MOX–Report No. 07/2014

Heterogeneity, school-effects and achievement gaps across Italian regions: further evidence from statistical modeling

Agasisti, T.; Ieva, F.; Paganoni, A.M.

MOX, Dipartimento di Matematica "F. Brioschi"
Politecnico di Milano, Via Bonardi 9 - 20133 Milano (Italy)

mox@mate.polimi.it http://mox.polimi.it

Heterogeneity, school-effects and achievement gaps across Italian regions: further evidence from statistical modeling

Tommaso Agasisti*, Francesca Ieva♯, Anna Maria Paganoni♮

January 31, 2014

* DIG - Department of Management, Economics and Industrial Engineering, Politecnico di MIlano, via Lambruschini 4/b, Milano, Italy
tommaso.agasisti@polimi.it
♯ Department of Mathematics "Federigo Enriques",
Università Statale di Milano, via Saldini 50, Milano, Italy
francesca.ieva@unimi.it
♮ MOX– Modelling and Scientific Computing, Department of Mathematics, Politecnico di MIlano, via Bonardi 9, Milano, Italy
anna.paganoni@polimi.it

Keywords: Child development; Multilevel models; School effectiveness; Value-added model; Contextual effects; Education

Abstract

Catching the differences in educational attainments between groups of students and across schools is becoming increasingly interesting. With the aim of assessing the extent of these differences in the context of Italian educational system, the paper applies multilevel modeling to a new administrative dataset, containing detailed information for more than $500,000$ students at grade 6 in the year 2011/12, provided by the Italian Institute for the Evaluation of Educational System. The results show that the national averages hide considerable heterogeneity both within and between schools, and that it is possible to estimate statistically significant "school effects", i.e. the positive/negative impact of attending a specific school on the student's test score, after a case-mix adjustment. Therefore, the paper's most important message is that school effects are different in terms of magnitude and types in the three geographical macro-areas (Northern, Central and Southern Italy) and are dependent upon specific students' characteristics.

1 Introduction and Motivation

The institutional organization of the Italian educational system is based on strong assumptions about its equality purposes, among which a key role is assigned to the presumption that all schools provide similar educational standards. Commonly, families perceive that quality of different schools is quite homogenous, especially at lower levels, e.g. primary and junior secondary. Therefore, recent aggregate data provided by the Italian Institute for the Evaluation of Educational System (hereafter, Invalsi) show that it is not the case, and that a significant portion of variance in students' test scores is attributable to structural between-schools differences. For instance, the Figure 1 illustrates some Invalsi's estimates for 2012-2013 (junior secondary schools), arguing that between-schools differences account for 15.3% and 18.3% of reading and mathematics test scores, respectively. Put simply: even in the Italian "egalitarian" educational system, school matters - and attending the school A instead than the school B can have a strong effect on achievement.

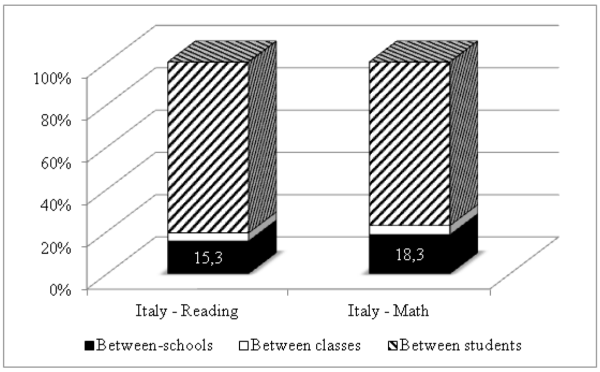

Figure 1: Differences in reading and mathematics test scores accounted by schools, classes and students. Invalsi's estimates, 2012-2013 (junior secondary schools).

This evidence is accompanied by a specific feature of the Italian educational system, namely a strong difference in educational attainment and results in different geographical macro-areas [2], with students in Northern Italy obtaining (on average, and all else equal) higher scores than their counterparts in Central and Southern part of the country. While the determinants of this gap are still not completely clear - even though some authors also propose possible explanations based, for example, on different resources and/or social capital, see [6] - the empirical evidence illustrates that also between-schools differences are stronger in the South than in the North[1]. In this perspective, a study of school effects on achievement for the different areas of the country seems worthy of specific attention. Additionally, it is important to investigate the relationship between achievement and variables measuring students and schools' characteristics, to-

[1]In a recent paper, [1] also showed that schools in the South practice within-school segmentation (e.g. between classes) than their counterparts in the North. In this paper, we explore between-schools differences, but we are aware that similar mechanisms (that is to say, differential effects on achievement between classes of the same school) are also operating within schools.

gether with estimates of the relative weights of the two levels of grouping (classes and schools). It is also likely that the various elements contribute differently to students' achievement in the different geographical areas. All these degrees of heterogeneity should be kept into consideration adequately when studying the determinants of Italian students' academic results. This study is inserted into a stream of the applied statistics literature, which uses multilevel models to investigate the relative impact of different sets of observable variables on students' achievement. Some studies used these methods in measuring specific phenomena, such as the differences between performances of native and immigrant students (see, for instance, [20] or the role of school resources [24]). The approach of focusing on a single, particular problem is also common in the economics of education literature, and especially in that branch devoted to the (empirically-based) policy evaluation.

Nonetheless, recent interesting statistical studies aim at describing the relationship between various sets of different individual level and school level factors and students' results [22]. The present work applies a multilevel model for estimating the school impact on student achievement; we define the "impact" as the effect exerted by attending a specific school on its students' achievement (the latter measured through the scores in the standardized test administered by Invalsi), after adjusting for student's characteristics.

This paper is innovative for several reasons. The first is that it is among the first attempts of using administrative Invalsi data (and, more generally, Italian data) in a value-added fashion. While other contributes employed Value Added Models (VAMs), they sometimes refer to very small experiences, involving few schools and students (see [9], [10]) or worked with data at school level (see [8]). To the best of our knowledge, the first paper using extensively the whole Invalsi dataset with a VAM approach is that of [1], and the present one is the first going into this direction through multilevel statistical modeling. A second aspect of innovation refers to the investigation of the "school effect", that is to say the positive or negative independent effect associated with the attendance of a specific school. This objective, which is typically related to the problems of evaluation, led to use a set of methods, which must isolate the effect of the school from other confounding factors. Typically, these studies then result in defining rankings of schools, albeit this activity is questionable on the methodological side for many reasons (see [16]). Therefore, in our paper we try to characterize the variables that are associated to positive/negative school effects. Operationally, we regress the school effects (obtained through the estimates of school-level random effects from a multilevel model) against a set of explanatory variables, which describe some important factors of schools themselves. Lastly, the very fact that the paper presents the results about the determinants of students' results for Italy is innovative per se; being the bulk of the existent literature is about English schools (see, among others, [11], [12], [14] and references therein). Few papers still exist about the Italian case (notable exceptions are [2] and [6]).

To anticipate the main results of our paper, the most important messages contained in it are two. First, the paper empirically demonstrates that the differences in the determinants of student achievement in the three macro-areas of the country (Northern, Central and Southern Italy) are so profound that it is impossible to specify a single empirical model for investigating them; as a consequence, this study promotes the idea of using three different models, one for each area. Even if the method and the list of variables used in the three cases are identical, the extremely different estimates and statistical significance of parameters obtained confirm the necessity to treat the three realities as structurally different. Obviously, this finding has clear and relevant policy implications, which are discussed in detail - especially, on the ground of equity and achievement gaps between geographical areas. The other major message from this paper is that the so-called "school effect" is actually very heterogeneous, in other words it is very dependent upon specific students and schools' characteristics. With the aim of providing evidence of such heterogeneity, we employ a graphically based method that highlights how the school effect is stronger/weaker for specific types of students' profiles. We believe that describing the school effects diversity is useful for policy purposes, as it reduce the emphasis on the "average" effects, and instead stimulates policy makers and school administrators to look at specific circumstances that can facilitate or impede the influence of schools on students' experiences and results.

