MOX-Report No. 18/2015

Bivariate multilevel models for the analysis of mathematics and reading pupils' achievements

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

MOX, Dipartimento di Matematica
Politecnico di Milano, Via Bonardi 9 - 20133 Milano (Italy)

mox-dmat@polimi.it http://mox.polimi.it

Bivariate multilevel models for the analysis of mathematics and reading pupils' achievements

Chiara Masci[♮], Francesca Ieva[♯], Tommaso Agasisti[*], Anna Maria Paganoni[♮]

April 9, 2015

[♮] MOX– Modelling and Scientific Computing, Department of Mathematics,
Politecnico di MIlano, via Bonardi 9, Milano, Italy
chiara.masci@mail.polimi.it
anna.paganoni@polimi.it
[*] 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

Keywords:Pupils' achievement; Multilevel models; Bivariate models; School and class effects; Value-added.

Abstract

The purpose of this paper is to identify a relationship between pupils' mathematics and reading test scores and the characteristics of students themselves, stratifying for classes, schools and geographical areas. The dataset of interest contains detailed information about more than 500,000 students at the first year of junior secondary school in the year 2012/2013, provided by the Italian Institute for the Evaluation of Educational System (INVALSI). The innovation of this work is in the use of multivariate models, in which the outcome variable is bivariate: reading and mathematics achievements. Using the bivariate outcome enables researchers to analyze the correlations between achievement levels in the two fields and to estimate statistically significant school and class effects after adjusting for pupil's characteristics. The statistical model employed here explicates account for the potential covariance between the two topics, and at the same time it allows the school effect to vary among them. The results show that while for most cases the direction of school's effect is coherent for reading and mathematics (i.e. positive/negative), there are cases where internal school factors lead to differential performances in the two fields.

1 Introduction and Motivation

Nowadays, the analysis of the differences in educational attainments between groups of students and across schools and classes is becoming increasingly interesting. Due to the increasing demand for public education accountability, studies on this topic are carried out in order to test and improve the educational system and to understand which variables mostly affect it (see [3],[7],[17]). The Programme for International Student Assessment (PISA) is a project promoted by the Organization for Economic Co-operation and Development (OECD) that was created in year 2000 in order to analyze the educational level of the teenagers in the main industrialized countries. Its purpose is to compare the results of the tests, identifying which are the aspects that distinguish countries with high and low performances. It is quite common in the most industrialized countries to have standardized tests administrated primarily to assess the achievement levels of students of various grades.

In Italy, the Italian Institute for the Evaluation of Educational System (hereafter INVALSI, founded in 2007), assesses students abilities in reading and mathematics at different stages. This happens at the end of the second and fifth year of primary school (when pupils are aged 7 and 10, respectively), at the end of the first and third year of lower secondary school (aged 11 and 13) and at the end of the second year of upper secondary school (aged 15). Students are requested to answer questions with both multiple choices and open-ended questions, that test their ability in reading and mathematics. This is a way to test knowledge and reasoning that pupils should have learned in their school career. Also, they are requested to compile a questionnaire about themselves, their family, their parents' educational level and their socio-economic situation.

The institutional organization of the Italian educational system is based on the strong assumptions about its equality purposes, among which all schools and classes provide similar educational standards. Therefore, recent aggregate data provided by the Invalsi show that this is not the case, and that a significant portion of variance in students' test scores is attributable to the structural between-schools differences. Multilevel models have enabled researchers to catch the stratification between schools, but, up to this time, only in univariate case.

Several studies are present in literature on the mathematics and reading achievements of Italian students, where they are treated as separate (see [1]), applying univariate multilevel linear models (see [4], [5], [12], [13]) to analyze how the outcome variables (mathematics or reading achievement) depend on the students' characteristics, and which are the schools/classes affecting students' achievements most. Regarding the characteristics of students' profile, it emerged that, for instance, 1^{st} and 2^{nd} generation immigrants students obtain, on average, worse test scores than native Italian students, both in reading and mathematics; females have better average results than males in reading, but worse ones in mathematics; early/late-enrolled students have worse performances than "regular" students. Also, big differences elapse between North, Center and South of Italy: students in the South have lower mean results than students in the North and the aspects of students' profile weigh in a different way across the geographical macro-areas, emphasizing the need to have three different models to explain the completely different phenomena. Furthermore, the results asserted that the choice of the school and especially of the class can influence the students' performances, acting beyond the pure effect of their socio-economic background. In both the studies emerged some common behaviors about students' characteristics and we noted a good correlation between the reading and mathematics achievements. As we explained before, there are aspects of student's profile that weigh in the same way both in reading and mathematics. It has also been possible to identify a coherence between the school and contained class effects, but only considering the two topics as separate. At this point, we held that could be interesting to analyze the relationships and the interactions between the school/class effects in reading and mathematics. So that, the main purpose of this work is to fit multilevel linear models, for the bivariate outcome of reading and mathematics achievements, in order to assess this issue. By means of the bivariate models, we can jointly study the results of mathematics and reading and compare them; we can estimate the school/class effects of the two topics and examine if they are affected by the same type of variables at school/class level; and, moreover, we can analyze the correlations between the values-added by schools in the two topics and point out if there is heterogeneity within schools.

