INAPP
ISTITUTO NAZIONALE PER L'ANALISI DELLE POLITICHE PUBBLICHE

WORKING PAPER

INAPP WP n. 112

New evidence on what influences student performance along the distribution of test scores

Luca Bonacini
Irene Brunetti
Giovanni Gallo

ISSN 2784-8701

DICEMBRE 2023

New evidence on what influences student performance along the distribution of test scores

Luca Bonacini
Università di Bologna, Bologna, Italia
Global Labor Organization (GLO), Essen, Germany
l.bonacini@unibo.it

Irene Brunetti
Istituto nazionale per l'analisi delle politiche pubbliche (INAPP), Roma, Italia
Global Labor Organization (GLO), Essen, Germany
i.brunetti@inapp.gov.it

Giovanni Gallo
Università degli studi di Modena e Reggio Emilia (UNIMORE), Modena, Italia
Global Labor Organization (GLO), Essen, Germany
Corresponding author: giovanni.gallo@unimore.it

DICEMBRE 2023

Il working paper è stato realizzato nell'ambito del PTA Inapp 2021-2023, con riferimento alle linee di attività afferenti alla Struttura Imprese e lavoro

The opinions expressed here reflect only the authors' views and not their institutions'. The INAPP is not responsible for any use that can be made of the present results. The other usual disclaimers apply.

CONTENTS:

INAPP – Istituto nazionale per l'analisi delle politiche pubbliche

Corso d'Italia 33	Tel. +39 06854471	
00198 Roma, Italia	Email: urp@inapp.gov.it	www.inapp.gov.it

ABSTRACT

New evidence on what influences student performance along the distribution of test scores

This paper investigates the influences engendered by both the type and localization of schools on the distribution of test scores in four European countries (i.e. Germany, Italy, Portugal, and Slovenia). Based on PISA 2018 data and applying the unconditional quantile regressions, results show that the type and localization of schools significantly influence the students' performances distribution, as well as their inequality levels. In particular, replacing general schools in small cities with vocational schools (regardless their localization) tends to decrease mean value and increase the Gini index of test scores. These influences appear greater in Slovenia and smaller in German.

KEYWORDS: education inequalities, PISA, schooling tracking, student performance, RIF regressions
JEL CODES: C21, I21, I24

L'articolo analizza l'effetto che il tipo e la localizzazione della scuola hanno sulla distribuzione delle performance, misurate dai punteggi ottenuti ai test di matematica e lettura, degli studenti delle scuole superiori, in quattro diversi Paesi europei (Germania, Italia, Portogallo e Slovenia). A partire dai dati PISA 2018, i risultati ottenuti applicando stime quantiliche non condizionate, suggeriscono che entrambe i fattori influenzano significativamente sia la distribuzione dei risultati degli studenti che il livello di disuguaglianza. In particolare, rispetto a coloro che frequentano un liceo localizzato in piccoli centri urbani, gli studenti iscritti in scuole professionali o tecniche (indipendentemente dalla loro localizzazione) mostrano punteggi medi più bassi e il livello di disuguaglianza, misurato dall'indice di Gini, tende ad essere più alto. In termini comparativi, questi effetti appaiono molto pronunciati in Slovenia, meno invece in Germania.

PAROLE CHIAVE: Disuguaglianza di istruzione, PISA, performance degli studenti, regressioni RIF

CODICI JEL: C21, I21, I24

DOI: 10.53223/InappWP_2023-112

Cite as:
Bonacini L., Brunetti I., Gallo G. (2023), *New evidence on what influences student performance along the distribution of test scores*, Inapp Working Paper n.112, Roma, Inapp