The paper is organised as follows. The next section presents the empirical model (§2.1) and the list of variables contained in the Invalsi dataset (§2.2), while Section 3 contains the main results. Section 4 discusses the main policy and managerial implications, and concludes. All the analyses are carried out using `R` [21] statistical software.

2 Models, Methods and Data

In this section we formulate a general mixed effect model for pupils attainment. Pupils (level 1) are nested within schools (level 2). We consider only variables at student level with a random effect on schools, then we fit a linear model using only school level information in order to explain the observed distribution of the random effects. Our general model decomposes the total variability in pupils test scores at secondary level into the parts that vary between schools and pupils (see [13], [14]). The purpose of such modeling strategy is to describe each of the conditional associations between the pupil and school contexts on attainment and how these vary by stratification groups of interest.

2.1 School influence model

First we introduce the traditional two-level school effectiveness model which provides value added estimates of secondary school performance. So consider the simple variance components model for pupils Corrected Math Scores (CMSs), where we treat pupils (level 1) as nested within schools (level 2). The model, for $i = 1, \dots, n_j$ and $j = 1, \dots, J$ can be written as:

$$y_{ij} = \beta_0 + \sum_{k=1}^{K} \beta_k x_{kij} + b_j + \epsilon_{ij} \tag{1}$$

$$b_j \sim \mathcal{N}(0, \sigma_b^2) \quad \epsilon_{ij} \sim \mathcal{N}(0, \sigma_\epsilon^2) \tag{2}$$

where y_{ij} is the CMS for the $i-th$ pupil within the $j-th$ school, x_{kij} is the corresponding value of the $k-th$ predictor variable at student's level, $\beta = (\beta_0, \dots, \beta_K)$ is the $(K+1)$ dimensional vector of parameters to be estimated and ϵ_{ij} is the zero mean gaussian error. The random effect b_j for the $j-th$ school is assumed to be Gaussian distributed and independent of any predictor variables that are included in the model[2]. Estimates of school effects are derived from Maximum Likelihood (ML) optimization tools, provided in `lme4` [4] and `nlme` [19] R-packages.

In a second stage, we use variables describing schools characteristics to model the factors affecting the estimated random effects:

$$\hat{b}_j = \gamma_0 + \sum_{l=1}^{L} \gamma_l z_{lj} + \eta_j \tag{3}$$

$$\eta_j \sim \mathcal{N}(0, \sigma_\eta^2) \tag{4}$$

being z_{lj} the value of the $l-th$ predictor variable at school's level, $\gamma = (\gamma_0, \dots, \gamma_L)$ is the $(L+1)$ dimensional vector of parameters and η_j the zero mean gaussian error.

Once a mixed effect model has been fitted, a deeper study of the estimated random effects distribution ($\hat{b}_j$, for $j = 1, ..., J$) is needed in order to explain the overdispersion due to the grouped nature of data. For example, when the parametric assumptions are too restrictive and questionable, agglomerative algorithms (i.e., k-means, PAM,...) on the estimated values of random effect can be implemented to detect clustering structure and establish how many clusters of institutions, in terms of suitable similarity indexes, might exist (see [25] and [15], among others). Another case is when the possible correlation between $\hat{b}_j$ and group's-level covariates is a proxy of the causal relationship between observed phenomena. In the case of interest, the latter situation seems be suitable to happen, so we model this dependence not only to explain the observed variability, but also to quantify the contribution of different covariates, possibly accomodating for stratifications of interest. The use of this model is strongly

[2]While the distribution of b_j is checked later, its characteristics of being exogenous (in the sense of [24]) is not empirically verifiable in this setting.

legitimated when, like in this case, there is a statistical evidence for normality of the distribution of the random effects estimates (see below, Section 3). The great advantage is that when we come to analyze the results of fitted model we can interpret the influence of each significant school's level covariate in generating the random effects estimates.

2.2 Data

The Math Score Invalsi database collects the achievement in math tests of pupils attending the first year of junior secondary school[3]. Several information are provided at pupil, class and school level. A complete description of these variables is reported in Table 1.

[3]Since 2007-2008, Invalsi administers standardized test scores at various stages of the pupils' educational career (grades 2, 5, 6, 8 and 10). Unfortunately, until now, nor the tests are statistically anchored, neither a full retention of pupils' identity is guaranteed. As a consequence, longitudinal approach is prevented, even though this paper uses some features of the dataset to provide VAM estimates.

Variables of the Invalsi Database

Level	Type	Variable Name	Mean	Missing
Student	–	Student ID	—	0%
	(Y/N)	Female	48.93%	0.15%
	(Y/N)	1^{st} generation immigrant	6%	8.11%
	(Y/N)	2^{nd} generation immigrant	5%	8.11%
		ESCS - Socioeconomic background indicator	0.14	8.29%
	(Y/N)	Early-enrolled student	2.16%	0.15%
	(Y/N)	Late-enrolled student	7.29%	0.15%
	(Y/N)	Student who does NOT live with both parents	13.8%	3.63%
	(Y/N)	Student who has siblings	84.2%	3.54%
	%	Cheating	0.038	0.01%
Class	–	Class ID	—	0%
		Mean Number of Students	23	0%
		Mean ESCS	0.14	8.05%
	%	Female percentage	43.56	0.02%
	%	1^{st} generation immigrant percentage	4.86	7.95%
	%	2^{nd} generation immigrant percentage	4.35	7.95%
	%	Early-enrolled student percentage	1.96	0.02%
	%	Late-enrolled student percentage	6.38	0.02%
	%	Disabled percentage	5.42	0%
	count	Number of students	23	0%
	(Y/N)	*Tempo Pieno**	2.40%	7.95%
School		School ID	—	0%
		Mean ESCS	0.13	5.83%
	%	Female percentage	43.23	0%
	num	Number of Classes	6.5	0%
	%	1^{st} generation immigrant percentage	4.73	5.88%
	%	2^{nd} generation immigrant percentage	4.23	5.88%
	%	Early-enrolled student percentage	1.93	0%
	%	Late-enrolled student percentage	6.36	0%
	count	Number of Students	144	0%
	count	Average number of students per class	23	0%
	(Y/N)	NW - North West	25.31%	0%
	(Y/N)	C - Center	17.71%	0%
	(Y/N)	S - South	39.04%	0%
	(Y/N)	Private	2.91%	0%
	(Y/N)	*Istituto Comprensivo***	60.64%	3.63%
Outcomes	[0-100]	MS - Math Score	46.16	0%
	[0-100]	CMS - Math Score Corrected for cheating	44.93	0.01%
	[0-100]	CMS5 - 5^{th} year of Primary School Math Score	70.29	46.48%

* Indicates whether the class timetable ends at lunchtime or afternoon activities are scheduled.

** Indicates if both primary and junior secondary schools are present in the same building, under the same direction.

Table 1: List of variables contained in the Invalsi Database

When considering characteristics referred to the single student, the following information is available: gender, immigrant status (Italian, first generation or second generation immigrant), if the student is early-enrolled (i.e. was enrolled for the first time when five years-old, the norm being to start the school when six years-old), or if the student is late-enrolled - this is the case when the student must repeat one grade, or if he/she is admitted at school one year later if immigrant. Also, the dataset contains information about the family's background: if the student lives or not with both parents (i.e. the parents are not died, or are separated/divorced), and if the student has siblings or not. Lastly, Invalsi collects information about the socioeconomic status of the student, by deriving an indicator (called ESCS - Economic and Social Cultural Status), which is built in accordance to the one proposed in the OECD-PISA framework (see [17] and [18]), in other words by considering (i) parents' occupation and educational titles, and (ii) the possession of certain goods at home (for instance, the number of books). Once measured, this indicator has been standardized, i.e., it has mean = 0 and variance = 1. The minimum and maximum observed values in the Invalsi dataset are -3.11 and 2.67. The $5-$th and $95-$th empirical percentiles are equal to -1.53 and 1.90, respectively. In general, pupils with ESCS equal to or greater than 2 are very socially and culturally advantaged (high family's socioeconomic background). The dataset also allows to explore several characteristics at class level, among which the class-level average of several individuals' characteristics (for example: class-average ESCS, the proportion of immigrant students, etc.). Of particular importance, there is a dummy for schools that use a particular schedule for lessons ("*Tempo Pieno*" classes comprise educational activities in the afternoon, and no lessons on Saturday, while traditional classes end at lunchtime, from Monday to Saturday). Also the variables at school level measure some school-average characteristics of students, such as the proportion of immigrants, early and late-enrolled students, etc. Two dummies are included to distinguish (i) private schools from public ones, and (ii) "*Istituti Comprensivi*" which are schools that include both primary and lower-secondary schools in the same building/structure. This last variable is relevant to understand if the "continuity" of the same educational environment affects (positively or negatively) students results. Some variables about dimension (number of students per class, average size of classes, number of students of the school) are also included to take size effects into account. Lastly, we include two dummies for schools located in Central and Southern Italy; some previous literature, indeed, pointed at demonstrating that students attending the schools located in Northern Italy tend to have higher achievement scores than their counterparts in other regions, all else equal [2].

