Educational Research for Policy and Practice (2022) 21:69–84
https://doi.org/10.1007/s10671-021-09292-y

ORIGINAL ARTICLE

Italian students' performance and regional decomposition

Marilena Furno[1]

Received: 14 July 2019 / Accepted: 4 January 2021 / Published online: 20 January 2021

Abstract
We relate students' math scores in the OECD-PISA test to school characteristics. The average math score for Italian students has been increasing in 2009. The determinants of this growth are analyzed by the Oaxaca–Blinder decomposition, that is particularly useful in comparing groups. The progress in educational attainments shows a different composition between northern and southern schools. In the North-Center regions, improvements are explained by school endowments, while in the South they are also driven by external factors that are not explained by the estimated model and are linked to improvements in students' attitude to education leading to a more favorable disciplinary climate. The regional gap decreases but does not disappear.

Keywords Regional difference · Students' performance · Decomposition

1 Introduction

There are many tests aiming to detect students' performance, like the US accountability tests, the British Cohort Study, or the Australian NAPLAN program. They provide data for educational monitoring as input to achieve reliable literacy rates and enables a sensible comparison across countries with a focus on resources allocation and policymaking of institutions and funding agencies (UNESCO 2006). Cross-national studies, by disclosing disparities in school resources, school organization, and teaching practices, encourage investments in education (Postlethwaite 2004).[1] Indeed, students' performance can be regarded as an indicator of economic growth, and schooling is important for boosting the rate of growth of the economy (Barro 1997). For instance, mathematical skills are considered a particularly good proxy for labor force quality. For instance, Hanushek (2006)

[1] Unfortunately, the cross-countries assessment programs have also prompted works interpreting discrepancies and below average scores as caused by lower IQ (Lynn and Vanhanen 2006; Kanazawa 2006a, b; Whetzel and McDanie, 2006). The lower IQ would be responsible for the national/regional differences in economic growth and for the gap in per capita income between rich and poor nations/regions. This highly debatable link between lower scores and IQ contradicts, among others, the findings by Barro (1997), Hanushek (2006), Daniele and Malanima (2011), Jamison et al. (2007), and Daniele (2015). All of them, among many others, stress the relevance of education as a key factor of economic prosperity. If below average scores were really driven by low IQ, any effort to improve proficiency would be useless.

✉ Marilena Furno
marfurno@unina.it

[1] Università Degli Studi di Napoli Federico II, Napoli, Italy

relates a one standard deviation difference in students' math performance to a 1 percent difference in annual per capita GDP growth rates.

There are, of course, mixed feelings about the effectiveness of those cross-countries assessment programs. According to Breakspear (2012), PISA methodologies have been embedded within national systems to evaluate school organization, and large-scale evaluation has been introduced for systems that did not have any previous national assessment. The PISA test has complemented national data and has promoted the comparison of national results against international benchmarks (Sellar and Lingard 2018). Mulongo and Amod (2019), for instance, analyzing Kenya, Tanzania, and South Africa policies, find that at least 11 policy/strategic documents formulated between the years 2000 and 2015 respond to recommendations or findings emanating from the cross-national learning assessments.

To the contrary, other studies criticize cross-countries assessment programs suggesting that, to emulate the PISA performance of the 'best education systems', important contextual and cultural differences existing within and between education systems are often ignored or sidelined (Harris and Jones 2017). In other words, the same strategy implemented in very different contexts results in highly variable outcomes and impact. Harris and Jones (2018), analyzing the school structure in seven different countries (Australia, England, Indonesia, Hong Kong, Malaysia, Russia, and Singapore), conclude that explanations of performance differentials need much more attention and acknowledgement of cultural and contextual influences: The context in which policy is implemented requires great consideration and understanding.

In addition, other authors are concerned about the occurrence of non-standard gradings approaches. They show variations across macro-geographical areas in the degree of strictness in teachers' standards. These studies find empirical evidence that teachers' school marks and students' performance in the standardized tests are not homogenous across geographical areas, with a focus on the North-Center versus South divide in Italy (Argentin and Triventi 2015; Battistin et al. 2017). Diverging grading standards provide differing information about the student's competencies and imply diverse opportunities between students who attend school in different regions.

In what follows, the scores of math test in the international OECD-PISA questionnaire for 15-year-old Italian students are analyzed. The OECD-PISA is a cross-national learning assessment program aiming to gain information on countries educational outcomes in support to policymakers (Knight et al. 2012). In this analysis, PISA test scores are related to explanatory variables describing school characteristics such as size, funding, student–teacher ratio, to link student performance and school effectiveness. Our focus is on the link between student performance and school environment in a setting characterized by a marked regional divide, with Italian Southern regions lagging.

We look at the evolution of this link both over time and across regions, to seek the determinants of scores improvements and to find the best devices to close the educational gap across Italian regions. The analysis particularly focuses on regional discrepancies by implementing a decomposition approach that provides fruitful insights wherever a divide occurs. Disparities between high/low developed regions, or between urban versus rural areas, or different school's tracks, can be fruitfully analyzed by decomposition. The focus of the Oaxaca–Blinder (1973) decomposition is to point out the source of the discrepancy. Is it due to a difference in the variables of the model across regions/areas/schools' organization, or is it due to external factors not comprised in the equation? In the former case, the discrepancy is explained by a difference in endowments, that is in the covariates of the model. When this is the case, the policy effort to close the gap attempts to increase endowments of one of the groups and assessing which factor has a wide impact on students' scores helps selecting the most appropriate intervention. Vice versa, a discrepancy due to external factors does not provide any potential factor of policy interest. The Italian data set provides a valid example on how to investigate within country discrepancies. The decomposition approach allows to identify which components could be changed in order to improve the general attainments.

