Johnes G, Agasisti T (2018) Student and school performance across countries: A machine learning approach. European Journal of Operational Research 269: 1072–1085

Mathis W (2003) Financial Challenges, Adequacy, and Equity in Rural Schools and Communities. Journal of Education Finance 29(2): 119–136

Mocetti S (2012) Educational choices and the selection process: before and after compulsory schooling. Education Economic 20(2): 189–209

Murat M, Bonacini L (2020) Coronavirus pandemic, remote learning and education inequalities. GLO Discussion Paper 679

Nisar N, Muhammad KM, Dogar AH (2017) Determinants of Students' Academic Achievement at Secondary School Level. Bulletin of Education and Research 39(1): 145–158

OECD (2010) PISA 2009 Results: Overcoming Social Background: Equity in Learning Opportunities and Outcomes (Volume II). OECD Publishing, Paris

OECD (2016) Low-Performing students: Why they fall behind and how to help them succeed. OECD Publishing, Paris

OECD (2019a) Learning in rural schools: insights from PISA, TALIS and the literature, OECD Education Working Paper 196. OECD Publishing, Paris

OECD (2019b) Economic Policy Reforms 2019: Going for Growth. OECD Publishing, Paris

Pholphirul P (2016) Pre-primary education and long-term education performance: evidence from programme for international student assessment (PISA) Thailand. Journal of Early Childhood Research 15: 410–432. doi: 10.1177/1476718x15616834

Raitano M, Vona F (2013) Peer heterogeneity, school tracking and students' performances: evidence from PISA 2006. Applied Economics 45(32): 4516–4532

Schizzerotto A, Barone C (2006) Sociologia dell'istruzione. Il Mulino, Bologna

Schnepf SV (2007) How Different Are Immigrants? Cross-Country and Cross-Survey Analysis of Educational Achievement. Journal of Population Economics 20: 527–545

Semke C, Sheridan S (2012) Family–school connections in rural educational settings: A systematic review of the empirical literature. The School Community Journal, 22(1), 21–48

Smidova Z (2019) Educational outcomes: A literature review of policy drivers from a macroeconomic perspective. OECD Economics Department Working Papers 1577. OECD Publishing, Paris. https://doi.org/10.1787/990801aa-en

Tonello M (2016) Peer effects of non-native students on natives' educational outcomes: mechanisms and evidence. Empirical Economics 51(1): 383–414

Woessmann L (2016) The importance of school systems: Evidence from international differences in student achievement. Journal of Economic Perspectives 30: 3–32

Woessmann L (2003) Schooling resources, educational institutions, and student performance: The international evidence. Oxford Bulletin of Economics and Statistics 65(2): 117–170

Woessmann L, Luedemann E, Schuetz G, West MR (2013) School Accountability, Autonomy and Choice around the World. In: Dronkers J, Van Der Velden R, Windzio M (ed) Integration and Inequality in Educational Institutions. Springer, Dordrecht, pp. 123–152

Appendix

Table A.1. Variable descriptions

Variable	Definition	Mean	Std. Dev.	Min.	Max.
Dependent variables					
Log(Read score)	Continuous variable representing the logarithmic transformation of student scores in reading	6.185	0.207	5.081	6.672
Log(Math score)	Continuous variable representing the logarithmic transformation of student scores in mathematics	6.199	0.185	5.257	6.635
Control variables					
Age	Continuous variable representing student age at the time of interview (months included as a decimal)	15.796	0.288	15.250	16.330
Female	Binary variable taking a value of 1 for female and 0 for male	0.491	0.500	0.000	1.000
Foreign language	Binary variable taking a value of 1 for those speaking a foreign language at home and 0 otherwise	0.141	0.348	0.000	1.000
Low occupational level	Binary variables representing the highest parental occupation level. Based on the ISEI classification of occupations, the occupational level is high if at least one parent attained an occupation coded between 70 (senior government official) and 89 (medical doctor), middle if the highest occupation attained ranges between 40 (pharmaceutical technician and assistant) and 69 (chief executive, senior official and legislator), and low for occupations in lower levels.	0.393	0.488	0.000	1.000
Average occupational level		0.359	0.480	0.000	1.000
High occupational level		0.249	0.432	0.000	1.000
11 books or fewer	Binary variables representing the number of books at home	0.117	0.322	0.000	1.000
11–25 books		0.158	0.365	0.000	1.000
26–100 books		0.292	0.455	0.000	1.000
101–200 books		0.191	0.393	0.000	1.000
201–500 books		0.152	0.359	0.000	1.000
500 books or more		0.089	0.284	0.000	1.000
General - big city	Binary variables combining the type of school and the size of the municipality where the school is located. A municipality is considered as a big city if it has more than 15,000 inhabitants and as a small city otherwise.	0.403	0.491	0.000	1.000
General - small city		0.096	0.295	0.000	1.000
Vocational - big city		0.346	0.476	0.000	1.000
Vocational - small city		0.155	0.362	0.000	1.000
15 students or fewer	Binary variables representing the class size of the interviewed student	0.028	0.164	0.000	1.000
16–20 students		0.159	0.366	0.000	1.000
21–25 students		0.428	0.495	0.000	1.000
26 students or more		0.385	0.487	0.000	1.000
Public school	Binary variable taking a value of 1 if the interviewed student attends a public school and 0 otherwise	0.937	0.242	0.000	1.000
No. of computers	Continuous variable representing the number of computers per student in the school attended by the interviewed student	0.599	0.764	0.000	14.667

Declarations

Funding

No.

Conflicts of interest/Competing interests

No.

Availability of data and material

PISA survey data are available at the following link: https://www.oecd.org/pisa/.

Code availability

Stata codes used by the authors are available upon request.

RECent Working Papers Series

The 10 most RECent releases are:

No. 149	COVID-19, RACE, AND GENDER (2021) G. Bertocchi, A. Dimico
No. 148	MACROECONOMIC UNCERTAINTY AND VECTOR AUTOREGRESSIONS (2020) M. Forni, L. Gambetti, L. Sala
No. 147	COMMON COMPONENTS STRUCTURAL VARS (2020) M. Forni, L. Gambetti, M. Lippi, L. Sala
No. 146	ASYMMETRIC EFFECTS OF MONETARY POLICY EASING AND TIGHTENING (2020) D. Debortoli, M. Forni, L. Gambetti, L. Sala
No. 145	COVID-19, RACE, AND REDLINING (2020) G. Bertocchi, A. Dimico
No. 144	BITTER SUGAR: SLAVERY AND THE BLACK FAMILY (2020) G. Bertocchi, A. Dimico
No. 143	IS THE UNEMPLOYMENT INFLATION TRADE-OFF STILL ALIVE IN THE EURO AREA AND ITS MEMBER COUNTRIES? IT SEEMS SO (2019) A. Ribba
No. 142	GENDER GAPS IN EDUCATION (2019) G. Bertocchi, M. Bozzano
No. 141	THE DYNAMIC EFFECTS OF MONETARY POLICY AND GOVERNMENT SPENDING. SHOCKS ON UNEMPLOYMENT IN THE PERIPHERAL EURO AREA COUNTRIES (2019) P. Dallari, A. Ribba
No. 140	LABOUR MARKET EFFECTS OF CROWDWORK IN US AND EU: AN ENPIRICAL INVESTIGATION (2018) M. Cantarella, C. Strozzi

The full list of available working papers, together with their electronic versions, can be found on the RECent website: http://www.recent.unimore.it/site/home/publications/working-papers.html