1. Introduction

Many studies have investigated the role played by different variables on the quality of students' performance (Hanushek 1996; Fuchs and Woessmann 2007; Hanushek and Woessmann 2011; Wang 2021). The type (general or vocational) and the localization (in a small or big city) of the school are viewed as important determinants and possible sources of inequality between students. Indeed, students with vocational education tend to perform worse than those with a general education both in terms of labour market outcomes and level of basic skills (Brunello and Rocco 2015; Bratti *et al.* 2007). Moreover, small areas have long been associated with lesser opportunities, one of which is the provision of education (Arnold *et al.* 2005). Students at school located in small cities appear disadvantaged in terms of academic achievement with respect to their urban counterparts (OECD 2010). Researchers have generally analysed the impact and the effectiveness of schools or socio-economic factors on students' achievements relying on OLS or IV estimation approaches. However, as discussed by Eide and Showalter (1998), these methods allow observing the conditional effect only on at the average and they do not permit the inspection of the effect heterogeneity along the test scores distribution, thus missing important implications for policy intervention. Among few studies using a quantile regression method, Levin (2001) and Rangvid (2007) argue that, unlike what happens for students at the top of the reading test scores distribution, the peers' effect is stronger for students at the lower hand. Wang (2021), analysing the heterogeneous effects of peers' parental education on students' scholastic performance, shows an opposite result: students who have high-educated parents benefit more from peer groups characterized by higher levels of parental education compared to students with medium- or low-educated parents.

Therefore, following the strand of literature investigating the heterogenous effects of school's variables on students' performance, this paper uses the unconditional quantile regression (UQR) proposed by Firpo *et al.* (2009). We want to test to what extent the type and localization of schools have a significant influence on the distribution of test scores, focusing on the Gini index, the mean and decile values as distributional statistics.

To develop the empirical analysis, we use the OECD-PISA 2018 data which collects information on 15-year-old students, their households, and school's characteristics, as well as their levels of reading, mathematics, and science literacy. The results of each test 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. Among the countries participating to the survey, we focus on four European countries (i.e. Germany, Italy, Portugal, Slovenia) where the school tracking begins when students are 15 years old or before[1].

[1] Starting by the sample containing all the European Union countries, we have excluded those countries where school systems establish that the tracking begins after 15 years old (source: <https://bitly.ws/35PL3>) and those in which the math or the reading performance of students still in lower secondary level is statistically different to the ones reported by those already in an upper secondary school. Among the eight countries satisfying these criteria, for the sake of brevity, we further excluded Austria, Croatia, Hungary and Slovakia as the results for these countries are overall in line with those reported by the four countries here analysed. More details are available upon request to the authors.

This country selection allows us to analyze students' performances once they have already chosen the upper secondary school to attend.

2. Econometric methods

For each country, let F be the distribution function of test scores and $v(F)$ denote a distributional statistic, such as the mean or a quantile. We identify four different types of schools (i.e. general school in big city, general school in small city, vocational school in big city, and vocational school in small city), so that F can be expressed as

$$F(y) = \sum_{x=1}^{4} s_x F_x(y),$$

where y is the test score (i.e. the outcome variable), Fx is the test score distribution among students attending the type of school x, and sx is the proportion of the total population of students attending that type of school.

The UQR method proposed by Firpo *et al.* (2009) aims to evaluate the impact of marginal changes in the distribution of the explanatory variables on the distributional statistic $v(F)$. To be applied, this method involves the calculation of the Recentered Influence Function (RIF) which is defined as

$$\mathrm{RIF}(y; v, F) = v(F) + \mathrm{IF}(y; v, F) = v(F) + \lim_{t \downarrow 0} \frac{v\big((1-t)F + t\Delta_y\big) - v(F)}{t},$$

where the IF(y;v,F) is the influence function initially introduced by Hampel (1974). According to Firpo *et al.* (2009), once the values of RIF(y;v,F) are computed for all observations, the effects of a marginal change in the distribution of the variable of interest (i.e. type of school) on the distributional statistic $v(F)$ can be correctly calculated through a simple OLS estimation. Specifically, following Choe and Van Kerm (2018), we estimate the 'unconditional effect' assuming as marginal change a 10% swapping share of students from one type of school (i.e. general school in small city) to the another one (to be noted, in this 'shares swap' scenario, within-groups test score distributions remain constant). The core idea of this methodology is the following: if the described marginal change engenders significant effects on distributional statistics, then the type of school influences the test score distribution. In other words, the more the estimated coefficients are bigger and distant from zero the more the type of school attended plays an important role in the test score distribution of a country.