The outputs (MS, i.e., the score in the Mathematics standardized test administered by Invalsi) are expressed as "cheating-corrected" scores (CMS): Invalsi estimates the propensity-to-cheating as a percentage, based on the variability of intra-class percentage of correct answers, modes of wrong answers, etc.; the resulting estimates are used to "deflate" the raw scores in the test[4]. Among data, there is also the score in the Math test at grade 5, which is used as a control in the multilevel model to specify a Value-Added estimate of the school's fixed effect. It is well known from the literature that education is a cumulative process, where achievement in the period t exerts an effect on results of the period $t+1$. The empirical analysis can be then better conducted in a Value-Added (VA) fashion, namely considering the role of variables statistically correlated with test scores. In a cross-section setting, like the one of this paper, it is then important to include prior achievement among independent variables; in this case, we have information about the test score of the $i-th$ individual in the prior academic year and we use it when estimating the model parameters presented in the Section 2.1. However, the procedure of matching individual data longitudinally at student level is new in Italy, and the Ministry of Education and Invalsi are still improving it - the main problems are related to the transmission of information from schools to the Ministry. Unfortunately, for the year under scrutiny this procedure led to the loss of around half of the observations (precisely, 46.5%).

The database consists of $509,360$ records, within $25,922$ classes and $5,311$ schools. They represent the entire population of children from the first year of junior secondary schools in Italy. If we consider only statistical units with no missing information, the database reduces to $259,757$ records, within $18,761$ classes and $4,119$ schools. We will discuss representativeness of such subsample in Section 2.3, together with motivations, pros and cons of focusing on it for the analyses. In what follows, we will refer to the subsample as the *reduced* database.

2.3 Representativeness of the subsample

As mentioned before, in the *reduced* database we mainly discard all the statistical units for which the 5^{th} year of Primary School Math Score (CMS5) is missing. It is worth to evaluate from a statistical point of view the representativeness of this subsample with respect to the entire population. Since, the sample size at

[4]The Invalsi dataset reports an achievement score obtained in the standardized test as a percentage of correct answers, on a scale $[0;100]$. Nevetheless, the first exercise conducted in 2007-2008 revealed that many schools did act strategically, with (i) teachers suggesting the right answers and (ii) allowing the students to collaborate on answering the questions. Such behaviour ("cheating") is still documented in some schools (see [5]) even if year after year the proportion of schools that adopt it is steadily decreasing. Since then, Invalsi adopts a statistically-based model - run at class level - that allows estimating a "propensity to cheating index", based on the distribution of correct answers, missing answers, etc. and the modalities of wrong answers. The profile of answers of each class is indeed compared with a benchmark (standard) as obtained by the answers provided by a statistically representative sample of classes and students where Invalsi conducts the test in a controlled setting (i.e. with external examiners, etc.). The propensity to cheating index is then used to correct individual students' raw scores, and in this paper we use this indicator of output which is "net of (estimated) cheating", called here Corrected Math Score (CMS).

pupils level is extremely high, is quite impossible to find a non-significative difference in statistics summarizing the student's level features. Moreover the CMS5 score is a data transmitted to Invalsi at the school level. For these reasons we checked the representativeness of the subsample studying the distributions of the school's-level variables. For the continuous ones we performed non parametric comparison test (Wilcoxon test) to detect possibly differences in the stochastic distributions generating data. For the dichotomic ones we performed a comparison between proportions.

In particular the mean ESCS in the subsample seems to be slightly higher that in the population (p-value = 0.0004). There is no strong statistical evidence for difference with respect to female percentage (p-value = 0.1259), 1^{st} generation immigrant percentage (p-value = 0.0329), 2^{nd} generation immigrant percentage (p-value = 0.0199), early-enrolled student percentage (p-value = 0.2925) and late-enrolled student percentage (p-value = 0.7271). The number of students in the schools and the proportion of *Istituti Comprensivi* of the subsample are greater than the correspondents in the population (p-value = 0.0025, p-value = 0.0008, respectively), maybe reflecting the fact that it is easier for schools that share administrative offices to transmit complete and coherent information about the same pupils over time and across grades. There is no statistical evidence for difference in the proportion of private schools (p-value = 0.4467). Focusing on the geographic distribution there is a high statistical evidence that the South area is under-represented (p-value = $6.687 * 10^{-5}$). While the magnitude of the phenomenon is not worrying, it must be kept in mind when assessing the overall picture from results. Nevertheless, it is the case that schools in the South not only tend to have lower performances than those in the North, but also have less ability to transmit administrative information to Invalsi and to the Ministry of Education for allowing longitudinal comparisons. Being this behavior voluntarily-driven or not, it represents a problem for the evaluation of the national educational system. Overall, the *reduced* dataset used in this paper is substantially representative of the original population, with the only exception of the proportion of schools in the South, even if the magnitude of the bias is definitely low. Albeit the use of the *reduced* sample can be criticized on this ground, it has two major advantages that justify our choice. First, the performance at grade 5 is strongly predictive of the test score at grade 6, and the dismissal of such important control can generate a problem of omitted variables that is statistically more serious than the problem of macroareas' representativeness. Second, as anticipated in the Section 2.2, the inclusion of prior achievement allows considering the analysis in a VA fashion; as a consequence, the effects of attending a specific $j-$th school is not estimated through a simple cross-sectional variation, but also with reference to a longitudinal variation in relative test scores.

3 Results

In what follows, we will consider the *reduced* database defined above, containing also the information about the achievement of the 5^{th} year of the primary school.

3.1 Descriptives

The output of interest in our analyses is the Math Score (corrected for cheating, namely CMS) of students attending the first year of junior secondary school. It is a normalized score ranging from 0 to 100, with median equal to 46.94, first and third quantile respectively equal to 4.56 and 47.71. The mean value (std.dev.) is equal to 61.05 (17.74). The histogram reporting the distribution of the CMS is shown in Figure 2.

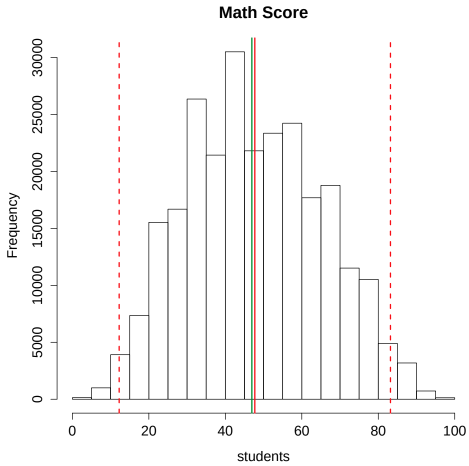

Figure 2: Histogram of the Corrected Math Score (CMS) of pupils in the *reduced* Invalsi database. The red solid line refers to the mean, the green one to the median. Dashed red lines are located 2 standard deviations away from the mean.

As we mentioned before, we would like to model CMS by means of suitable student's level variables, accounting for the school effect. Then explanations will be given for the random effect estimates using school level variables.