2 Schools endowments

There is a growing interest among academics and policymakers on school effectiveness research. The central hypothesis of this research area postulates that certain characteristics of school impact student educational attainment, and this impact holds even after controlling for the students' socioeconomic, academic, and demographic characteristics (Goldstein and Woodhouse 2000; Opdenakker and Van Damme 2001). The school effect, once the socioeconomic level of the students has been controlled for, ranges between 10 and 30%; it is higher in mathematics than in either languages or science, and it is also higher in primary education than in secondary education. It appears that the most important source of school inefficiency is the resource endowment. However, there are mixed results on the impact of specific educational elements. Kruger (1999) in the STAR (Student–Teacher Achievement Ratio) experiment, where students are randomly assigned to classes differing in size, concludes that students of small size classes yield better performances, and later on, by adding college records (Kruger and Whitmore 2001), he finds that students from small classes are more likely to apply to college, particularly so for Afro-American. Vice versa, Hoxby (2000) does not find any significant effect of class size on student achievements, together with Woessman and West (2002) who analyze TIMMS (Third International Mathematics and Science Study) data. Class size is not the only one element of school endowments having an ambiguous impact on school attainment. Student–teacher ratio, library facilities, books availability, and funding are in turn positively or negatively related to student performance or statistically non-significant. In the literature there are, indeed, mixed findings so that rise in school inputs does not necessarily imply students improved scores. Barro and Lee (2001), looking at scholastic districts, find a positive impact of school endowments on student scores in internationally comparable tests. Card and Kruger (1992, 1996) find a positive impact of school resources as well, in terms of both (i) students staying longer at school and thus (ii) receiving afterward higher returns to education. They find that an average 10% increase in educational expenditure per student yields a 1–2% increase in future annual earnings. On the other hand, Hanushek (1996) finds a stable student performance against 3.5% annual increase in US educational expenditure, while Gundlack et al. (2001) in the OECD countries find a decline in student attainment, annually by a 2–4%. Interestingly, Fiske and Ladd (2000) take advantage of school reforms in New Zealand that shifts from a centralized to a decentralized school system occurred in 1991. They find an increased polarization of good students in the best schools, with the low performing schools generally located in deprived areas. In what follows, the links between school endowments and students score are analyzed, particularly focusing on regional discrepancy in the school system.

3 The PISA test and the Italian school

The OECD-PISA test of student proficiency was administered since year 2000. [2] The math scores of 15-year-old Italian students have been slightly above the OECD average score in 2009, 495 versus the OECD average of 491. In earlier years, the average Italian score was below average, although showing an improving trend over time, and it declined again below average in 2012. [2]

According to the OECD country note,[2]

> Italy's mean performance improved between 2003 and 2012 by an average of 20 score points, moving substantially closer to the OECD average… Most of the improvement in mathematics performance was observed between 2006 and 2009… Italy is one of the fastest improving countries in mathematics performance among those countries that participated in every PISA assessment since 2003… Italy shows above-OECD-average equity in education outcomes, with 10% of the variation in student performance in mathematics attributable to differences in student socio-economic status… While performance improved, equity remained stable… The improvement in mathematics performance is observed among all socio-economic groups: disadvantaged students improved by 27 score points and advantaged students by 17 score-points.

This suggests focusing on the changes occurred in 2009 to examine its main determinants. We relate the OECD-PISA test math score achieved by Italian students, *y*, to the explanatory variables describing school characteristics: field; school size; funding; student–teacher ratio; library facilities, number of computers, lab equipment, and textbooks. This allows to estimate the link between student performance and school environment.

As mentioned, growth analysis emphasizes school attainment and shows its strong links to differences in economic growth across countries. Mathematical skills are considered a particularly good proxy for labor force quality, and Hanushek (2006) relates a one standard deviation difference in students' math performance to a 1 percent difference in annual per capita GDP growth rates.

3.1 The data set

The estimated equation explains math test scores, *y*, as a function of funding represented by a dummy variable, *private,* which assumes unit value for private schools which in Italy are remedial and very few in number; school track which uses the dummy variables *academic* and *technical* to distinguish vocational fields; a dummy for gender which takes a unit value for *boys* who generally perform better than girls in math but worse in reading proficiency; *school size* a variable for the number of students enrolled; number of *computers* in the school; *student–teacher ratio*; *proportion of certified teachers* in the school; *shortage of teachers*; four categorical variables for *poor library* facilities, *shortage* of *computers*, *shortage* of *textbooks*, *shortage* of *lab equipment*. These four variables are coded as: a lot, some, a little, none at all. Each of them is transformed into a numerical variable assigning values ranging from 1 (none at all) to 4 (a lot).[3] The sample comprises $n = 52{,}067$

[2] http://www.oecd.org/pisa/keyfindings/PISA-2012-results-italy.pdf.

[3] The numerical conversion follows Likert (1932) who proposes a numerical scale in research that employs questionnaires.

Table 1 Summary statistics

	Sample	Mean	SD	Median	10th	25th	75th	90th	OECD mean	
Dependent variable: math scores										
2000	2113	469.56	89.760	474.34	341.86	418.08	531.99	581.15	500	
2003	10,640	498.51	88.577	501.74	381.86	440.05	560.96	610.56	500	
2006	16,279	481.90	90.475	483.21	364.53	420.38	544.78	597.36	497	
2009	23,819	495.03	84.438	497.05	384.18	436.83	555.38	604.15	491	
All	52,851	490.88	87.575	492.76	376.45	431.54	552.58	602.81		
$\Delta_{2009–00}$		25.47		22.71	42.32	18.75	23.39	23	−9	
	2000		2003		2006		2009		All	
	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD
Explanatory variables										
Sh. textbook	1.51	0.75	1.74	0.82	1.76	0.78	1.75	0.74	1.73	0.774
Sh. computer	1.95	0.96	1.94	0.89	1.77	0.83	1.92	0.88	1.88	0.879
Sh. library	1.98	0.97	1.99	0.88	1.92	0.88	2.21	0.94	2.06	0.923
Sh. lab equip	2.15	1.08	2.16	0.99	2.28	1.03	2.25	1.00	2.24	1.02
Computer	68.97	55.34	87.66	61.36	93.82	65.43	141.10	79.34	111.72	75.54
School size	753.27	375.1	632.35	360.1	698.58	413.22	681.12	369.69	683.12	383.22
Stud–teach ratio	9.22	2.30	9.28	4.87	8.86	2.68	9.12	2.77	9.083	3
Prop. cer. teacher	0.76	0.30	0.90	0.16	0.88	0.18	0.89	0.18	0.884	0.19
Teach. short	0.27	0.99	0.11	0.85	0.11	0.92	0.16	0.84	0.149	0.88