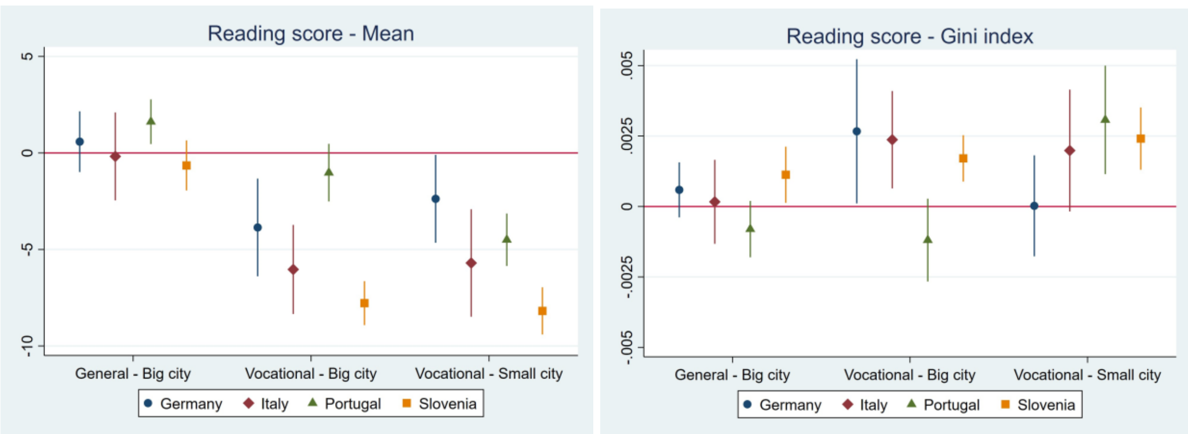
The UQR method also allows for considering relevant characteristics which may diverge among students of different types of school and therefore potentially lead to incorrect effects on the distributional statistics. We then regressed RIFs on a vector of type of school dummies and a vector of covariates including: gender (i.e., female or male), language spoken at home (i.e., local or foreign), age (both year and months), highest parental occupational status (i.e., high, average, and low occupational level on the basis of the ISEI classification), 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), class size (i.e., 15 students or fewer, 16-20 students, 21-25 students, 26 or more students), public school dummy, and the number of computers per student[2].

In this analysis, we estimate the unconditional effects of types of school on the test scores distribution focusing on the following distributional statistics: the mean, the Gini index, and the nine deciles. The main analysis looks at the influences on students' scores in reading, while those regarding test scores in mathematics are available in the appendix (figures A1 and A2). Both test scores are scaled to fit approximately normal distributions, with means for OECD countries around 500 score points and standard deviations around 100 score points. As usual in empirical studies using PISA data, all descriptive statistics and estimates take into account individual sample weights provided.

3. Results

Figure 1 highlights that a marginal change of students in general schools from a small to a big city would not change the average reading score in Germany, Italy, and Slovenia, while it would have a positive influence in Portugal. A shift of students toward vocational schools would instead engender a reduction in average reading scores. This relationship is particularly strong in Slovenia and Italy, where the average reading score would decrease by more than 5 points. As regards the results on Gini index, a shift of students from general schools in a small city to the other cases would increase the inequality levels of students' performances. Specifically, this positive influence appears always significant (at 5% level) in Slovenia, while it is significant only in case of specific changes in the other countries: shift toward vocational schools in big cities for Germany and Italy, and shift toward vocational schools in small cities for Portugal.

Figure 1. Influences on mean value and Gini index of students' performance in reading (with 95% CI)