The steps that we carry out in this paper are: (a) to examine the relationship between pupils' characteristics, such as profile, socio-cultural background, household, cultural resources, and pupil's achievement, (b) to discover how the school and class effects positively or negatively influence specific types of students' profile and how they are correlated when considering the joint "production" of reading and mathematics achievements. We will figure out that there are big differences both between school and class effects and between mathematics and reading, showing discrepancies between and within schools and across the country. In particular, we will discover that the class effects weigh more than the school ones, proving that the choice of the class may influence the students' performances more than the choice of the school and that the school/class effects in mathematics and reading may be influenced by different aspects. Furthermore, while the school effects in the two topics will prove to be pretty coherent, identifying good/bad schools, the two class effects will not, proving to be not correlated. Lastly, we will figure out that the scholl effect is independent of the contained class effects, suggesting that good (or bad) schools do not necessarily contain good (or bad) classes.

The paper is organized as follows: in Section 2 we present the dataset and the models that we have used; in Section 3 we analyze students nested in schools by means of bivariate two-level linear models, stratifying the models by macro-areas; in Section 4 we analyze students nested in classes, which in turn nested in schools by means of a three-level model; Section 5 contains discussion and conclusions. All the bivariate multilevel models are implemented using the software Asreml [6] and the other models and analysis are implemented in the statistical software R [14].

2 Dataset and Models

In this section we present the dataset of interest and the statistical tools requested to make the analysis.

Our resources are two separate set of data, containing information about more than 500,000 students attending the first year of junior secondary school in the year 2012/2013, provided by INVALSI. The former contains the mathematics achievements and the latter the reading ones, followed by the corresponding information about students, classes and schools. The reading dataset contains information about 510,933 students and the mathematics one about 509,371. We linked these two dataset, retaining only the students that have both the test scores of mathematics and reading, followed by all the variables presented in the two set of data. The deterministic linkage of the two dataset is possible thanks to the anonymous student ID that is known for each pupil. We then obtain a new dataset containing 507,229 students, whose both the achievements in mathematics and reading are known, and 50 variables, loosing very few individuals.

At pupil's level, the following information is available: gender, immigrant status (Italian, first generation, 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). The dataset contains also information about the family's background: if the student lives or not with both parents (i.e. the parents are 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. In other words, it is built considering (i) parents' occupation and educational titles, and (ii) possession of certain goods at home (for instance, the number of books). Once measured, this indicator has been standardized to have mean zero and variance one. The minimum and maximum observed values in the Invalsi dataset are - 3.11 and 2.67, respectively. In general, pupils with ESCS equal to or greater than 2 are very socially and culturally advantaged (high family's socioeconomic background). Among data, there are also the Invalsi scores in the Mathematics and Reading tests at grade 5 of the previous year (ranging between 0 and 100), which are 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 dataset also allows us 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 latest variable is relevant to understand if the "continuity" of the same educational environment affects (positively or negatively) students results. Some variables about size (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, regarding geographical location, we include two dummies for schools located in Central and Southern Italy and the district in which the school is located; 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. As we have the anonymous student ID, we have also the encrypted school and class IDs that allow us to identify and distinguish schools and classes. The outputs MS and RS (hereafter, respectively the score in the Mathematics and Reading standardized test administered by Invalsi) are expressed as "cheating-corrected" scores (CMS and CRS). In fact, 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. These variables take values between 0 and 100.

Unfortunately, there are lots of missing data in the score at grade 5, both in mathematics and reading achievements. This kind of data may have been lost in the passage of information between primary and junior secondary schools. Since having longitudinal data is very important for this study, we omit the individuals with missing data at grade 5, loosing almost 300,000 students. The final and reduced dataset collects 221,529 students, almost half of the initial dataset, within 16,246 classes, within 3,920 schools.

Hereafter, all the analysis are made on this reduced dataset. The variables and some related descriptive statistics are presented in Table 1.

Level	Type	Variable Name	Mean	sd
Student	-	Student ID	-	-
Student	(Y/N)	Female	49.8%	-
Student	(Y/N)	1^{st} generation immigrants	4.4%	-
Student	(Y/N)	2^{nd} generation immigrants	4.9%	-
Student	num	ESCS	0.24	1.02
Student	(Y/N)	Early-enrolled student	1.6%	-
Student	(Y/N)	Late-enrolled student	2.8%	-
Student	(Y/N)	Not living with both parents	12.6%	-
Student	(Y/N)	Student with siblings	83.3%	-
Student	%	Cheating	0.016	0.05
Student	num	Written reading score	9.41	2.74
Student	num	Oral reading score	6.80	1.13
Student	num	Written mathematics score	9.48	2.75
Student	num	Oral mathematics score	6.88	1.35
Student	num	CMS5-5^{th} year Primary school mathematics score	70.5	16.30
Student	num	CRS5-5^{th} year Primary school reading score	74.5	13.50
Class	-	Class ID	-	-
Class	num	Mean ESCS	0.18	0.48
Class	%	Female percentage	43.7	10.07
Class	%	1^{st} generation immigrant percent	5.4	6.47
Class	%	2^{nd} generation immigrant percent	4.7	5.83
Class	%	Early-enrolled student percent	1.4	3.24
Class	%	Late-enrolled student percent	6.2	6.11
Class	%	Disable percentage	5.8	5.58
Class	count	Number of students	23	3.49
Class	(Y/N)	"Tempo pieno"	0.023%	-
School	-	School ID	-	-
School	num	Mean ESCS	0.18	0.41
School	%	Female percentage	43.3	5.46
School	%	1^{st} generation immigrant percent	5.4	4.65
School	%	2^{nd} generation immigrant percent	4.6	4.06
School	%	Early-enrolled student percent	1.5	2.23
School	%	Late-enrolled student percent	6.3	3.94
School	count	Number of students	143	76.52
School	count	Average number of students	22.6	2.94
School	count	Number of classes	6.2	3.05
School	(Y/N)	North	52%	-
School	(Y/N)	Center	18%	-
School	(Y/N)	South	30%	-
School	-	District	–	-
School	(Y/N)	Private	3.1%	-
School	(Y/N)	"Istituto comprensivo"	65.8%	-