Table 1 (continued)

	2000	2003	2006	2009	All
	Mean	Mean	Mean	Mean	Mean
Dummy variables					
Boy	0.49	0.48	0.49	0.50	0.493
Vocational	0.25	0.24	0.22	0.20	0.219
Technical	0.38	0.36	0.35	0.32	0.340
Academic	0.37	0.39	0.43	0.47	0.438
South	0.42	0.27	0.42	0.36	0.368
Private	0.04	0.04	0.03	0.03	0.037

Table 2 OLS estimates nationwide

	2000–03–06		2009		2000–2009	
	Coeff	SE	Coeff	SE	Coeff	SE
Academic	90.09	1.47	84.37	1.62	89.68	1.061
Technical	45.26	1.33	57.09	1.37	50.35	0.963
Stud.–teach. ratio	1.97	0.121	3.61	0.255	1.99	0.099
Short. lab equip	−6.07	0.592	*0.336*	0.600	−3.67	0.429
Short. library	2.23	0.659	−4.64	0.652	*0.677*	0.467
Boys	23.73	0.983	19.35	1.02	22.64	0.719
Teacher short	2.08	0.498	5.561	0.622	3.295	0.379
Shortage computer	9.52	0.659	−2.38	0.685	5.637	0.476
Private	−32.18	1.69	−31.56	2.02	−32.64	1.28
Prop. cert. teacher	−6.41	2.26	*−1.07*	2.92	−3.58	1.73
School size	−0.012	0.002	−0.067	0.005	−0.017	0.001
Computer	0.262	0.011	0.343	0.022	0.224	0.007
Short. textbook	−14.26	0.749	−6.24	0.881	−12.58	0.561
Constant	397.17	2.91	414.13	3.65	402.84	2.19
Sample size	28,479		23,588		52,067	

In italics are the non-significant estimated coefficients

observations. The top section of Table 1 presents the summary statistics for the dependent variable: sample size, mean, standard deviation, quantiles, in each year of the data set.[4] In this sample, the mean and the median reach a peak in 2003 and then decline to rise again in 2009. The median is always larger than the mean, signaling left skewness. Over time, comparing the scores for 2000 and 2009, the 10th quantile increases more than the 90th: $y(0.10)_{2009}–y(0.10)_{2000}=42.32$, $\Delta y(0.10)_{2009–00}$ in the table, $y(0.50)_{2009}–y(0.50)_{2000}=22.71$, $\Delta y(0.50)_{2009–00}$ in the table, $y(0.90)_{2009}–y(0.90)_{2000}=23$, $\Delta y(0.90)_{2009–00}$ in the table. These values are reported in the last row of the top section of this table and reveal that low scoring students show the greatest improvement over time, which is in line with the OECD findings of above average equity in education in Italy.

The bottom sections of Table 1 collect the summary statistics for the explanatory variables. The curricula show a steady increase over time of students enrolled in the academic track while vocational and technical track decrease steadily. The variables measuring shortage in lab equipment, in textbook, and in library facilities do not show any improvement over time.

3.2 The empirical model

We estimate the model for the initial years 2000–2003–2006 in the first column, for 2009 in column 2, and for the entire period 2000–2009 in the last column of Table 2. The math score test y is explained by *academic, technical, boys, school size, number of computers, student–teacher ratio, proportion of certified teachers*, *private, shortage of library*

[4] Due to missing values in some of the variables of the selected model, the sample size and summary statistics reported in Table 1 do not necessarily coincide with their OECD analogues.

facilities, teacher shortage, shortage of computers, shortage of textbook, shortage of lab equipment.[5]

y=*f*(*academic, technical, boys*, *school size, number of computers, student–teacher ratio*, *proportion certified teacher, private, shortage of library facilities, teacher shortage, computer shortage, textbook shortage*, *lab equipment shortage*).

Almost all the estimated coefficients are statistically significant. *School size, proportion of certified teachers*, *private* schools together with the *poor facilities* coefficients (*shortage of textbook*, *shortage of lab equipment*) have a negative impact on students' scores. Comparison over time shows a sizable reduction in year 2009 of the math gender gap and an improvement in the *technical* track coefficient. S*hortage of lab equipment* and *proportion of certified teachers* are not statistically significant in 2009.

4 Oaxaca–Blinder decomposition

4.1 Decomposition over time

Based on the selected covariates of the math scores equation, the changes over time reported in Table 2 can be decomposed into covariates and coefficients effect by the Oaxaca (1973) and Blinder (1973) decomposition approach. Other authors have implemented a decomposition approach with PISA scores. For instance, Gigena et al. (2011) analyze the discrepancy in PISA reading scores between Chile, Mexico, and Argentina. They show how differently covariates and coefficients effects perform in the three countries. In Mexico, the decrease in covariates is balanced by the increase in coefficients. Chile shows a positive effect in both endowment and coefficients, while the opposite occurs for Argentina. Barrera-Osorio et al. (2011) consider PISA math scores in Indonesia, finding a key role for the improved performance in an adequate teacher supply.