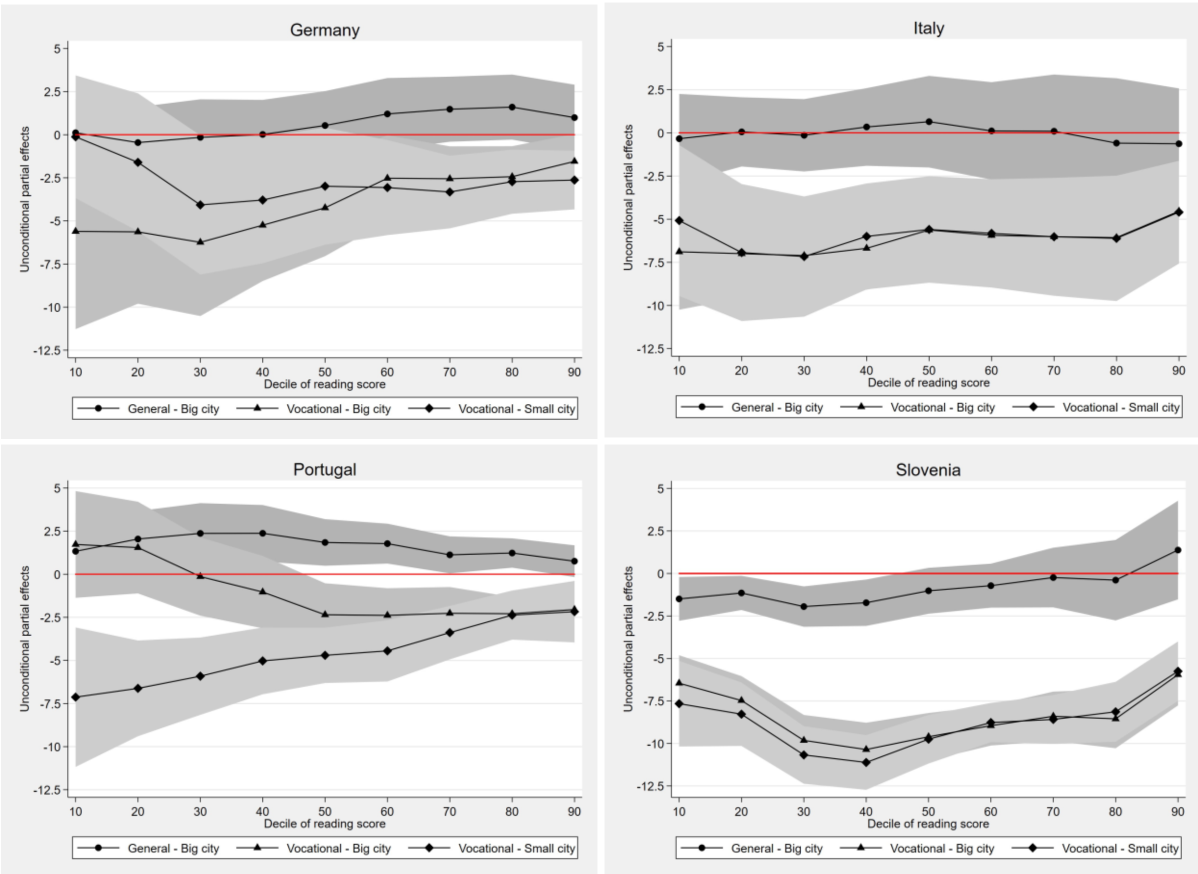


Note: full estimates are provided in the appendix (table A1).
Source: elaborations of the Authors on PISA 2018 data

[2] From the variables provided by the PISA survey dataset and somewhat related to the students' performance, we exclude here having repeated almost a school year, the percentage of full professors, the percentage of qualified professors, the percentage of government expenditure, and the percentage of student fees because of their large extent of missing values.

Looking at the estimated influences along the score distributions further interesting findings emerge (figure 2). First, a shift of students toward general schools in big cities would have a significant negative influence in the left side of score distribution in Slovenia. At the opposite, a marginal change of students toward vocational schools would have a significant negative influence especially from the sixth decile onwards in Germany (and Portugal if we look at vocational schools in big cities only). Moreover, influences on reading scores related to a shift of students are overall stable along the distribution in Italy, where swapping students from general to vocational schools would decrease all decile values of about 6-7 percentage points.

Figure 2. Influences along the distribution of students' performance in reading (with 95% CI)



Source: elaborations of the Authors on PISA 2018 data

In conclusion, our results show that the type and localization of schools play a relevant role in the distribution of students' performances in all countries analysed, and thus in the levels of inequality among students. In fact, despite a marginal change in the distribution of students may engender influences in scores which are different in terms of magnitude and direction (and by country), these influences tend to be statistically significant in most cases.

4. Conclusion

We apply the UQR method to explore the differences on students' performances by type and localization of schools in a cluster of European countries. Our analysis points out that these two dimensions have heterogeneous influences on the distribution of test scores across countries. In general, differences are particularly strong between general and vocational students in all countries analysed, and they are also significant between big and small cities in Portugal and Slovenia (especially in the left-tail of distribution). These findings are useful in terms of policies aimed to reduce the educational gap and to contrast the inequalities related to the school tracking and the type of municipality in which the school is located. Differences along test scores' distribution are probably related to the fact that students, with an educationally rich home environment, are likely to do relatively well in most school environments. In contrast, the performance of students from more educationally impoverished backgrounds may depend more heavily on school factors such as school localization or school type.