Level	Type	Variable Name	Mean	sd
Outcome	num	CMS-Mathematics Score corrected for Cheating	47.4	17.67
Outcome	num	CRS-Reading Score corrected for Cheating	65	14.65

Table 1: Variables of the database

The main statistical tools requested to make this study are bivariate multilevel linear models. We develop two-level models in which students (level 1) are nested in schools (level 2) and then three-level models in which students (level 1) are nested in classes (level 2), that are nested in schools (level 3). We consider only variables at student level with random effects on schools or classes. This allows us to individuate the relationships between the test results and the characteristics of student's profile and to estimate the random effects, such as school and class effects.

As we introduced before, the outcome variable of these models is bivariate: reading and mathematics achievements. Figure 1 shows the histograms of the two test results.

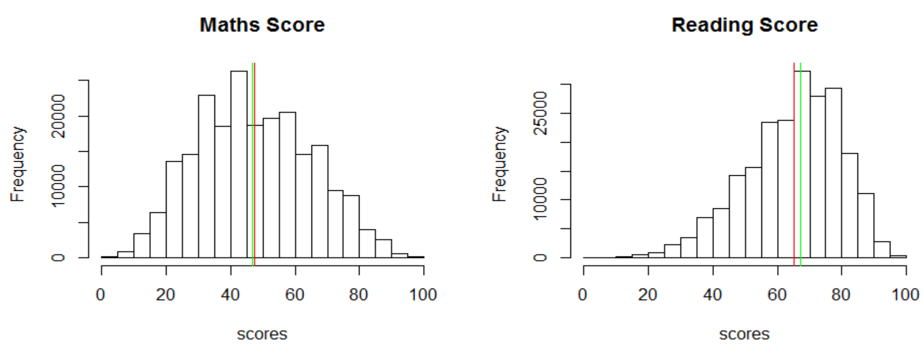

Figure 1: Histogram of Corrected Reading and Mathematics Score of pupils in the Invalsi database. The red lines refer to the mean, the green ones to the median.

It is immediately clear that the two distributions are different and that the CRSs are on average higher than the CMSs. There is a positive correlation between the performances of students in the two topics, CMS and CRS. By a test of correlation (Pearson's product moment), we obtain a coefficient of correlation of 0.59 with a high significance (p-value $< 2.2e-16$).

3 Bivariate two-level linear model: students nested in schools

Let's start with a bivariate two-level linear model in which pupils (level 1) are nested in schools (level 2). Pupil i, i = 1,..., n_j; $n = \sum_j n_j$ (first level) is in school $j, j = 1, ..., J$ (second level):

$$\mathbf{y}_{ij} = \boldsymbol{\beta}_0 + \sum_{k=1}^{K} \boldsymbol{\beta}_k x_{kij} + \mathbf{b}_j + \boldsymbol{\epsilon}_{ij} \tag{1}$$

where

$\mathbf{y}_{ij}$ is the bivariate outcome with mathematics and reading achievements of pupil i in school j;

$\boldsymbol{\beta} = (\boldsymbol{\beta}_0, ..., \boldsymbol{\beta}_K)$ is the bivariate (K+1)-dimensional vector of parameter;

x_{kij} is the value of the k-th predictor variable at student's level;

$\mathbf{b} \sim N_2(\mathbf{0}, \Sigma)$ is the matrix of the bivariate random effects (mathematics and reading) at school level;

$\boldsymbol{\epsilon} \sim N_2(\mathbf{0}, W)$ is the error.

We assume $\mathbf{b}$ independent of $\boldsymbol{\epsilon}$.

Using the software *AsReml*, we can fit this bivariate model and obtain the estimates of the coefficients and the random effects. Table 2 shows the results. Note that we managed to use the CRS5/CMS5 as a regressor only for the reading/mathematics achievement respectively, because the reading achievement doesn't depend on the mathematics score at grade 5 and the mathematics achievement doesn't depend on the reading one.

Fixed Effects	Mathematics coeff	Reading coeff
Intercept	14.91	30.44
Female	−2.211	2.134
1^{st} generation immigrant	−1.511	−3.921
2^{nd} generation immigrant	−2.281	−3.548
South	−6.437	−4.670
Center	−2.699	−1.163
Early-enrolled student	−0.793	−0.792
Late-enrolled student	−2.744	−3.638
ESCS	2.625	2.211
Not living with both parents	−1.463	−1.104
Student with siblings	0.049	−0.644
CS5	0.505	0.476
Variance/Covariance matrix of random effects	$\begin{pmatrix} 23.04 & 5.51 \\ 5.51 & 13.08 \end{pmatrix}$	
Variance/Covariance matrix of error	$\begin{pmatrix} 180.5 & 63.13 \\ 63.13 & 132.25 \end{pmatrix}$	
Size		
Number of observations	221,529	
Number of groups (School)	3,920	

Table 2: ML estimates of model (1) fitted to the dataset.