We begin our analysis by comparing data for years 2000–2003–2006, that are indexed 0, with the data for year 2009 indexed 1, namely ($y_0 - y_1$). Consider the comparison between y_0 and y_1 on average, $\mathrm{E}(y_0 - y_1)$. By adding and subtracting the term $\mathrm{E}x_1\,\hat{\beta}_0$ and by rewriting $\mathrm{E}y_0=\mathrm{E}x_0\,\hat{\beta}_0$ and $\mathrm{E}y_1=\mathrm{E}\,x_1\,\hat{\beta}_1$, the average difference can be written as

$$\begin{aligned}\mathrm{E}(y_o - y_1) &= \mathrm{E}y_0 - \mathrm{E}x_1\hat{\beta}_0 + \mathrm{E}x_1\hat{\beta}_0 - \mathrm{E}y_1\\ &= \mathrm{E}x_0\hat{\beta}_0 - \mathrm{E}x_1\hat{\beta}_0 + \mathrm{E}x_1\hat{\beta}_0 - \mathrm{E}x_1\hat{\beta}_1\\ &= (\mathrm{E}x_0 - \mathrm{E}x_1)\hat{\beta}_0 + \mathrm{E}x_1(\hat{\beta}_0 - \hat{\beta}_1).\end{aligned}$$

The term $\mathrm{E}x_1\,\hat{\beta}_0$ is the so-called counterfactual, an unobservable term that computes how would y_1 be in case x_1 would have the same impact as group 0, impact measured by $\hat{\beta}_0$. Thus, the counterfactual $\mathrm{E}x_1\,\hat{\beta}_0$ measures the effect of the group 1 covariates, i.e., their recent value, evaluated at the group 0 estimated coefficients, i.e., their past coefficients. The decomposition breaks the average difference $y_0 - y_1$ into changes due to coefficients, $\mathrm{E}x_1\,(\hat{\beta}_0 - \hat{\beta}_1)$, and changes explained by the covariates, $(\mathrm{E}x_0 - \mathrm{E}x_1)\hat{\beta}_0$. A change in the covariates ($\mathrm{E}x_0 - \mathrm{E}x_1$) signals a variation in math scores explained by group differences, in this case explained by changes in schools' resources. This provides a potential factor of policy interest. Vice versa, a change in the estimated coefficients $(\hat{\beta}_0 - \hat{\beta}_1)$

[5] The estimates are computed using Stata version 15.

represents a change in students' performance that is not explained by the covariates of the equation and that is outside policy intervention. This unexplained part is often used as a measure for discrimination.[6] More in general, it computes the effects of group differences in unobserved predictors.

In addition, an interaction term, $(\mathrm{E}x_0 - \mathrm{E}x_1)\ (\hat{\beta}_0 - \hat{\beta}_1)$, can be added to account for simultaneous differences in endowments and coefficients:

$$\mathrm{E}y_o - \mathrm{E}y_1 = (\mathrm{E}x_0 - \mathrm{E}x_1)\hat{\beta}_0 + \mathrm{E}x_1(\hat{\beta}_0 - \hat{\beta}_1) + (\mathrm{E}x_0 - \mathrm{E}x_1)(\hat{\beta}_0 - \hat{\beta}_1).$$

The estimates of the above decomposition are reported in the first column of Table 4.[7] The results show an overall improvement in students' scores over time that is sizable and statistically significant: The difference $\mathrm{E}y_0 - \mathrm{E}y_1 = -20.08$ (se = 0.775) tells of a general significant improvement in year 2009. This improvement is due to endowments, with schools' characteristics accounting for the greatest part $(\mathrm{E}x_0 - \mathrm{E}x_1)\hat{\beta}_0 = -15.04$ (se = 1.02), while the increase in the coefficients effect is relatively small, $\mathrm{E}x_1\ (\hat{\beta}_0 - \hat{\beta}_1) = -4.69$ (se = 0.888). The interaction term is not statistically significant, with an increment of $(\mathrm{E}x_0 - \mathrm{E}x_1)\ (\hat{\beta}_0 - \hat{\beta}_1) = -0.343$ (se = 1.11). School characteristics improved in 2009, and the unexplained change due to the coefficients is comparatively small.

The first column of Table 5 shows for each explanatory variable the detailed contribution of each covariate and of each coefficient to the global improvement occurred in year 2009. *Academic*, *computer*, *boys*, *private* schools are among the endowments that contributed most to the increased performance in year 2009. The coefficients of *technical* schools, *student–teacher ratio*, *shortage of lab equipment*, *shortage of textbooks*, *shortage of teachers*, *computers* and *proportion of certified teachers* have all improved their coefficient effect in 2009 regardless of their endowment. The interaction term, that overall is not statistically significant, is significant for many of the variables of the model individually considered. At times, the interaction term reinforces the improvement signaled by covariates and coefficients effects, at times mitigates it.

4.2 Regional inequality

The OECD country notes [2] state that:

> In Italy more than half (51.7%) of the overall variation in student performance lies between schools: this means that two students who attend different schools can be expected to perform at very different levels. The comparatively large between-school variation in performance to an extent reflects the large regional differences in performance which can be observed in Italy, although large between-school differences can be observed even when regional differences are considered.

Therefore, we look at the regression explaining math scores in each macro region, North-Center and South. Table 3 collects the OLS estimates for the entire time period, 2000–2009. The North-Center regions perform generally better than the southern ones, as shown by the greater coefficients of *academic* and *technical* truck, *proportion of certified*

[6] This is often the case in men/women or white/non-white comparisons, but it is not applicable to the actual analysis.

[7] The decompositions are computed in Stata15. Jann (2005a, b, 2008) describes the implementation in Stata of the Oaxaca–Blinder decomposition. The Stata codes to carry out the decomposition can be downloaded at http://fmwww.bc.edu/repec/bocode/o/oaxaca.ado, http://fmwww.bc.edu/repec/bocode/o/oaxaca.hlp.