Appendix

Table A1. Full estimates of RIF regressions on the mean value and Gini index of students' performance in reading

Variables	Mean value				Gini index			
	DE	IT	PT	SI	DE	IT	PT	SI
General - Big city	0.584	-0.180	1.618***	-0.647	0.001	0.000	-0.001	0.001**
Vocational - Big city	-3.861***	-6.036***	-1.021	-7.781***	0.003**	0.002***	-0.001	0.002***
Vocational - Small city	-2.376**	-5.702***	-4.497***	-8.184***	0.000	0.002*	0.003***	0.002***
Age	2.744***	1.402***	1.597***	1.255***	-0.001*	-0.001*	-0.000	-0.000
Female	1.644***	0.360	1.893***	1.911***	-0.001***	-0.001**	-0.001***	-0.001***
Foreign speaking	-4.714***	-1.874***	-3.914***	-2.254***	0.004***	0.001***	0.004***	0.002***
Average occupational level	1.353***	1.677***	1.929***	0.223	-0.001**	-0.001*	-0.001***	-0.000
High occupational level	3.453***	1.817***	3.773***	1.479***	0.001**	-0.000	-0.001**	-0.000
11-25 books	2.648***	2.176***	1.505***	2.516***	-0.003***	-0.003***	-0.001***	-0.003***
26-100 books	5.636***	4.329***	4.483***	3.695***	-0.005***	-0.004***	-0.003***	-0.003***
101-200 books	7.476***	4.799***	6.024***	5.024***	-0.005***	-0.004***	-0.003***	-0.003***
201-500 books	9.213***	5.549***	6.867***	5.341***	-0.004***	-0.003***	-0.001***	-0.002***
500 books or more	9.845***	5.708***	6.576***	5.130***	-0.003***	-0.002**	-0.001	-0.000
16-20 students	1.887	3.076*	0.035	-0.357	-0.002	-0.001	-0.002	-0.000
21-25 students	5.504***	5.866***	2.004	0.726	-0.005***	-0.003	-0.004*	-0.001
26 students or more	7.894***	6.240***	4.299***	2.951***	-0.007***	-0.004	-0.005**	-0.002*
Public school	0.529	0.144	-1.441***	-0.357	0.000	0.001	0.001**	-0.001
No. of computers	-0.100	-0.664	-0.053	-0.280	-0.000	0.001	-0.000	0.000
Constant	-6.541	19.422***	15.591**	29.116***	0.035***	0.027***	0.018***	0.020***
Observations	3,761	10,631	5,316	5,395	3,761	10,631	5,316	5,395
R-squared	0.372	0.299	0.288	0.400	0.178	0.094	0.095	0.082

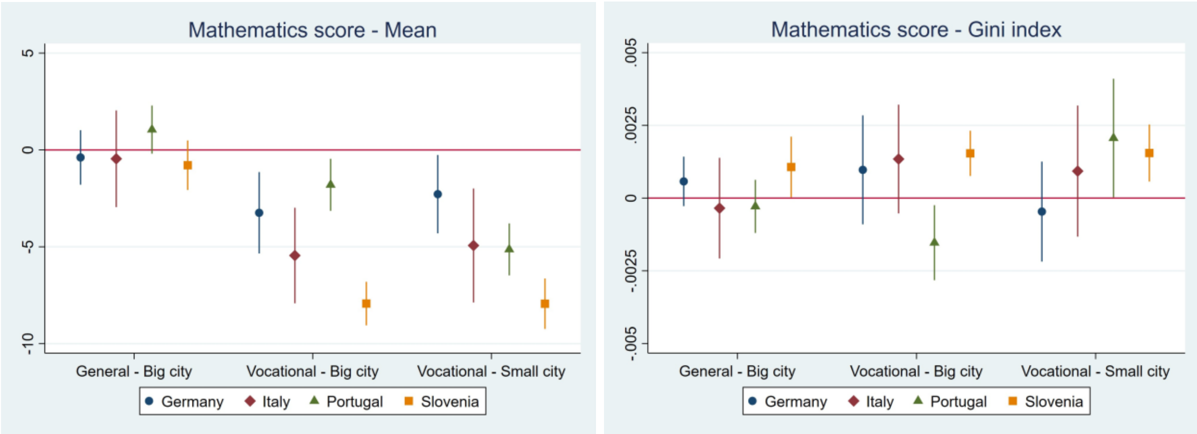
Source: elaborations of the Authors on PISA 2018 data