Due to the high number of variables available into the Invalsi database, and in order to identify the most relevant ones, an explorative analysis has been conducted. In particular, Wilcoxon tests for comparing CMS distributions stratified by different covariates (Gender, being a 1st or 2nd generation immigrant, being early or late enrolled, not living with both parents, having siblings) have been performed, showing in every case p-values lower than $4.17 \cdot 10^{-10}$. The boxplots corresponding to the most significant associations are reported in Figure 3.
It can be noticed how being a female student has a negative correlation with the CMS, as well as being late enrolled. Moreover, there is also a significant differ-

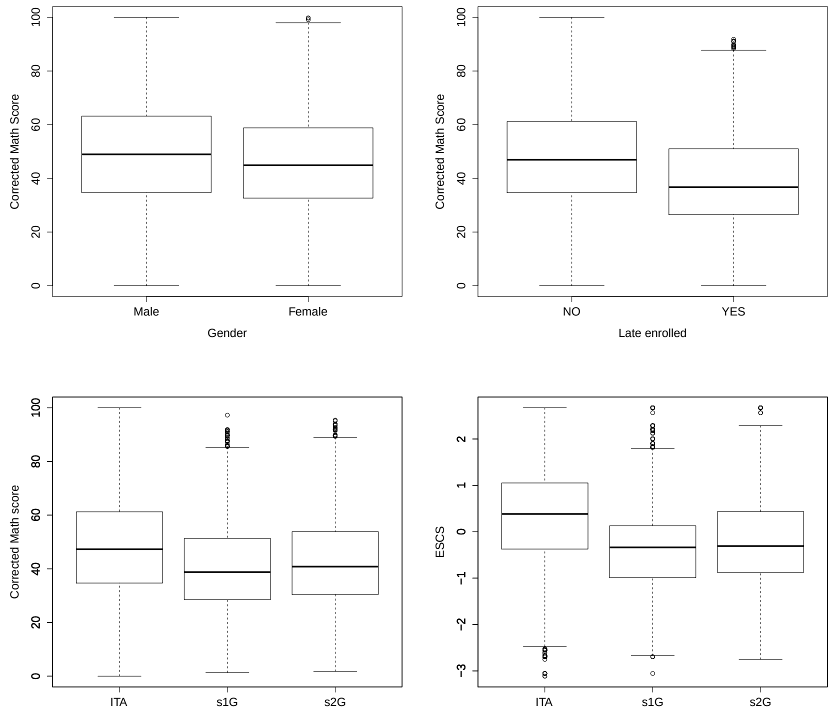

Figure 3: Flanked boxplots of CMS (upper panels and lower left panel) stratified by gender, being late enrolled and being first or second generation immigrant student. This last stratification is also adopted for ESCS (right lower panel).

ence among CMS and ESCS of italian and 1st generation immigrant students. Such a difference tends to reduce comparing the 2nd generation immigrants.

Significant positive associations between the CMS and the ESCS, and between CMS and CMS5 (the Corrected Math Score that pupils obtained in the Primary School 5^{th} grade evaluation) have also been detected. In particular, correlation between CMS and ESCS is equal to 0.28 (p-value of the Spearman correlation test $2.2 \cdot 10^{-16}$) and correlation between CMS and CMS5 is equal to 0.53 (p-value of the Spearman correlation test $2.2 \cdot 10^{-16}$) Finally, there is also evidence for stratifying the ESCS by indicators of being 1st generation immigrant and late enrolled (both Wilcoxon test p-value are lower than $2.2 \cdot 10^{-16}$).

Also the schools features do affect math attainments: attending a private school seems to have the effect of increasing the CMS (Wilcoxon test p-value $< 2.2 \cdot 10^{-16}$), as well as belonging to an *Istituto Comprensivo* (Wilcoxon test p-value equal to $3.70 \cdot 10^{-09}$)

3.2 Variation in math attainments between schools and across geographical areas

As we mentioned in the Introduction, there is a strong difference in educational attainment and results in different geographical macro-areas. A deeper investigation of the determinants of this gap is the main aim of the modeling approach adopted in this paper. Moreover, since between-school differences are stronger in the South than in the North, studying school effects on achievement for the different areas of the country seems worthy of interest. These are the main reasons we applied the models presented in Section 2.1 stratifying models by geographical areas (after having tested model (1) for the whole national population of students). In fact, this enables not only the investigation of the relationship between achievement and variables adjusted for students characteristics, but also a comparison among how each factor acts within the geographic subgroup. In fact, it is likely that the various elements contribute differently to students' achievement in the different geographical areas. All these degrees of heterogeneity have been kept into consideration in modeling the students' results. We think that this analysis is more informative and interesting with respect to the choice of just introducing a categorical variable indicating the geographical area in the fixed part of the model (which is our baseline strategy, see below), since it would only quantify the differences between areas in the mean outcome for a pupil with all the covariates equal to zero.

We then start fitting model (1) to the entire population of students, including a dummy for each of the three geographical areas. Results are reported in Table 2. It is worth noting the difference in Percentage of Variation captured by Random Effects (PVRE) over the three geographical areas. PVRE is obtained as the proportion of random effects variance over the total variation, i.e.,

$$\frac{\sigma_b^2}{\sigma_b^2 + \sigma_\epsilon^2} \tag{5}$$

All the levels of the categorical variable indicating which geographical area the student belongs to result significant with respect to the level taken as baseline (i.e., Central area). The estimated variance between schools (around 14.4%) is somewhat of the same magnitude of that reported by Invalsi (see Figure 1). Nevertheless, according to what we said before, this modelling approach captures the structural differences in math scores between students in different areas given the other variables at student and school level, i.e., assuming that these have an identical marginal effect on the students. In other words, this is equivalent to claim that the education production function is not different across areas. This is why we prefer to split the database in subsamples arising from the geographical division, focusing on comparing determinants of each area.

In the original database, the schools are distributed among the different geo-

Fixed effects	Estimate	(Std. Err.)
Intercept	7.7711***	(0.2582)
Female	-2.1067***	(0.0536)
1^{st} generation Immigrant	-0.8469***	(0.1499)
Late-enrolled student	-2.8484***	(0.1778)
ESCS	2.5045***	(0.0291)
Student NOT living with both parents	-1.3271***	(0.0811)
CMS5	0.5696***	(0.0018)
North-East	2.5965***	(0.3046)
North-West	2.6102***	(0.2892)
South	-3.6796***	(0.2719)
Random effects		
σ_b	5.552	
σ_ϵ	13.528	
PVRE	14.41%	
Size		
Number of Observations	259,757	
Number of Groups (schools)	4,119	

Table 2: ML estimates (with standard errors) for model (1), fitted to the *reduced* dataset. Asterisks denote different levels of significance: . $0.01 < p - val < 0.1$; $*0.001 <$ p-value < 0.01; $**\,0.0001 <$ p-value < 0.001; $***$ p-value < 0.0001

graphic areas as shown in Table 3 (upper part). Also the corresponding stratification of the students (lower part) is reported. The definition of *Northern area* is to be intended as comprehensive of both North East and North West.

School	**Area**		
Original DB	*Northern*	*Central*	*Southern*
4,119	1,843	712	1,564
100%	44.74%	17.29%	37.97%

Pupils	**Area**		
Original DB	*Northern*	*Central*	*Southern*
259.757	130.256	46.529	82.972
100%	50.15%	17.91%	31.94%

Table 3: Distribution of schools and pupils in the Original database (left column) and stratified by geographical area (right columns)

The Kruskal-Wallis Sum Rank test carried out on CMS stratified by geographical area shows that CMS is significantly lower in the South (p-value $< 2 \cdot 10^{-16}$), as can be seen also in the flanked boxplots reported in Figure 4. This is consistent with results arising from Table 2.

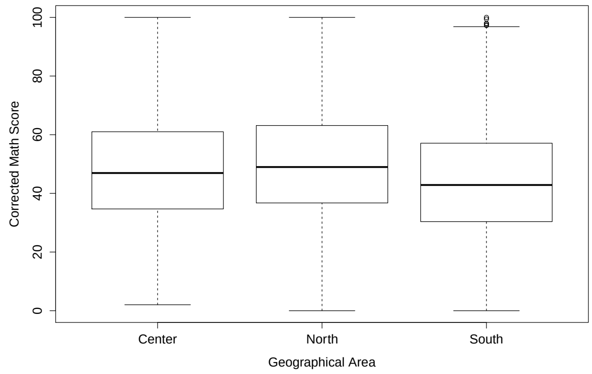

Figure 4: Flanked boxplots of CMS stratified by geographical location of schools.