Now, we can analyze the relationships between the test scores and the characteristics of students, comparing the estimates of the coefficients of the two topics. As we anticipated before, the coefficients of the variable “female” of the two topics are almost opposites: being a female has a good effect in reading and a bad one in mathematics, suggesting that on average males are better in mathematics, while females are better in reading. Being immigrants has a negative effect in both the fields, but especially in reading, since the main difficulty for immigrants students is the language. Being a student in the South of Italy has a worse effect in mathematics than in reading, while anyway has a negative effect in both the topics: students of the South have worse results than students of the North. Being early/late-enrolled has a negative effect in both the fields. The ESCS and the score at grade 5 are positively correlated with the achievements and have similar coefficients in both the fields. Regarding the ESCS, this means that the socio-economic status of student and the family background weigh positively on the performances: students with high value of ESCS have better performances. Lastly, the positive influence of the score at grade 5, suggests that there is a continuity in students' efficiency.

Looking at the variance/covariance matrix of the random effect, it is evident that the variability of the mathematics random effect is much higher than the reading one (23.04 vs 13.08), therefore, attending a specific school influences more the results in mathematics than in reading. The two effects are positively correlated (0.307). Figure 2 shows the variability of the marginal random effects.

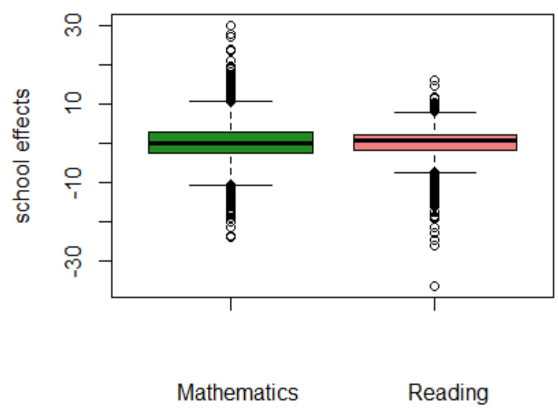

Figure 2: Estimated school effects $\hat{\boldsymbol{b}}_j$ in mathematics and reading.

To test this difference in variability, we implement a non-parametric Levene's test. We obtain a p-value less than $2.2e-16$, proving that the variances of the random effects of the two topics are different.

3.1 Differences across macro-areas

From all the previous studies, emerged that big discrepancies elapse between the three geographical macro-areas. Therefore, we fit model (1) for each macro-area, in order to point out the differences across the areas and the trends inside them:

$$\mathbf{y}_{ij}^{(R)} = \boldsymbol{\beta}_0^{(R)} + \sum_{k=1}^{K} \boldsymbol{\beta}_k^{(R)} x_{kij} + \mathbf{b}_j^{(R)} + \boldsymbol{\epsilon}_{ij}^{(R)} \tag{2}$$

where $R = \{North, Center, South\}$

The estimates of model (2) are reported in Table 3.

Fixed Effects	North mat	Center mat	South mat
Intercept	6.2	12	20
Female	-1.8	-2.8	-2.15
1^{st} generation imm	-1.2	-1.17	0.18
2^{nd} generation imm	-2.3	-1.4	-0.55
Early-enrolled student	-2.3	-0.5	-0.24
Late-enrolled student	-2.7	-1.7	-0.55
ESCS	2.1	2.56	3.28
not living with both parents	-1.37	-1.5	-1.57
student with siblings	0.15	-0.1	0
CR5	0.62	0.5	0.33
Fixed Effects	**North read**	**Center read**	**South read**
Intercept	24	31	33.6
Female	2.16	1.89	2.21
1^{st} generation imm	-3.9	-3.7	-1.6
2^{nd} generation imm	-3.7	-3.2	-1.16
Early-enrolled student	-2.04	-0.8	-0.37
Late-enrolled student	-3.4	-2.8	-1.16
ESCS	1.7	2.21	2.8
not living with both parents	-1	-1.4	-1
student with siblings	-0.5	-0.6	-0.7
CR5	0.56	0.45	0.36
	North	**Center**	**South**
Variance/covariance matrix of random effects	$\begin{pmatrix} 9.74 & 1.85 \\ 1.85 & 11.2 \end{pmatrix}$	$\begin{pmatrix} 14.8 & 5.31 \\ 5.31 & 12.7 \end{pmatrix}$	$\begin{pmatrix} 43.6 & 8.36 \\ 8.36 & 15.7 \end{pmatrix}$
Variance/covariance matrix of residuals	$\begin{pmatrix} 154 & 47 \\ 47 & 113 \end{pmatrix}$	$\begin{pmatrix} 182 & 64 \\ 64 & 159.6 \end{pmatrix}$	$\begin{pmatrix} 210 & 82 \\ 82 & 159 \end{pmatrix}$

Table 3: ML estimates of model (2) fitted for each macro-area.

Looking at the estimates of the three models, we observe that, in general, the coefficients of variables immigrants and late/early-enrolled students of the South are closer to zero than those of the North. Particularly, for immigrants students, this can be explained by the high presence of immigrants in the North respect to the South. The ESCS, instead, has a greater coefficients in the South than in the North (3.28 and 2.8 against 2.1 and 1.7), suggesting that in the South, the socio-cultural back-ground is very important in the students' achievements:

pupils with high ESCS have better results. Lastly, the score at grade 5 is more relevant in the North than in the South in both the topics (0.62 and 0.56 against 0.33 and 0.36), emphasizing a greater continuity in student performances.

Let's now look at the variance/covariance matrices of the random effects. The three matrices seem quite different: instead of the North and the Center, where the variances of the random effects of mathematics and reading are almost the same (respectively 9.74 vs 11.2 and 14.8 vs 12.7), in the South the variance of the random effects of mathematics is much higher than the reading one (43.6 vs 15.7), meaning that the school weighs more in mathematics than in reading. The coefficients of correlation between the two vectors of random effects are respectively 0.17 in the North, 0.39 in the Center and 0.32 in the South.