Table 3 OLS estimates within each region, time period 2000–2009

	South		North-Center	
	Coeff.	SE	Coeff.	SE
Academic	79.074	1.743	103.06	1.218
Technical	45.817	1.605	58.364	1.106
Stud.–teach. ratio	2.857	0.1396	1.3046	0.1462
Short. Lab equip	−2.881	0.7206	*0.2117*	0.487
Short. library	4.488	0.7440	−1.088	0.560
Boys	26.169	1.178	18.74	0.829
Teacher short	*−0.139*	0.5773	1.786	0.489
Shortage computer	8.635	0.7687	−1.640	0.561
Private	−48.97	2.476	−39.13	1.338
Prop. cert. teacher	7.459	2.897	10.719	1.988
School size	−0.006	0.0020	−0.006	0.002
Computer	0.1578	0.0118	0.115	0.009
Short. textbook	−10.16	0.8663	−3.38	0.716
Constant	344.48	3.641	412.69	2.635
Sample size	18,202		33,865	

In italics are the nonsignificant estimated coefficients

Table 4 Oaxaca–Blinder decompositions of math scores

	2000/03/06–2009		south over time		north over time	
	Nationwide		2000/03/06–2009		2000/03/06–2009	
	Coeff.	SE	Coeff.	SE	Coeff.	SE
Group 0	462.244	0.536	437.204	0.871	491.451	0.630
Group 1	482.326	0.560	457.349	0.927	499.764	0.670
Difference	−20.083	0.775	−20.145	1.27	−8.3132	0.920
Endowments	−15.045	1.02	−6.269	0.898	2.399	0.656
Coefficients	−4.694	0.888	56.760	1.05	39.789	1.07
Interaction	−0.*344*	1.11	3.755	0.876	0.225	0.605

In italics are the nonsignificant estimated coefficients

teachers, by the reduced gender gap in math scores, by the nonsignificance of *shortage of lab equipment* and by the milder impact of *private* schools and *shortage of textbook*.

A deeper analysis is offered by the regional Oaxaca–Blinder decomposition, comparing the time period 2000–2006, indexed 0, and year 2009, indexed 1, in the South and independently in the North-Center. Once again, the difference in math scores is decomposed into covariates, coefficients, and interaction effects by implementing the Oaxaca–Blinder decomposition separately in each region.[8]

[8] The North-Center/South inhomogeneity can be measured by a dummy variable assuming unit value for the South. The estimated coefficient measures the regional gap in the math scores. For instance, in year 2009 the coefficient of the dummy is −40.103 with a Student t of −7.47. However, the regional gap affects in a pervasive way the entire regression, i.e., also the other covariates like school track, school size, etc. This becomes clear if the regressions are separately computed in the two regions, but its pervasive effect is blurred if we consider the nationwide regression with a dummy variable for the regional gap.

Table 5 Detailed Oaxaca–Blinder decomposition of math scores

	2000/03/06–2009		South over time		North-Centre over time	
	Nationwide		2000/03/06–2009		2000/03/06–2009	
	Coeff.	SE	Coeff.	SE	Coeff.	SE
Endowments						
Academic	−3.362	0.367	−1.197	0.464	−6.38	0.568
Technical	1.457	0.244	−0.7041	0.351	4.478	0.371
Stud.–teach. ratio	1.040	0.134	3.632	0.437	−0.299	0.077
Short. lab. equip	*−0.031*	0.055	*0.0738*	0.130	−0.627	0.118
Short. library	1.392	0.198	0.882	0.431	0.604	0.217
Boys	−0.623	0.091	−0.706	0.146	−0.592	0.126
Teach. shortage	−0.173	0.048	−0.197	0.078	0.022	0.045
Short. computer	0.147	0.046	−0.684	0.167	0.272	0.073
Private	−0.762	0.099	0.092	0.114	−4.80	0.295
Prop. cert. teacher	*0.027*	0.073	*0.069*	0.047	1.18	0.242
School size	1.911	0.253	−0.337	0.178	4.63	0.467
Computer	−16.43	1.08	−5.32	1.83	−17.68	1.21
Short. textbook	0.368	0.067	−0.235	0.117	*0.217*	0.223
Coefficients						
Academic	2.31	0.883	8.150	1.50	*−1.258*	1.02
Technical	−4.19	0.678	−3.52	1.19	−4.568	0.765
Stud.–teach. ratio	−14.3	2.45	−15.27	4.05	−5.422	3.07
Short. lab. equip	−15.6	2.05	−15.20	3.80	−15.19	2.21
Short. library	15. 9	2.14	17.58	3.74	7.827	2.42
Boy	2.27	0.733	6.084	1.25	−2.681	0.831
Teach. shortage	−0.425	0.098	−0.458	0.124	−0.869	0.150
Short. computer	23.6	1.88	13.23	3.29	14.05	2.15
Private	*−0.061*	0.258	−7.14	0.531	1.741	0.294
Prop. cert. teacher	−4.77	3.29	−9.674	5.38	32.68	4.01
School size	1.73	2.98	10.75	4.93	38.40	3.59
Computer	−9.54	2.94	15.72	4.86	−32.12	3.50
Short. textbook	−14.7	2.12	4.356	3.50	−6.193	2.58
Constant	−16.96	4.67	−31.43	7.83	−22.53	5.65
Interaction						
Academic	−0.228	0.090	−0.388	0.165	*0.186*	0.152
Technical	−0.302	0.069	*0.138*	0.083	−0.877	0.162
Stud.–teach. ratio	−0.474	0.096	−1.33	0.369	*0.101*	0.061
Short. lab. equip	0.587	0.096	0.712	0.197	1.04	0.168
Short. library	−2.06	0.283	−2.75	0.592	−1.02	0.317
Boy	−0.141	0.049	−0.480	0.130	0.137	0.051
Teach. shortage	0.108	0.037	0.251	0.097	*−0.026*	0.054
Short. computer	−0.736	0.110	−0.835	0.223	−0.296	0.085
Private	*−0.015*	0.063	2.46	0.341	1.62	0.279
Prop. cert. teacher	0.*135*	0.093	*−0.053*	0.043	−2.37	0.301
School size	−1.58	0.229	*0.248*	0.153	−4.48	0.489