Then, for each area we fit the following model: for each geographic area $R = \{Northern, Central, Southern\}$, for $i = 1, \dots, n_j^{(R)}$ and $j = 1, \dots, J^{(R)}$

$$y_{ij}^{(R)} = \beta_0^{(R)} + \sum_{k=1}^{K} \beta_k^{(R)} x_{kij}^{(R)} + b_j^{(R)} + \epsilon_{ij}^{(R)} \tag{6}$$

$$b_j^{(R)} \sim \mathcal{N}(0, \sigma_b^{2(R)}) \quad \epsilon_{ij}^{(R)} \sim \mathcal{N}(0, \sigma_\epsilon^{2(R)}) \tag{7}$$

$$\hat{b}_j^{(R)} = \gamma_0^{(R)} + \sum_{l=1}^{L} \gamma_l^{(R)} z_{lj}^{(R)} + \eta_j^{(R)} \tag{8}$$

$$\eta_j^{(R)} \sim \mathcal{N}(0, \sigma_\eta^{2(R)}) \tag{9}$$

Table 4 shows the estimates carried out fitting model (6) separately to the datasets of Northern, Central and Southern data.
The findings highlight that the educational production functions look quite different across the three geographical areas. For instance, the effect of gender on academic achievement is negative in all the three areas, but stronger in the South; on the contrary, being first generation immigrant is negatively correlated with achievement in the North, but not in the South. This latter result can be driven by (i) a lower proportion of immigrant students in the South and (ii) by worse (absolute and relative) scores of Italian students in the area. The family's background matters: in all the three areas, the variable measuring the socioeconomic condition (ESCS) is statistically significant and positively correlated with achievement in mathematics, even if such dependency is lower in the North - suggesting a higher degree of the equality of the educational system in this part of the country, i.e. the students' results are less dependent upon parents' characteristics. The sign and magnitude of the correlations with the individual-level variables is coherent with previous literature about the determinants of

students' achievement (for example, [6]). What is interesting and innovative is to look at the effect of prior achievement (at grade 5): as described above, no other control for prior students' results in a VA approach. What emerges from this analysis is not only that scores at grade 5 are highly predictive of subsequent results (as expected), but that this influence is different across geographical areas: more specifically, the correlation with previous results is higher in the North. Overall, before turning to the estimates of the schools' random effects, it is important to underline here an important methodological point: if the analysis does not adequately consider the structural differences across areas in the students' achievement patterns, the results would be inconsistent, as they would attribute to random school effects (b_j elements) which, instead, are related to differences in the role of that individual students' characteristics play in influencing test scores in the different regions - not attributable to schools themselves.

Fixed effects	NORTH	CENTER	SOUTH
Intercept	1.157***	7.914***	16.833***
	(0.196)	(0.357)	(0.311)
Female	-1.695***	-2.659***	-2.141***
	(0.069)	(0.126)	(0.102)
1^{st} generation Immigrant	-0.623***	-0.590	0.436
	(0.169)	(0.323)	(0.485)
Late-enrolled student	-2.566***	-1.794***	-3.933***
	(0.215)	(0.394)	(0.413)
ESCS	1.943***	2.428***	3.181***
	(0.038)	(0.071)	(0.054)
Student NOT living with both parents	-1.216***	-1.335***	-1.485***
	(0.100)	(0.182)	(0.175)
CMS5	0.700***	0.571***	0.387***
	(0.002)	(0.004)	(0.003)

Random effects	NORTH	CENTER	SOUTH
σ_b	3.645	4.510	7.354
σ_ϵ	12.434	13.527	14.622
PVRE	7.91%	10%	20.18%

Size	NORTH	CENTER	SOUTH
Number of Observations	130,256	46,529	82,972
Number of Groups (schools)	1,843	712	1,564

Table 4: ML estimates (with standard errors) for model (1), fitted to data of Northern, Central and Southern area. Asterisks denote different levels of significance: . $0.01 < p-val < 0.1$; $*0.001 < p-val < 0.01$; $**0.0001 < p-val < 0.001$; $***p-val < 0.0001$

Looking at the estimates of the schools' effects $b_j^{(R)}$s, they are characterized by a greater variability in the Southern area. Figure 5 shows the distributions $\mathcal{L}(\hat{b}^{(N)})$, $\mathcal{L}(\hat{b}^{(C)})$ and $\mathcal{L}(\hat{b}^{(S)})$ of the random effects estimated by fitting model (6) to the North, Center and South database respectively. They reflect the differences in variation we appreciated from computing PVRE in Table 4.

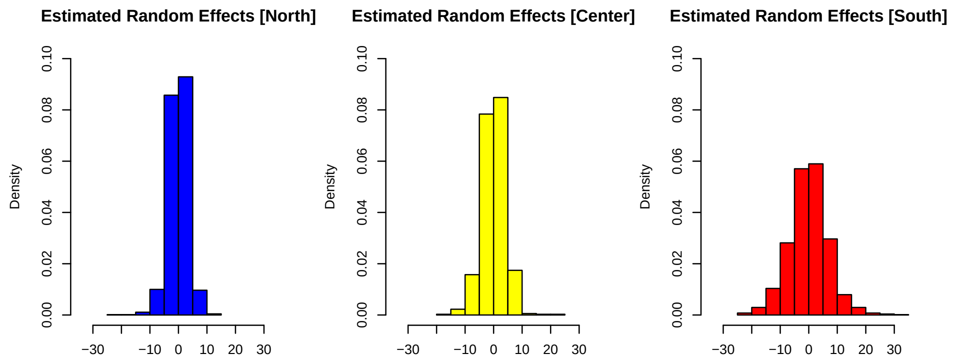

Figure 5: Distribution of the Random Effects arising from fitting model (6) to the databases of data concerning students and schools of Northern area (left panel), Central area (central panel) and Southern area (right panel).

The representation provided by model (6) enables us to highlight how the school effect acts on specific types of students profiles in different regions. Indeed, another key message of this paper is that the school effect is heterogeneous not only across geographical areas, but also across different individuals' profiles. For instance, the magnitude of the effect of attending a school A can be high for poor students, and low for a rich ones, or vice versa. With the aim of providing empirical evidence of such heterogeneity, we estimated the relationships between our variable of interest ($b_j^{(R)}$, $j = 1, \ldots, J^{(R)}$, the school random effect) and two individual characteristics which turned out to be strongly predictive of achievement, namely prior achievement ($CMS5$) an family's socioeconomic background ($ESCS$).

Figures 6 and 7 show how the CMS, estimated according to model (8), changes as a function of CMS5 and ESCS, in the different geographical areas. The estimation of CMS surfaces are given for 2 different set of values of the remaining covariates in model (8): the *best case* and the *worst case* scenario. The definitions of "best" and "worst" are given according to the signs of estimated coefficients in Table 4: Therefore, for Northern and Central areas the *best case* scenario means being Male, Italian, living with both parents, not late enrolled student, whereas it is being Male, 1st generation immigrant, living with both parents, not late enrolled student for the Southern area. On the contrary, the *worst case* scenario for the Northern and Central areas is represented by being Female, 1st generation immigrant, not living with both parents, late enrolled student, whereas it is being Female, Italian, not living with both parents, late enrolled student for the Southern area. Values are computed for mean values and mean values $\pm$ 2 standard deviation of the schools effects.

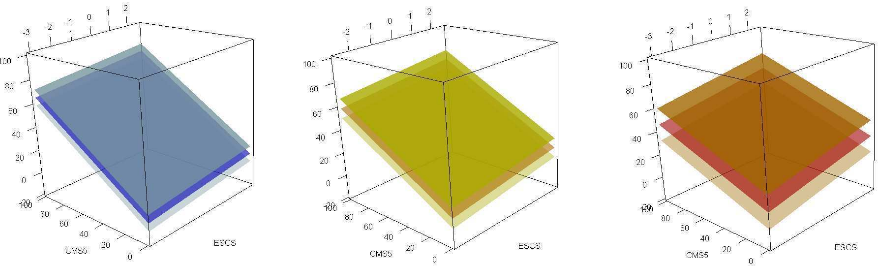

Figure 6: Distribution of the CMS in the *best case* scenario, stratified by geographical area (Northern is blue, Central is yellow and Southern is red). Random effects are fixed to their mean values (central plane) and to minus/plus 2 standard deviation (lower and upper shaded planes, respectively).