3.1.1 Comparing variance matrices

To test if there is really a significant difference between the three variance/covariance matrices of the three macro-areas, we use a distance-based test for homogeneity of multivariate dispersions.

Applying the method proposed in [2] to the three variance/covariance matrices estimated in model (2) and using the R package *vegan* (see [10]), we find that the means of the Euclidean distances between points and centroid within each group are 3.677 in the North, 4.238 in the Center and 6.329 in the South, showing that, as we saw below, the points of the South are more scattered. Similar results are obtained if we calculate the distances from the median within each group (repectively 3.645, 4.217 and 6.303). Both the tests ANOVA (with centroids and medians) give p-values less than 2.2e-16, proving that the three matrices are different, so that, there are different correlations between the school effects and different variance structures of random effects in the three macro-areas.

If we repeat this study on the variance/covariance matrices of the errors we notice the same trend of the random effects' matrices: the distributions and the variances of the residuals of the three macro-areas are different, the distances between the points and the centroids within each group are about 14 in the North, 15 in the Center and 16 in the South. The test ANOVA gives a p-value less than 2.2e-16. The big dispersion of the residuals in the South suggests that there is a big part of variability that remains unexplained.

3.2 Variables at School level across macro-areas

Now, it may be interesting to understand how the information at school level (number of students, percentage of female, immigrants..., private schools etc.) is correlated with the coefficients $\hat{\mathbf{b}}_j$ of the random effects. The variables at school level are divided into two groups: (i) the peers effects related to the composition of student body and (ii) managerial and structural features of the school. We fit now three bivariate linear models in which the outcome variables are the estimated school effects $\hat{\mathbf{b}}_j$ (models (2)) for each macro-area and the covariates are the variables at school level:

$$\hat{\mathbf{b}}_j^{(R)} = \boldsymbol{\gamma}_0^{(R)} + \sum_{k=1}^{K} \boldsymbol{\gamma}_k^{(R)} z_{jk}^{(R)} + \boldsymbol{\eta}_j^{(R)} \tag{3}$$

where

$R = \{North, Center, South\}$;

$j = 1, ..., J$ is the index of the school;

$\hat{\mathbf{b}}_j^{(R)}$ is the estimated random effect of the j-th school of models (2);

$z_{kj}^{(R)}$ is the value of the k-th predictor variable at school's level;

$\boldsymbol{\gamma}^{(R)} = (\boldsymbol{\gamma_0}^{(R)}, ..., \boldsymbol{\gamma}_L{}^{(R)})$ is the bivariate (L+1)-dimensional vector of parameters;

$\boldsymbol{\eta}_j{}^{(R)}$ is the zero mean gaussian error.

Using the Lasso regression method to select the variables (see [16]), we fit models with a reduced space of variables. Estimates of models (3) are reported in Table 4.

Model coefficients	North mat	Center mat	South mat
Intercept	$-1.467*$	$-3.899***$	$-3.946**$
Mean ESCS			$2.042***$
Female percentage	$0.038**$	$0.063**$	$0.066*$
1^{st} generation imm perc	-0.025		
2^{nd} generation imm perc		$0.146***$	
Early-enrolled student perc			
Late-enrolled student perc	$-0.057*$		$-0.208***$
Number of classes			
Number of students	$0.003**$	$0.004*$	
Average num of stud per class			$0.110*$
Private school		$-2.710**$	
IC			
Model coefficients	**North read**	**Center read**	**South read**
Intercept	0.156	-0.380	-0.753
Mean ESCS	$-1.727***$	-0.396	$1.095***$
Female percentage			$0.032*$
1^{st} generation imm perc			
2^{nd} generation imm perc		$0.153***$	
Early-enrolled student perc		$-0.203*$	$-0.097*$
Late-enrolled student perc	0.025		$-0.056*$
Number of classes			
Number of students	0.002.		
Average num of stud per class			
Private school	$-1.424***$	$-2.457**$	
IC			

Table 4: ML estimates of model (3) fitted to data of Northern, Central and Southern area, with the only variables selected by the LASSO. Asteriscs denote different levels of significance: . $0.01 < \text{p-val} < 0.1$; * $0.001 < \text{p-val} < 0.01$; ** $0.0001 < \text{p-val} < 0.001$; *** $\text{p-val} < 0.0001$.

The composition of the school's peers, such as female, early/late enrolled students percentage, weighs more in the South than in the North, in both the fields. The mean ESCS of the school is very signficant and weighs positively in the South, while it weighs negatively in the North. Lastly, being a private school is significant just in the North and in the Center and it weighs negatively, suggesting that Public schools are on average better than Private ones.

4 Bivariate three-level linear model: students nested in classes, nested in schools

At this point, it may be interesting to introduce also the class level into the model, therefore fitting a three-level linear model in which pupils are nested in classes, that are in turn nested in schools. In that way, we can analyze how much of the random effects is really due only to the school and how much only to the class. Previous studies show that the main differences in educational attainments elapse within schools, and not between schools: attending certain classes weighs more than attending certain schools (see [9]). We fit bivariate three-level model in which pupil $i, i = 1, ..., n_{lj}$; n $= \sum_{l,j} n_{lj}$ (first level) is in class $l, l = 1, ..., L_j$; L $= \sum_j L_j$ (second level) that is in school $j, j = 1, ..., J$ (third level):

$$\mathbf{y}_{ilj} = \boldsymbol{\beta}_0 + \sum_{k=1}^{K} \boldsymbol{\beta}_k x_{kilj} + \mathbf{b}_j + \mathbf{u}_l + \boldsymbol{\epsilon}_{ilj} \tag{4}$$

where

$\mathbf{y}_{ilj}$ is the bivariate outcome with mathematics and reading achievements of pupil i, in class l, in school j;