Table 5 (continued)

	2000/03/06–2009		South over time		North-Centre over time	
	Nationwide		2000/03/06–2009		2000/03/06–2009	
	Coeff.	SE	Coeff.	SE	Coeff.	SE
Computer	3.88	1.19	−6.70	2.07	12.07	1.33
Short. textbook	0.473	0.088	*0.059*	0.055	0.718	0.300

In italics are the nonsignificant estimated coefficients

$$[\mathrm{E}y_o - \mathrm{E}y_1]_\mathrm{i} = [(\mathrm{E}x_0 - \mathrm{E}x_1)\hat{\beta}_0 + \mathrm{E}x_1(\hat{\beta}_0 - \hat{\beta}_1) + (\mathrm{E}x_0 - \mathrm{E}x_1)(\hat{\beta}_0 - \hat{\beta}_1)]_\mathrm{i}\ \mathrm{i} = \text{North - Center, South.}$$

The impact of the covariates/coefficients/interaction terms for the southern regions over time is reported in Table 4 column 2. The third column of this table collects the results for the North-Center regions over time.

Both regions show a statistically significant improvement over time, estimated by $[\mathrm{E}y_0 - \mathrm{E}y_1]_{\mathrm{South}} = -\ 20.14$ (se = 1.27) in the South and by $[\mathrm{E}y_0 - \mathrm{E}y_1]_{\mathrm{North\text{-}Center}} = -\ 8.31$ (se = 0.920) in the North-Center. The overall improvement in the southern regions is quite sizable, more than double than the change occurred in the North-Center over time, moving toward a reduction in the regional gap in educational attainment. In addition, the composition of these changes is quite different. In the northern regions, the increase in math scores is exclusively due to the covariates effect, with an increase in endowment equal to $[(\mathrm{E}x_0 - \mathrm{E}x_1)\hat{\beta}_0]_{\mathrm{North\text{-}Center}} = -\ 18.99$ (se = 0.993), mostly counterbalanced by the worsening of coefficient and interaction effects, respectively $[\mathrm{E}x_1\ (\hat{\beta}_0 - \hat{\beta}_1)]_{\mathrm{North\text{-}Center}} = 3.86$ (se = 0.989) and $[(\mathrm{E}x_0 - \mathrm{E}x_1)\ (\hat{\beta}_0 - \hat{\beta}_1)]_{\mathrm{North\text{-}Center}} = 6.81$ (se = 1.06). In the southern regions, instead, the improvement is spread among covariates, coefficients, and interaction terms, and the estimates are all statistically significant, assuming, respectively, the values of $[(\mathrm{E}x_0 - \mathrm{E}x_1)\ \hat{\beta}_0]_{\mathrm{South}} = -\ 4.62$ (se = 2.08), $[\mathrm{E}x_1\ (\hat{\beta}_0 - \hat{\beta}_1)]_{\mathrm{South}} = -6.84$ (se = 1.61), and $[(\mathrm{E}x_0 - \mathrm{E}x_1)\ (\hat{\beta}_0 - \hat{\beta}_1)]_{\mathrm{South}} = -8.67$ (se = 2.33).

In sum, the regional decomposition provides different results for the North-Center and the southern regions. In the North-Center, the covariates, i.e., the school characteristics, show a significant improvement over time, which is largely curbed by the worsened impact of coefficients and interaction effects. In the South, the positive changes are driven by all the three terms: covariates, coefficients, and interaction effects. As mentioned, the coefficients effects refer to aspects beyond the schools' control. Among the many possible explanations for these results is students' attitude to education. The OECD country note [2] reports that:

> Between 2003 and 2012, the disciplinary climate in Italian schools improved significantly. In 2003, 39% of students reported that, in most or all lessons, the teacher has to wait a long time for students to quiet down; by 2012 that proportion had decreased to 31%. Similarly, in 2003, 42% of students reported that there is noise and disorder in most or all lessons. By 2012 this percentage had decreased to 36%.

The different behavior in the decomposition within regions confirms that both temporal and regional discrepancies should be analyzed.

The second and third columns of Table 5 detail the impact on the decomposition of each variable of the model. By this set of results, it is possible to find which covariate improves and which one worsens over time within each region and which variable provides a fruitful device to reduce the regional gap (or to achieve whatever other relevant target for policymakers). *Computer*, *teacher shortage*, *computer shortage* and *school size* are the endowments that most improved over time in the southern regions, while the *student–teacher ratio* sizably worsened. The former group of variables provides good tools to improve southern regions educational attainment and to reduce the educational gap across regions. The widest significant coefficients effect is for *technical* track, *student–teacher ratio*, and *proportion of certified teachers*, but these coefficients effects cannot be considered factors of policy interest since they are beyond school control.

In the North-Center group of schools, the greatest endowment increase is in *academic* track, *private* schools, and *computers*, but they are greatly offset by the sizable worsening in 2009 of the coefficient effects for *proportion of certified teachers* and *school size*.

Finally, the above results show that over time the regional gap has been reduced, although not completely closed, and a greater improvement in southern schools' characteristics would be very helpful.