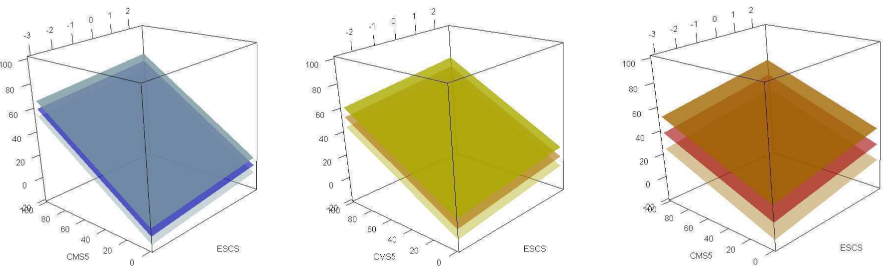

Figure 7: Distribution of the CMS in the *worst case* scenario, stratified by geographical area (Northern is blue, Central is yellow and Southern is red). Random effects are fixed to their mean values (central plane) and to minus/plus 2 standard deviation (lower and upper shaded planes, respectively).

Figures 6 and 7 confirm what it has been previously noticed, i.e., the greater variability of the school effect in the Southern area. Moreover, in the Northern area the prior achievement is definitively more important than in other areas; it is confirmed also that student's ESCS is much more influential in central and southern Italy.

Tables 5 and 6 highlights extremes cases of the CMS surfaces descrived above, again stratifying results by geographical area and considering the *best* and *worst* case scenario, respectively. Values are computed for mean values and mean values $\pm$ 1 standard deviation of the schools effects.

Estimated CMS - best case scenario

	-1σ rand.eff.		Mean rand.eff.		+1σ rand.eff.	
North						
	ESCS min	ESCS max	ESCS min	ESCS max	ESCS min	ESCS max
CMS5 min	0*	3.06	0*	6.34	1.77	12.90
CMS5 max	58.71	69.84	65.27	76.40	71.82	82.96
Center						
	ESCS min	ESCS max	ESCS min	ESCS max	ESCS min	ESCS max
CMS5 min	0*	6.22	1.38	14.39	9.55	22.57
CMS5 max	50.36	63.37	58.57	71.55	66.71	79.72
South						
	ESCS min	ESCS max	ESCS min	ESCS max	ESCS min	ESCS max
CMS5 min	0*	12.06	7.38	25.76	21.07	39.46
CMS5 max	32.39	50.78	46.09	64.47	59.79	78.17

* Indicates truncation to for non ammissible (i.e., < 0) predictions of the linear model in (8).

Table 5: Estimated CMS for Northern (upper rows), Central (central rows) and Southern (lower rows) area, according to model (8). Estimates refers to a 2×2 grid of values (min, max) of CMS5 and ESCS, and are reported for mean values of random effects (central columns) ± 1 standard deviation (left and right columns, respectively).

A further aspect that is interesting is to provide some empirical evidence about the main characteristics of the schools that exert a positive/negative effect on students' achievement. A potential approach for this purpose is to investigate substantially which are the main feature that can "explain" (in a correlational, not causal way) the schools' effect $b_j^{(R)}$, $j = 1, \ldots, J^{(R)}$.

Once the model in (6) is fitted to the data concerning each geographical area and the estimates for the random effects, we try to model them by means of suitable school-level covariates. Table 7 shows results obtained fitting the linear model in (8) to the North, Center and South datasets. The starting set of covariates is the same for all the three models but, according to previous findings, we also estimated the model separately for the three geographical areas. The choice of the variables has been guided by previous literature about the school-level factors that affect students' performance (as pointed out also in [2] and [6], among others) and refer to two main groups: (i) the peer effects related to the composition of student body (school-average ESCS, proportion of immigrant students or regular/early/later enrolled, etc.), and (ii) managerial and structural features of the school (size, *Istituto Comprensivo*, private, etc.).

Estimated CMS - worst case scenario

	-1σ rand.eff.		Mean rand.eff.		$+1\sigma$ rand.eff.	
North						
	ESCS min	ESCS max	ESCS min	ESCS max	ESCS min	ESCS max
CMS5 min	0^{*}	0^{*}	0^{*}	0.24	0^{*}	6.80
CMS5 max	52.60	63.74	59.16	70.29	65.72	76.86
Center						
	ESCS min	ESCS max	ESCS min	ESCS max	ESCS min	ESCS max
CMS5 min	0^{*}	0^{*}	0^{*}	8.02	3.18	16.19
CMS5 max	43.98	56.99	52.16	65.17	60.33	73.34
South						
	ESCS min	ESCS max	ESCS min	ESCS max	ESCS min	ESCS max
CMS5 min	0^{*}	4.07	0^{*}	17.76	13.07	31.46
CMS5 max	24.39	42.78	38.09	56.48	51.79	70.18

* Indicates truncation to for non ammissible (i.e., < 0) predictions of the linear model in (8).

Table 6: Estimated CMS for Northern (upper rows), Central (central rows) and Southern (lower rows) area, according to model (8). Estimates refers to a 2×2 grid of values (min, max) of CMS5 and ESCS, and are reported for mean values of random effects (central columns) $\pm$ 1 standard deviation (left and right columns, respectively).

Model Coefficients	NORTH	CENTER	SOUTH
Intercept	$-1.3175^{\cdot}$	-3.6470^{**}	-3.0748^{*}
mean school ESCS	0.4846^{*}	0.7090	1.9457^{***}
% Female	-0.0305^{*}	0.0549^{*}	0.0692^{**}
% 1st generation immigrants	-0.0553^{**}	$0.0822^{\cdot}$	0.1389^{**}
% Early-enrolled	-0.0905	-0.2115^{*}	-0.1594^{**}
% Late-enrolled	-0.0637^{**}	-0.0508	-0.2499^{***}
Number of students	-0.0040^{*}	0.0072^{*}	0.0118^{***}
Istituto Comprensivo	0.4328	0.5470	0.6935
Private	0.3899	-5.5368^{*}	1.5711
mean school ESCS : Private	-0.9840	3.1303^{*}	-1.1191

Table 7: ML estimates (with p-values) for model (6), fitted to data of Northern, Central and Southern area schools. Asterisks denote different levels of significance: $.\ 0.01 < p-val < 0.1$; $*0.001 < p-val < 0.01$; $**0.0001 < p-val < 0.001$; $***p-val < 0.0001$

It is interesting to notice that covariates act differently across the geographical areas. On the other hand, the low R^2s of the regressions (less than 20% for all the cases) suggest that a lot of variability remains unexplained considering the measurable variables only. Moreover, the design matrices result to be affected by a high correlation among their columns. In order to address the latter issue, we fitted a Lasso regression model [26] to the random effects estimates of each geographic area $R = \{Northern, Central, Southern\}$.

The Lasso regression is an efficient variable selection algorithm, which minimizes the residual sum of squares subject to the sum of the absolute value of the coefficients being less than a constant. So the Lasso is a regression method that involves penalizing the absolute size of the regression coefficients. By constraining the sum of the absolute values of the estimates we achieve a situation where some of the parameter estimates may be exactly zero. The larger the penalty applied, the more the estimates are shrunk towards zero. The penalization parameter λ is chosen by cross-validation techniques. Among the penalized regression methods, the Lasso one is very convenient when automatic variable selection is required and when dealing with highly correlated predictors.