$\boldsymbol{\beta} = (\boldsymbol{\beta}_0, ..., \boldsymbol{\beta}_K)$ is the bivariate (K+1)-dimensional vector of coefficients;

$\mathbf{u} \sim N_2(\mathbf{0}, \Sigma_u)$ is the matrix of the two random effects (mathematics and reading) at class level;

$\mathbf{b} \sim N_2(\mathbf{0}, \Sigma_b)$ is the matrix of the two random effects (mathematics and reading) at school level;

$\boldsymbol{\epsilon} \sim N_2(\mathbf{0}, W)$ is the error,

with $\mathbf{u}$ independent of $\boldsymbol{\epsilon}$ and $\mathbf{b}$ independent of $\boldsymbol{\epsilon}$.

Since the coefficients of the variables at student level are similar to the previous ones, we focus the attention on the random effects. Table 5 shows the variance/covariance matrices of the two random effects.

School	Class
$\begin{pmatrix} 10.4 & 4.30 \\ 4.30 & 3.50 \end{pmatrix}$	$\begin{pmatrix} 17.4 & -1.02 \\ -1.02 & 18.4 \end{pmatrix}$
cor $= 0.712$	cor $= -0.05$

Table 5: Variance/Covariance matrices of the two random effects estimated by model (4).

We notice that the variances of the class effect are higher than the school ones, both in mathematics and reading (17.4 vs 10.4 in mathematics and 18.4 vs 3.50 in reading), suggesting again that the effect of the class is stronger than the school one. While the class effects in the two fields have about the same variances, regarding the school the effect in mathematics is stronger than the reading one. Looking at the two coefficients of correlation, it's clear that the effects of the school in the two fields are quite correlated (coef. 0.712) and they may represent better or worse schools, that give coherent contributes in the two topics. On the other hand, the two class' contributes are totally uncorrelated (coef. -0.05), so that there are classes that give a good contribute in reading and a bad one in mathematics and viceversa, probably depending on teachers.

We fit now model (4) in each one of the three macro-areas:

$$\mathbf{y}_{ilj}^{(R)} = \boldsymbol{\beta}_0^{(R)} + \sum_{k=1}^{K} \boldsymbol{\beta}_k^{(R)} x_{kilj} + \mathbf{b}_j^{(R)} + \mathbf{u}_l^{(R)} + \boldsymbol{\epsilon}_{ilj}^{(R)} \tag{5}$$

with $R = \{North, Center, South\}$.

Table 6 shows the variance/covariances matrices of the two effects, in the three areas.

	North	Center	South
School	$\begin{pmatrix} 4.99 & 1.82 \\ 1.82 & 1.65 \end{pmatrix}$	$\begin{pmatrix} 5.89 & 3.49 \\ 3.49 & 3.56 \end{pmatrix}$	$\begin{pmatrix} 15.9 & 5.74 \\ 5.74 & 4.54 \end{pmatrix}$
	cor $= 0.63$	cor $= 0.76$	cor $= 0.67$
Class	$\begin{pmatrix} 6.20 & -1.13 \\ -1.13 & 17.5 \end{pmatrix}$	$\begin{pmatrix} 13.9 & -0.88 \\ -0.88 & 18.1 \end{pmatrix}$	$\begin{pmatrix} 40.9 & 0.07 \\ 0.07 & 20 \end{pmatrix}$
	cor $= -0.10$	cor $= -0.05$	cor $= -0.002$

Table 6: Variance/Covariance matrices of the two random effects across macro-areas estimated by models (5).

Again, we see that the effects of the class are stronger than the school ones, indeed the variances of the class effects are higher. Regarding the school effects, the contribute in mathematics is always stronger than the reading one and overall the random effects weighs more in the South than in the North. Regarding the class effects, while in the North it seems that the class weighs more in reading than in mathematics, in the South it is the opposite. Again, the random effects in the South are stronger than in the North. Looking at the coefficients of correlation we confirm that the two school effects are quite correlated in all the three macro-areas, while the two class effects are snugly uncorrelated.

Figure (3) shows the bivariate school and class effects estimated by model (4).