Table A2. Full estimates of RIF regressions on the mean value and Gini index of students' performance in mathematics

Variables	Mean value				Gini index			
	DE	IT	PT	SI	DE	IT	PT	SI
General - Big city	-0.387	-0.456	1.049*	-0.788	0.001	-0.000	-0.000	0.001**
Vocational - Big city	-3.244***	-5.449***	-1.800***	-7.932***	0.001	0.001	-0.002**	0.002***
Vocational - Small city	-2.282**	-4.932***	-5.135***	-7.937***	-0.000	0.001	0.002**	0.002***
Age	3.091***	1.410***	1.329***	0.584	-0.001*	-0.001***	0.000	-0.000
Female	-1.672***	-3.417***	-1.516***	-2.395***	0.000	-0.000	-0.001**	-0.000
Foreign speaking	-3.701***	-0.866**	-3.072***	-3.142***	0.003***	0.000	0.004***	0.003***
Average occupational level	1.376***	1.863***	2.077***	0.235	-0.001***	-0.001**	-0.001***	-0.001***
High occupational level	3.618***	2.289***	4.002***	1.873***	0.000	-0.000	-0.001***	-0.000
11-25 books	2.483***	1.429***	1.819***	1.968***	-0.003***	-0.002***	-0.002***	-0.002***
26-100 books	4.927***	3.766***	5.469***	3.030***	-0.004***	-0.004***	-0.004***	-0.003***
101-200 books	6.815***	4.998***	6.693***	4.482***	-0.004***	-0.003***	-0.003***	-0.002***
201-500 books	8.026***	5.199***	7.613***	5.193***	-0.004***	-0.002***	-0.002***	-0.001*
500 books or more	8.499***	4.969***	6.807***	4.184***	-0.003***	-0.002**	-0.002***	-0.001
16-20 students	3.204	3.425*	-0.154	0.362	-0.005**	-0.002	-0.001	-0.001
21-25 students	6.571***	6.400***	1.771	1.380	-0.006***	-0.004*	-0.002	-0.002
26 students or more	8.638***	6.904***	4.032***	3.537***	-0.008***	-0.004*	-0.004	-0.003**
Public school	-0.553	0.211	-2.313***	-0.185	0.001**	0.000	0.001	-0.001
No. of computers	0.050	-0.120	-0.147	-0.103	-0.000**	0.001	-0.000	0.000
Constant	-9.360	21.098***	22.352***	42.909***	0.030***	0.033***	0.014**	0.016***
Observations	3,761	10,631	5,316	5,395	3,761	10,631	5,316	5,395
R-squared	0.360	0.285	0.313	0.435	0.147	0.069	0.096	0.077