Referring to the case of (8), we have

$$\gamma^{(R)} = \arg\min\{(\gamma_0^{(R)} + \sum_{l=1}^{L} \gamma_l^{(R)} z_{lj}^{(R)})^2\} \tag{10}$$

$$\text{subject to} \sum_l |\gamma_l| \leq \lambda \tag{11}$$

Because of the nature of the constraint it tends to produce some coefficients that are exactly 0 and hence gives interpretable models. Table 8 shows the resulting models selected by Lasso regression; we prefer to comment on it instead of the baseline regression because the results are more efficient and representative of the structural differences across areas.

LASSO Model Coefficients	NORTH	CENTER	SOUTH
Intercept	-0.6996	-3.5284***	-2.2368$^{\cdot}$
mean school ESCS		0.9171$^{\cdot}$	1.9452***
% Female	0.0312*	0.0627**	0.0686**
% 1st generation immigrants	-0.0601**	0.0547$^{\cdot}$	0.1383**
% Early-enrolled		-0.1958*	-0.1585**
% Late-enrolled	-0.0713**		-0.2474***
Number of students	0.0027*	0.0050*	0.0118***
Istituto Comprensivo			0.0085***
Private	-0.7481**	-2.570**	

Table 8: ML estimates (with p-values) for model (10), fitted to data of Northern, Central and Southern area schools. Asterisks denote different levels of significance: . $0.01 <$ p-value < 0.1; $*$ $0.001 <$ p-value < 0.01; $**$ $0.0001 <$ p-value < 0.001; $***$ p-value < 0.0001

Generally speaking, the composition of student body seems more relevant in the South than in the North, as the strong and substantially significant effect of school-average ESCS reveals, as well as the proportion of early/late students. The only exception is the proportion of 1^{st} generation immigrants, that seems to have a negative effect only in the North, but this is probably due to the higher proportion of these students in the area, and the positive effect in the South is arguably due to low performances of Italian students. The negative coefficient attached to private schools is coherent with previous studies, and can indicate that - net of other variables - the estimated school effect is (on average) lower for private than public schools (although it can also mask heterogeneity within the group of private schools; moreover, the magnitude of this effect is lower in the North, and the effect itself is not statistically significant in the South). Other "structural" variables (size and *Istituto Comprensivo*) are partially correlated with the estimated schools' random effects, but the magnitude of these effects is negligible. Summarizing, the main explanation of the school's specific effect can be attributed to the characteristics of students' composition.

Even if the collinearity issue can be addressed by using penalized regression techniques, the amount of unexplained variability remains high. This is probably due to the unobserved variables like those that reflect the kind of activities which are undertaken within classes of each school, together with those at school level. In other words, part of the school effect is actually driven by differences between classes of the same school, so exploring the variance between-classes (within-school) can add explanatory power to our empirical analysis.

We denote by y_{ijk} the attainment at stage 6 in mathematics (CMS) of pupil i, $i = 1, \ldots, n_{lj}^{(R)}; n^{(R)} = \sum_{l,j} n_{lj}^{(R)}$, in class l, $l = 1, \ldots, L_j^{(R)}; L^{(R)} = \sum_k L_j^{(R)}$, in school j, $j = 1, \ldots, J^{(R)}$. We then fit a three-level random effects model. The simplest such model allows the regression intercept to vary randomly across classes and schools [23]. So for each geographic area $R = North, Center, South$, we have

$$y_{ilj}^{(R)} = \beta_0^{(R)} + \sum_{k=1}^{K} \beta_k^{(R)} x_{kilj} + b_j^{(R)} + u_{lj}^{(R)} + \epsilon_{ilj}^{(R)} \tag{12}$$

$$b_j^{(R)} \sim \mathcal{N}(0, \sigma_{School}^{2^{(R)}}) \quad u_{lj} \sim \mathcal{N}(0, \sigma_{Class}^{2^{(R)}}) \quad \epsilon_{ilj} \sim \mathcal{N}(0, \sigma_{\epsilon}^{2^{(R)}}) \tag{13}$$

where x_{ijk} is the value of the $k-th$ predictor variable at student's level, $\beta^{(R)} = (\beta_0^{(R)}, \ldots, \beta_k^{(R)})$ is the $(K+1)$-dimensional vector of parameters referred to the R-th geographical area to be estimated and $\epsilon_{ilj}^{(R)}$ is the zero mean gaussian error. The random effects $u_{lj}^{(R)}$ for the l-th class within the j-th school and b_j for the j-th school is assumed to be independent of any predictor variables that are included in the model.

The results (see Table 9) show some interesting elements. First, part of the variance that was explained at school level, now is attributed to differences between classes, nevertheless variance between schools is still higher in the South than in the North - so, the school effects are either more relevant there. Of particular interest is the estimated variance between classes, which is substantial in all of the three areas (similar to the variance between schools in magnitude), highlighting that not only the chosen school matters, but also the specific class attended by the students. Such an effect is even more marked in the South (where the variance between classes is much higher than between schools) suggesting the presence of sorting phenomena (or different educational quality) even within each school, that can explain some unobserved components of school effects.

Fixed effects	NORTH	CENTER	SOUTH
Intercept	0.797^{***}	7.305340^{***}	16.524^{***}
	(0.194)	(0.348)	(0.294)
Female	-1.683^{***}	-2.638^{***}	-2.165^{***}
	(0.067)	(0.121)	(0.093)
1st generation Immigrant	-0.637^{**}	-0.377	0.389
	(0.166)	(0.311)	(0.440)
Late-enrolled stud.	-2.466^{***}	-1.827^{***}	-3.791^{***}
	(0.211)	(0.379)	(0.375)
ESCS	1.879^{***}	2.268^{***}	2.676^{***}
	(0.037)	(0.068)	(0.050)
noMF	-1.182^{***}	-1.256^{***}	-1.276^{***}
	(0.098)	(0.176)	(0.158)
MS5	0.706^{***}	0.581^{***}	0.391^{***}
	(0.002)	(0.004)	(0.003)
Random effects	NORTH	CENTER	SOUTH
σ_{School}	3.13	3.58	5.77
σ_{Class}	3.68	5.19	8.17
σ_{ϵ}	12.00	12.75	12.86
Size	NORTH	CENTER	SOUTH
Number of Observations	130,256	46,529	82,972
Number of Groups (schools)	1,843	712	1,564
Number of Groups (classes)	8,615	3,485	6,661

Table 9: ML estimates (with standard errors) for model (12), fitted to data of Northern, Central and Southern area. Asterisks denote different levels of significance: . $0.01 <$ p-value < 0.1; $*0.001 <$ p-value < 0.01; $**\ 0.0001 <$ p-value < 0.001; $***$ p-value < 0.0001

Table 9 illustrates another interesting feature of the geographical gap, as the "class-effect" is again higher in the South than in the North of the country, suggesting that in that area not only the chosen school matters, but also the class that the student attends has a higher and significative effect on the student's test scores. As a further step, we first calculated an average "class-effect" for each school, then we computed correlations between class-effects and school-effects. Such correlations are very high in all the country, more specifically 89.9%, 85.7% and 84.3% in Northern, Central and Southern Italy respectively; this result suggests that those schools which effect is higher (positive/negative) on achievement tend to put into practice differential situations between their classes, and consequently this has an higher (positive/negative) effect on achievement at class level. It is interesting to note that the direction of these effects is the same (i.e. the correlation is positive); in other words, in those schools which effect on achievement is positive (negative), the (average) class-effect also tend to be positive (negative).

4 Concluding Remarks

This paper explored the school effects associated with achievement of Italian students who attended grade 6 (first year of junior secondary school) in 2011/12. A multilevel model has been proposed and used for this purpose, and a large and new dataset provided by Invalsi allowed us to control for individual-level covariates, among which also test score in grade 5 is available - thus, the empirical exercise can be considered among the first real Value Added Model (VAM) exercises in the context of the Italian educational system. The school effect, defined as the independent statistical effect of attending a specific school on a student's test score, has been modeled as a random effect b_j; then, this has been regressed against a set of school-level variables, with the aim of characterising the features of those schools that exert a positive (negative) effect on academic performance. The results show that the average socioeconomic condition of students attending the school has a negative effect on their performance, confirming that peers' characteristics play a role in the achievement process, and suggesting that some school-level tracking is likely to happen, e.g. with better-off and worse-off students sorting themselves in different institutions.