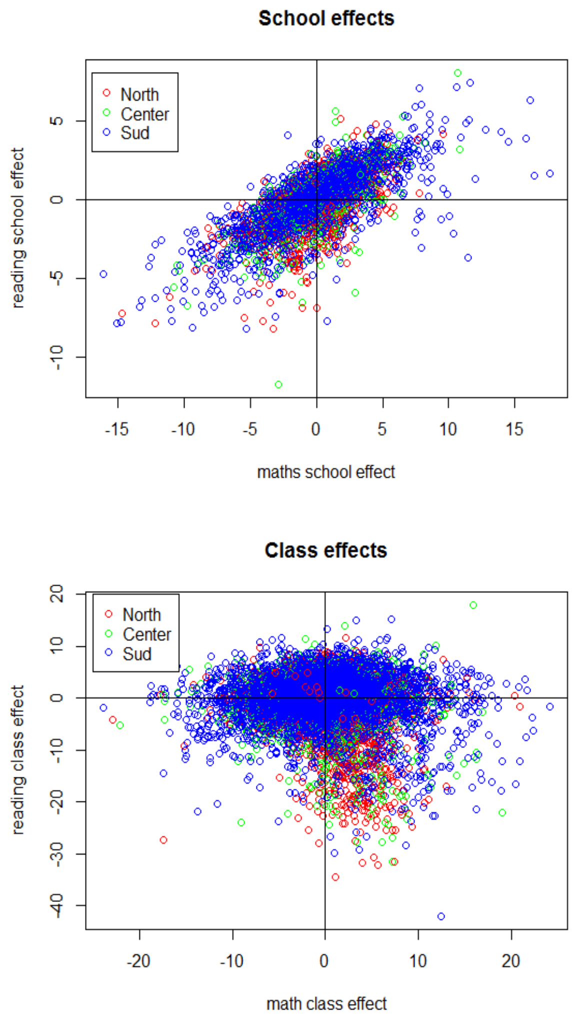

Figure 3: Plots of the school and class effects in mathematics and reading, estimated by model (4). Colours identify the three macro-areas: blue for the South, red for the North and green for the Center.

Regarding the school effects, it is clear that the variability of the value-added of schools in the South is much higher than the ones in the Center and in the North. This confirms that the impact the school has on students' performances in the South is stronger than the ones in the North and in the Center, that are similar. The same trend can be observed in the class effects, where the points of the South are more scattered, suggesting again that the effect of the class is stronger in the South than in the rest of Italy. Looking at the two shapes of points distributions, we see that there is a strong correlation (coefficient 0.712) between the school effects in reading and mathematics, that is most of the schools give coherent contributes in reading and mathematics, both positive or negative. On the other hand, there is not correlation (coefficient -0.05) between the two class effects. The class effect may be influenced by the teachers and a class may has a "good" teacher of mathematics and a "bad" one in reading or viceversa.

Now, it may be interesting to analyze if there is any kind of correlation between the school and class effects, such as if different kinds of schools (better or worse) contain different kinds of classes (better or worse). For this reason, from figure (3) we can identify the best schools in the first quarter, where schools give positive values-added in both the topics, and the worst ones in the third quarter, where schools give negative values-added in both the topics. In the same way, we can identify better and worse classes. We have computed the mean percentage of virtuous classes, that is classes with the effects in the first quarter, in each sector of schools, such as schools with values-added in all the four quarter. Figure (4) shows the distribution of these percentages.

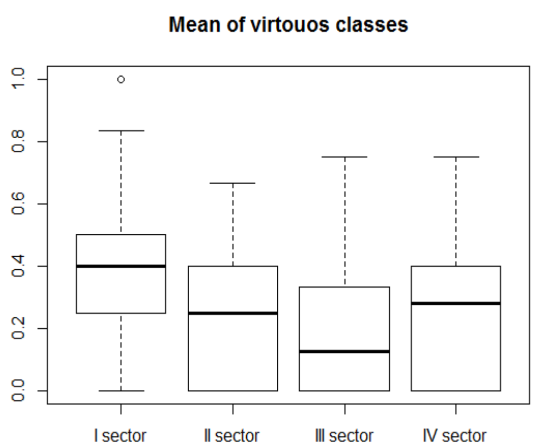

Figure 4: Boxplots of the mean percentages of virtuous classes stratified by sectors of schools.

The school effect is snugly uncorrelated with the effects of the classes that the school contains, indeed there are good percentages of virtuous classes in all kinds of schools. This means that there are schools with a positive (or negative) value-added containing classes with negative (or positive) mean value-added, that is the goodness of a school is uncorrelated with the goodness of the contained classes.

5 Concluding Remarks

This work explores how the students' achievements depend on students' characteristics and which may be the effects of attending specific schools. The data concerns INVALSI reading and mathematics test scores of students attending the first year of junior secondary school in the year 2012/2013.

The innovation of this work has been in the use of bivariate multilevel models, that allow us to explore the interaction and the correlation between the effects of schools and classes on reading and mathematics achievements and to compare the relationships between students' profile and their performances in the two fields. Previous studies have mainly analyzed school effects in reading and mathematics as separate, denying the possibility to have a complete view of the effectiveness of the school. With the bivariate models, we can estimate school/class effects in both the topics, analyze how they are correlated, if the class effects are coherent within schools and if they depend from the same variables at school/class level.

Univariate multilevel models had been already applied in previous studies in order to explain how the reading or mathematics (separately) achievements depend on students' characteristics. From this point of view, the results obtained by the new approach are in line with existing research. What has emerged are some recurrent dependencies between the outcome variable and regressors: females have, on average, better results than males in reading but worse one in mathematics, 1^{st} and 2^{nd} generation immigrants have more difficulties than Italian students, especially in reading, and being early/late-enrolled student decreases the average result. Furthermore, the pupil's ESCS, index of the socioeconomic status of the student, has a strong positive influence on the achievement and the CS5 (the student's achievement at grade 5) is positively correlated with the current score and this claims for a value-added (VA) specification of the statistical model. Lastly, from the models emerged that students of the Central and especially of the Southern Italy, have worse mean results than students of the North, showing big heterogeneities within the Country, that brought to the need to have three different models. What is interesting is that students' characteristics are different across the three geographical macro-areas, Northern, Central and Southern Italy, which can be considered as three different educational systems. The variables at student level that more influence the CRS and CMS are heterogeneous across macro-areas: the ESCS is much more relevant in the South than in the North and being 1^{st} and 2^{nd} generation immigrants decreases the mean result less in the South than in the North.