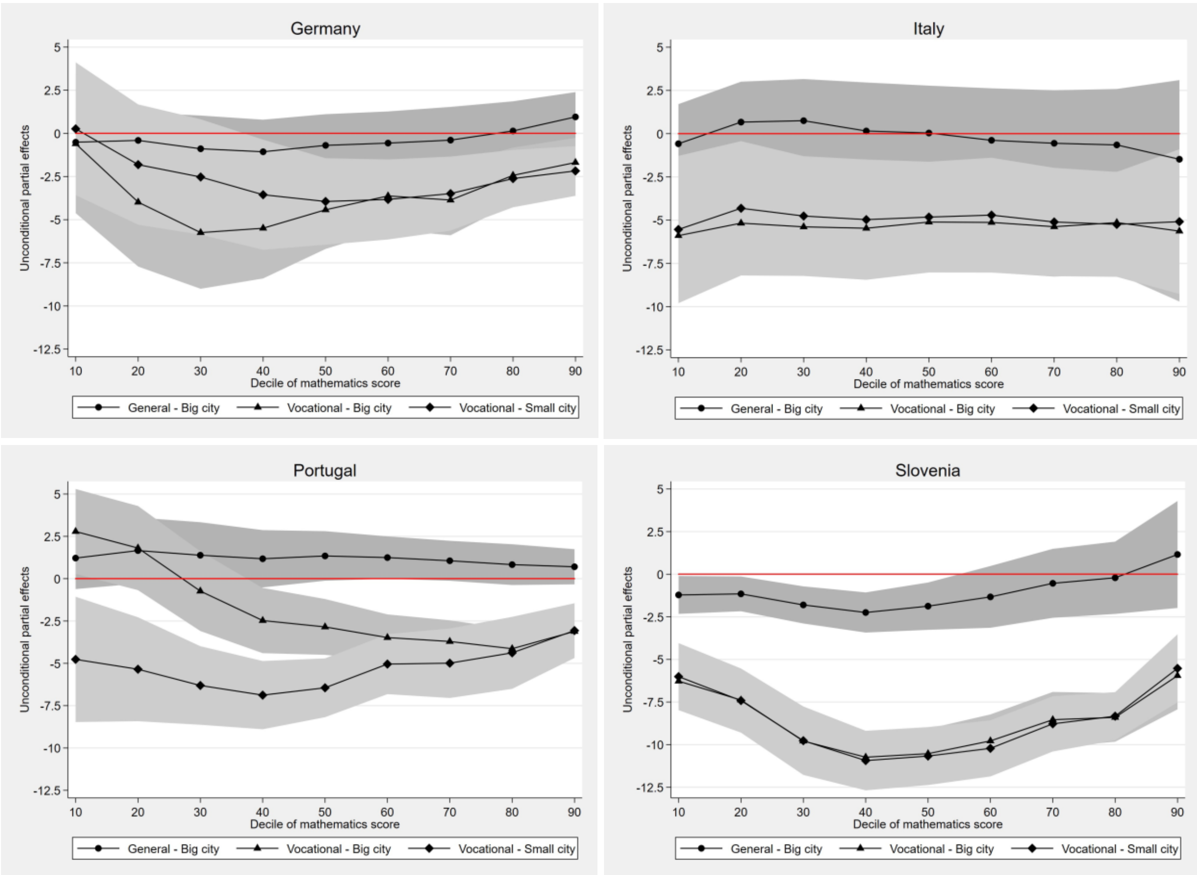
Source: elaborations of the Authors on PISA 2018 data

Figure A1. Influences on mean value and Gini index of students' performance in mathematics (with 95% CI)



Note: full estimates are provided in table A2.
Source: elaborations of the Authors on PISA 2018 data

Figure A2. Influences along the distribution of students' performance in mathematics (with 95% CI)



Source: elaborations of the Authors on PISA 2018 data

References

Arnold M., Newman J.H., Gaddy B.B., Dean C.B. (2005), A look at the condition of rural education research: Setting a direction for future research, *Journal of Research in Rural Education*, 20, n.6, pp.1-25

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, n.3, pp.299-333

Brunello G., Rocco L. (2015), *The effects of vocational education on adult skills and wages: What can we learn from PIAAC?*, OECD Social, Employment and Migration Working Papers n.168, Paris, OECD

Choe C., Van Kerm P. (2018), Foreign workers and the wage distribution: What does the influence function reveal?, *Econometrics*, 6, n.3, pp.1-26

Eide E., Showalter M.H. (1998), The effect of school quality on student performance: A quantile regression approach, *Economics Letters*, 58, n.3, pp.345-350

Firpo S., Fortin N.M., Lemieux T. (2009), Unconditional quantile regressions, *Econometrica*, 77, n.3, pp.953-973

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

Hampel F.R. (1974), The influence curve and its role in robust estimation, *Journal of the American Statistical Association*, 69, n.346, pp.383-393

Hanushek E. (1996), School resources and students performance, in Burtless G. (ed.), *Does Money Matter? The Effect of School Resources on Student Achievement and Adult Success*, Washington DC, Brookings Institution Press, pp.43-73

Hanushek E.A., Woessmann L. (2011), The Economics of International Differences in Educational Achievement, in Hanushek E.A., Machin S., Woessmann L. (eds.), *Handbook of the Economics of Education. Vol. 3*, Amsterdam, Elsevier, pp.89-200

Levin J. (2001), For whom the reductions count: A quantile regression analysis of class size and peer effects on scholastic achievement, *Empirical Economics*, 26, pp.221-246

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

Rangvid B.S. (2007), School composition effects in Denmark: Quantile regression evidence from PISA 2000, *Empirical Economics*, 33, pp.359-388

Wang T. (2021), Classroom Composition and Student Academic Achievement: The Impact of Peers' Parental Education, *The B.E. Journal of Economic Analysis & Policy*, 21, n.1, 273-305

ISTITUTO NAZIONALE PER L'ANALISI
DELLE POLITICHE PUBBLICHE