4.3 Overall results

The analysis of the behavior over time and with respect to differing regional characteristics provides the following results:

(i) The average Italian math score has been growing over time: In 2000–2006, it was below the OECD average but has been growing between 2006 and 2009, as stated in the OECD country note: 'Most of the improvement in mathematics performance was observed between 2006 and 2009.'[9] Indeed in 2009, for the first time since the test started, the average Italian math score raised above the OECD average. The analysis of year 2009 attempts to define the source of the improvement, and the decomposition over time compares 2009 with the previous years. This decomposition shows that school covariates have improved greatly, while the coefficients show a not so wide increase. Thus, the increase in math scores is mostly explained by improvements in the school environment, while the effect of the coefficients, which gathers components of this change not explained by the variables in the model, is not too sizable.

(ii) However, within regions this behavior is quite different. The Italian economy is split between South and North-Center with the southern regions lagging (Di Caro 2014).[10] This economic inequality is reflected in the education system (see for instance Seta et al. 2014), with southern students scoring less than students enrolled in North-Center schools.[11] Therefore, student performance is decomposed not only over time

[9] Source OECD-PISA at http://www.oecd.org/pisa.

[10] With 36.2 thousand euros in 2018 (35.7 thousand in 2017), North-West has the largest annual nominal per capita GNP. Next is North-East with 35.1 thousand euros (34.3 thousand euros in 2017) and Center, with 31.6 thousand euros (31.1 thousand euros in 2017). Southern Italy value is 19 thousand euros (with 18.7 thousand euros in year 2017), that is almost half of the North-West value (source: ISTAT Conti Economici Territoriali).

[11] More than a regional IQ difference, the economic gap affects schools' attainment through the labor market. The unemployment rate in southern regions is particularly high for the young generations and discourages southern students. For instance, in 2009 the southern unemployment rate is 28.5% for young workers aged between 15 and 29 years, to be compared with the nationwide value of 18.5%. The lack of motivations accounts for a reduction in the test scores (source: ISTAT at http://dati.istat.it/Index.aspx?DataSetCode=DCCV_TAXDISOCCU).

but also to compare North-Center versus southern regions. While changes in scores over time in the North-Center are related to improvements in the explanatory variables, i.e., explained by improvements in school characteristics, in southern schools the recent improvements in math scores are due to changes in the covariates and even more in the coefficients. The coefficients include effects which modify the conditional distribution of student performance and which are not directly related to school characteristics. They may be linked, for instance, to an improved attitude among students toward education.

(iii) Unfortunately, the southern regions improvements do not close the regional gap.

These results suggest that, by complementing the southern coefficients effect with an improvement in southern schools' characteristics, the regional educational gap could be further reduced.

5 Conclusions

Among the many tests aiming to detect students' performance, like the US accountability tests, the British Cohort Study, or the Australian NAPLAN program, the PISA cross-national learning assessment program is the one here considered. The analysis focuses on Italian students' scores in math with its sizable enhancement in year 2009. The Italian peculiarity is a marked regional divide that makes interesting to investigate the source of the improvement. We focus on student performance linked to school characteristics rather than socioeconomic variables. The average math scores in the OECD-PISA test in Italy have been growing, particularly in year 2009, to overtake the OECD average. The determinants of this growth are here analyzed and decomposed into terms explained and unexplained by the schools' features selected in the estimated model, respectively, the covariate and coefficient effects. A change in covariates mirrors the change in the impact of school characteristics, while a change in the coefficients shows a variation that cannot be ascribed to school characteristics. The decomposition analysis is generally implemented to analyze groups discrepancy like regional divide, urban versus rural gap, or, for instance, academic versus technical schools' attainments. It allows to point out potential factors of policy interest to curb/enhance discrepancies. The analysis here implemented considers both temporal and regional changes, the latter linked to the Italian regional divide that is fully reflected in the educational attainments.

The temporal decomposition shows a large improvement mostly due to the covariates, that is to increased schools' endowment over time. The regional decomposition shows differing behaviors. Over time, the northern regions present a steady improvement in the covariates (endowment) largely curbed by the coefficients effect. The southern schools, instead, show an improvement in both covariates and coefficients. Thus, the rise in students' math scores is due to a higher impact of the school covariates nationwide, particularly sizable over time, and to improvements in the southern school coefficients, i.e., to changes in characteristics outside the schools' control.

While a change in the covariates is directly attributable to the school environment, the change in the coefficients is less straightforward. A possible explanation lies in variables such as students' attitude to education which has improved greatly in recent years. The latter is measured by variables such as the number of skipped classes that are generally beyond the school's control.

The different behavior in the decomposition within regions confirms that both temporal and regional discrepancies should be analyzed.

The combined regional and temporal analysis shows that the regional gap occurring in the 2000–2003–2006 period is reduced in 2009, but there is still a North-Center versus South educational gap.

The finding that southern student improvements in math scores are partly related to school characteristics leaves room for interventions. Improvements in southern school covariates could reduce the regional divide at least in the educational attainment.

Funding Open Access funding provided by Università degli Studi di Napoli Federico II.

References

Argentin, G., & Triventi, M. (2015). The north–south divide in school grading standards. *Italian Journal of Sociology of Education, 7*(2), 157–185.

Barrera-Osorio, F., Garcia-Moreno, V., Patrinos, H., & Porte, E. (2011) Using the Oaxaca–Blinder decomposition technique to analyze learning outcomes changes over time: An application to Indonesia's results in PISA mathematics. The World Bank, Policy Research Working Papers No. 5584, pp. 1–23.

Battistin, E., De Nadai, M., & Vuri, D. (2017). Counting rotten apples: Student achievement and score manipulation in Italian elementary Schools. *Journal of Econometrics, 200,* 344–362.

Barro, R. (1997). *Determinants of economic growth: A cross-country empirical study*. Cambridge: MIT Press.

Barro, R., & Lee, J. W. (2001). Schooling quality in a cross-section of countries. *Economica, 68,* 465–488.

Blinder, A. (1973). Wage discrimination: reduced form and structural estimates. *Journal of Human Resources, 8,* 436–455.