What is more interesting on the policy ground is that the school effects are different in terms of magnitude and types in the three geographical macro-areas: Northern, Central and Southern Italy. Such differences are so marked that we estimate three different models, assuming that the characteristics of individuals and schools in the educational process do not influence in the same way - in other words, they can be considered as three different educational systems. The empirical evidence corroborates this intuition: not only school effects are much stronger in the South than in the North (that is, the former geographical area is characterized by a more diversified quality of schools) but also the interplay between individual and school characteristics is not uniform across the country. Indeed, we used our estimations for simulating how school effects b_j are heterogeneous and dependent upon individual and schools' features, and we discovered for example that the differential school effect on disadvantaged and advantaged students is much more pronounced in Southern Italy than in the North. Overall, our findings claim that not only schools are not of the same quality (contrary to the institutional presumption at the basis of the Italian educational system), but also differences between schools in the South tend to increase instead of reducing inequalities between more disadvantaged and advantaged students, raising serious issues about equality. The policy implications of this work deal with the necessity of (i) providing some kind of information to parents to make them informed about the broad differences between schools' quality, and (ii) focusing the attention to the schools in the South, given that even the more recent interventions funded through European Funds do not seem to have had the effect of helping them in raising educational results - indeed, the gap with the North is still evident.

Acknowledgments

We are grateful to Invalsi for having provided the original dataset, and P. Falzetti for the statistical assistance in building the specific database used in this paper.

References

[1] Agasisti, T., Falzetti, P. (2013) Between-classes sorting within schools and test scores: an empirical analysis of the Italian junior secondary schools. Paper presented at the *XXVIII National Conference of Labour Economics*, Rome, September 2013, 4 (1): 33–53

[2] Agasisti, T., Vittadini, G. (2012) Regional economic disparities as determinants of students' achievement in Italy, *Research in Applied Economics*, 4 (1): 33–53

[3] Andersson, C., Johansson, P. (2013). Social stratification and out-of-school learning. *Journal of the Royal Statistical Society - Series A*, 176 (3): 679–701

[4] Bates, D., Maechler, M., Bolker, B., Walker, S. (2013) `lme4`: Linear mixed-effects models using Eigen and S4. *R package version 3.1-111.*

[5] Bertoni, M., Brunello, G., Rocco, L. (2013) When the cat is near, the mice wont play: the effect of external examiners in the Italian schools. *Journal of Public Economics*, 104 (1): 65–77

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

[7] Chowdry, H., Crawford, C., Dearden, L., Goodman, A. (2013). Widening participation in higher education: analysis using linked administrative data. *Journal of the Royal Statistical Society - Series A*, 176 (2): 431–457

[8] Cipollone, P., Montanaro, P., Sestito, P. (2010) Value-added measures in Italian high schools: problems and findings. *Giornale degli Economisti e Annali di Economia*, 69 (2): 81–114

[9] Conti, E., Duranti, S., Maitino, M.L., Rampichini, C., Sciclone, N. (2013). The future has early roots. Learning outcomes and school's effectiveness in Tuscany's primary education system. Workshop Youth and their future: Work, Education and Health October 17th -18th, 2013 - University of Salerno.

[10] De Simone, G., Gavosto, A. (2013) Patterns of value-added creation in the transition from primary to lower secondary education in Italy. Paper presented at the *XXVIII National Conference of Labour Economics*, Rome, September 2013.

[11] Goldstein, H., Rasbash, J., Yang, M., Woodhouse, G., Pan, H., Nuttall, D., Thomas, S. (1993) A multilevel analysis of school examination results. *Oxford Reviw of Education*, 19: 425-433

[12] Goldstein, H., Sammons, P. (1997) The influence of secondary and junior schools on sixteen year examination performance: a cross-classified multilevel analysis. *School Effectiveness, School Improvement*, 8: 219-230

[13] Goldstein, H. (2003) Multilevel Statistical Models, 3rd edn. London: Arnold.

[14] Goldstein, H., Burgess, S., McConnell, B. (2007) Modelling the effect of pupil mobility on school differences in educational achievement. *Journal of the Royal Statistical Society - Series A*, 170: 941-954

[15] Grieco, N., Ieva, F., Paganoni, A.M. (2012). Performance assessment using mixed effects models: a case study on coronary patient care. *IMA Journal of Management Mathematics*, 23 (2): 117–131

[16] Leckie, G., Goldstein, H. (2009) The limitations of using school league tables to inform school choice. *Journal of the Royal Statistical Society - Series A*, 172 (4): 835–851

[17] OECD (2010) PISA 2009 results: what students know and can do. *Oecd - Paris*

[18] OECD (2013) PISA 2012 results in focus: what 15-year-olds know and what they can do with what they know. *Oecd - Paris*

[19] Pinheiro, J., Bates, D., DebRoy, S., Sarkar, D. and the R Development Core Team (2013). `nlme`: Linear and Nonlinear Mixed Effects Models. *R package version 3.1-111.*

[20] Plewis, I. (2011). Contextual variations in ethnic group differences in educational attainments. *Journal of the Royal Statistical Society - Series A*, 174 (2): 419–437

[21] R Development Core Team (2009), R: A language and environment for statistical computing, R Foundation for Statistical Computing, Vienna, Austria. [online] `http://www.R-project.org`

[22] Rasbash, J., Leckie, G., Pillinger, R., Jenkins, J. (2011). Childrens educational progress: partitioning family, school and area effects. *Journal of the Royal Statistical Society - Series A*, 173 (3): 419–437

[23] Snijders, T., Bosker, R. (1999) Multilevel Analysis. London: Sage.

[24] Steele, F., Vignoles, A., Jenkins, A. (2007) The effect of school resources on pupil attainment: a multilevel simultaneous equation modelling approach. *Journal of the Royal Statistical Society - Series A*, 170 (3): 801–824

[25] Struyf, A., Hubert, M., Rousseeuw, P.J. (1996) Clustering in an object-oriented environment. *Journal of Statistical Software*, 1 (4): 1–30

[26] Tibshirani, R. (1996) Regression Shrinkage and Selection via the Lasso. *Journal of the Royal Statistical Society - Series B*, 58 (1):267–288

MOX Technical Reports, last issues

Dipartimento di Matematica "F. Brioschi",
Politecnico di Milano, Via Bonardi 9 - 20133 Milano (Italy)

07/2014 Agasisti, T.; Ieva, F.; Paganoni, A.M.
Heterogeneity, school-effects and achievement gaps across Italian regions: further evidence from statistical modeling

06/2014 Benzi, M.; Deparis, S.; Grandperrin, G.; Quarteroni, A.
Parameter estimates for the relaxed dimensional factorization preconditioner and application to hemodynamics

05/2014 Rozza, G.; Koshakji, A.; Quarteroni, A.
Free Form Deformation Techniques Applied to 3D Shape Optimization Problems

04/2014 Palamara, S.; Vergara, C.; Catanzariti, D.; Faggiano, E.; Centonze, M.; Pangrazzi, C.; Maines, M.; Quarteroni, A.
Patient-specific generation of the Purkinje network driven by clinical measurements: The case of pathological propagations

03/2014 Kashiwabara, T.; Colciago, C.M.; Dede, L.; Quarteroni, A.
Well-posedness, regulariy, and convergence analysis of the Finite Element approximation of a Generalized Robin boundary value problem

02/2014 Antonietti, P.F.; Sarti, M.; Verani, M.
Multigrid algorithms for high order discontinuous Galerkin methods

01/2014 Secchi, P.; Vantini, S.; Zanini, P.
Hierarchical Independent Component Analysis: a multi-resolution non-orthogonal data-driven basis

67/2013 Canuto, C.; Simoncini, V.; Verani, M.
On the decay of the inverse of matrices that are sum of Kronecker products

66/2013 Tricerri, P.; Dede ,L; Quarteroni, A.; Sequeira, A.
Numerical validation of isotropic and transversely isotropic constitutive models for healthy and unhealthy cerebral arterial tissues

65/2013 Ambrosi, D.; Ciarletta, P.
Plasticity in passive cell mechanics