Breakspear, S. (2012). The policy impact of PISA: An exploration of the normative effects of international benchmarking in school system performance. *OECD working papers*, *71*, 1–31.

Card, D., & Krueger, A. (1992). Does school quality matter? Returns to education and characteristics of public schools in the United States. *Journal of Political Economy, 100*(1), 1–39.

Card, D., & Krueger, A. (1996). Labor market effects of school quality: theory and evidence. In G. Burtless (Ed.), *Does money matter? The effect of school resources on student achievement and adult success*. Washington: Brooking Institution Press.

Daniele, V. (2015). Two Italies? Genes, intelligence and the Italian North-South economic divide. *Intelligence, 49,* 44–56.

Daniele, V., & Malanima, P. (2011). Are people in the South less intelligent than in the North? IQ and the North-South disparity in Italy. *The Journal of Socio-Economics, 40,* 844–852.

Di Caro, P. (2014). Recessions, recoveries and regional resilience: Evidence on Italy. *Cambridge Journal of Regions, Economy and Society, 8,* 273–291.

Fiske, E., & Ladd, H. (2000). *When schools compete: A cautionary tale*. Washington: Brooking Institution Press.

Gigena, M., Vera, M., Giuliodori, R., & Gertel, H. (2011). Exploring the gap difference in 2000–2009 Pisa test scores between Argentina, Chile and Mexico. *Regional and Sectoral Economic Studies, 11–3,* 85–96.

Goldstein, H., & Woodhouse, G. (2000). School effectiveness research and educational policy? *Oxford Review of Education, 26*(3), 353–363.

Gundlack, E., Woessman, L., & Gmelin, J. (2001). The decline of school productivity in OECD countries. *Economic Journal, 111,* C135–C147.

Hanushek, E. (1996). Measuring investment in education. *Journal of Economic Perspectives, 10*(4), 9–30.

Hanushek, E. (2006). School resources. In E. Hanushek & F. Welch (Eds.), *Handbook of Economics of Education* (Vol. II). Amsterdam: North Holland.

Harris, A., & Jones, M. (2017). Leading educational change and improvement at scale: Some inconvenient truths about educational reform. *International Journal of Leadership in Education, 20,* 1–10.

Harris, A., & Jones, M. (2018). Why context matters: A comparative perspective on education reform and policy implementation. *Educational Research for Policy and Practice, 17,* 195–207.

Hoxby, C. (2000). The effects of class size on student achievement: New evidence from population variation. *Quarterly Journal of Economics, 115*(4), 1239–1285.

Jamison, E. A., Jamison, D. T., & Hanushek, E. A. (2007). The effects of education quality on mortality decline and income growth. *Economics of Education Review, 262,* 772–789.

Jann, B. (2005a). devcon: Stata module to apply the deviation contrast transform to estimation results. Boston College Department of Economics, Statistical Software Components S450603. Downloadable from http://ideas.repec.org/c/boc/bocode/s450603.html.

Jann, B. (2005b). Standard errors for the Blinder–Oaxaca decomposition. 2005 German Stata Users Group Meeting. http://repec.org/dsug2005/oaxaca_se_handout.pdf.

Jann, B. (2008). The Blinder–Oaxaca decomposition for linear regression models. *Stata Journal, 8,* 453–479.

Kanazawa, S. (2006). IQ and the wealth of states. *Intelligence, 34,* 593–600.

Kanazawa, S. (2006). Mind the gap…in intelligence: Re-examining the relationship between inequality and health. *British Journal of Health Psychology, 11,* 623–642.

Knight, P., Lietz, P., Nugroho, D., & Tobin, M. (2012). *The impact of national and international assessment programmes on educational policy, particularly policies regarding resource allocation and teaching and learning practices in developing countries*. London: EPPI-Centre, Social Science Research Unit, Institute of Education, University of London.

Kruger, A. (1999). Experimental estimates of education production functions. *Quarterly Journal of Economics, 114*(2), 497–532.

Kruger, A., & Whitmore, D. (2001). The effect of attending a small class in the early grades on college-test taking and middle school test results: Evidence from project STAR. *Economic Journal, 111,* 1–28.

Likert, R. (1932). Technique for the measure of attitudes. *Archives of Psychology*, *22*, 140–155.

Lynn, R., & Vanhanen, T. (2006). *IQ and global inequality*. Augusta: Washington Summit Publishers.

Mulongo, G., & Amod, Z. (2019). Nexus between participating in cross-national learning assessments and educational policy reforms: case study of Kenya. *Tanzania and SouthAfrica Educational Research for Policy and Practice, 18,* 203–223.

Oaxaca, R. (1973). Male–female wage differentials in urban labor markets. *International Economic Review, 14,* 693–709.

Opdenakker, M. C., & Van Damme, J. (2001). Relationship between school composition and characteristics of school process and their effect on mathematics achievement? *British Educational Research Journal, 27*(4), 407–432.

Postlethwaite, N. (2004). *Monitoring educational achievement*. Paris: The International Institute for Educational Planning, UNESCO.

Sellar, S., & Lingard, B. (2018). International large-scale assessments: Affective worlds and policy impacts in education. *International Journal of Qualitative Studies in Education, 31,* 367–381.

Seta, L., Pipitone, V., Gentile, M., & Allegra, M. (2014). A model to explain Italian regional differences in PISA 2009 outcomes. *Procedia-Social and Behavioral Sciences, 143,* 185–189.

UNESCO. (2006). Cross-national studies of the quality of education; planning their design and managing their impact. The International Institute for Education Planning.

Whetzel, D., & McDaniel, M. A. (2006). Prediction of national wealth. *Intelligence, 34,* 449–458.

Woessman, L., & West, M. (2002). Class size effects in school systems around the world: Evidence from between-grade variations in TIMSS. IZA discussion paper 485, Bonn.