As anticipated, the new contribution of this approach has been in the estimates of the bivariate random effects, with their correlations. The school effect, defined as the effect of attending a specific school on a student's test score, has been modeled as a random effect $\hat{b}_j$ and has been regressed against a school level variables with the aim of characterizing the features of those schools that exert a positive/negative effect on academic performance. Even the school effects are different across macro-areas: in the South, they are much more scattered, suggesting that the school effect is much stronger. Therefore, while being private or public school influences the school effect in the North, in the South the mean ESCS of the school is one of the most relevant variables that adds positive value-added, showing that in the South the differences between schools tend to increase the inequalities between disadvantaged and advantaged students. Furthermore, it emerged that, in Italy, the school effect in mathematics is much stronger than the one in reading. In the macro-areas, this behavior occurs also in the South of Italy, while in the North is less pronounced. The coefficients of correlation show a coherency between the school effects in the two topics, proving that generally the contributes of the school in reading and mathematics are positively correlated and this defines which are "good" ("bad") schools. Therefore, it is possible to identify "good" ("bad") schools, knowing that they give positive (negative) value-added in both the topics.

In the same way, also the class effects $\hat{u}_l$ have been modeled, showing a trend that is similar to the one of school effects, even stronger. The main difference between the two effects is that the reading and mathematics class effects are not correlated like the school ones, denying the possibility to identify "good" ("bad") classes. This arises from the fact that such kind of contributes at class level are probably given by the teachers, and students in a class may have a good teacher of mathematics and a less good one of reading or viceversa, without any kind of correlation. Anyway, the class effects follow similar trends of the school effects: the contributes in mathematics are more pronounced than in reading and in the South they are again stronger than in the North in both the topics, being different across macro-areas. At last, it has been impossible also to find a correlation between school and contained class effects, that prove to be independent. This means that good (or bad) schools may contain bad (or good) classes. From the univariate cases, instead, we obtained that there was a dependence between school and classes (i.e. school with a positive value-added in mathematics/reading contained classes with a mean positive value-added in mathematics/reading). In the bivariate case, we loose this dependence because within each school, the contributes that classes give in mathematics and reading are very heterogeneous and not coherent with the contribute of the school. Therefore, the effectiveness of schools is independent from the effectiveness of the contained classes. We can therefore conclude that sometimes it is possible to identify and choose a good school, but within it there is still variability between and within classes and this variability changes across the three geographical macro-areas.

Further studies may be done to explore other aspects of the Italian educational system. It could be interesting to deepen the geographical differences, analyzing the districts; to explore if there is a sort of homogeneity of the variables within the schools and within the classes; to discover how much the teachers influence the class effects; to provide a way to treat the missing data and, particularly, to explore if there is a way to reduce the geographical heterogeneity, in order to provide a good educational level for all Italian students.

Policy and Managerial Implications

The results of this analysis bring to the need to make some considerations about policy and managerial implications.

First of all, the use of school-effect estimates should be done in a proper way. As we have seen, the contributes that schools give in reading and mathematics may be different and sometimes even opposite. It can be the case that the school-effect is positive in mathematics and negative in reading, or viceversa. In this sense, policy makers should be clear about what they want to promote and reward. Sanctions and rewards can induce substantial consequences. For this reason, we should be sure about how the estimates are robust and stable, so that reliable to imply such policy and managerial consequences (see [18]).

Secondarily, more research is needed for brightening the factors behind the differences in the schools' profiles and for explaining the coherence/incoherence between the effects in mathematics and reading within schools. As we have seen, the two effects are affected by different variables. This heterogeneity may be due to various aspects, such as students' attitudes, teachers' effectiveness or schools' strategies and activities.

Lastly, our findings highlight that schools are intrinsically multi-output organizations. While this is acknowledged for universities, that produce teaching and research (see [8]), this is under-investigated for schools. Our study is a first step in this direction and in the next future we will concentrate on typical implications for multi-output organizations, such as checking for economies of scope and developing new indicators for unmeasured outputs (as non-cognitive skills).

Acknowledgments

This work is conducted within FARB - Public Management Research: Health and Education Systems Assessment, funded by Politecnico di Milano. The authors are grateful to Invalsi for having provided the original dataset.

References

[1] T. Agasisti, F. Ieva, and A. M. Paganoni. Heterogeneity, school-effects and achievement gaps across italian regions: further evidence from statistical modeling. *MOX, Dipartimento di Matematica F. Brioschi, Politecnico di Milano, https://www.mate.polimi.it/biblioteca/add/qmox/07-2014.pdf*, January 31, 2014.

[2] M. J. Anderson. Distance-based tests for homogeneity of multivariate dispersions. *Biometrics*, 62(1):245–253, 2006.

[3] Jaap Dronkers and Péter Robert. Differences in scholastic achievement of public, private government-dependent, and private independent schools a cross-national analysis. *Educational Policy*, 22(4):541–577, 2008.

[4] John Fox. Linear mixed models. *Appendix to An R and S-PLUS Companion to Applied Regression, URL http://cran. r-project. org/doc/contrib/Fox-Companion/appendix-mixed-models. pdf*, 2002.

[5] John Fox and Sanford Weisberg. *An R companion to applied regression.* Sage, 2010.

[6] A. R. Gilmour, B. R. Cullis, S. J. Welha, and Thompson. R. 2002 asreml reference manual 2nd edition, release 1.0 nsw. *Agriculture Biometrical Bulletin 3, NSW Agriculture, Locked Bag, Orange, NSW 2800, Australia.*

[7] E. A. Hanushek and L. Woessmann. *Handbook of the Economics of Education, Volume 3*, chapter The Economics of International. Differences in Educational Achievement, pages 91–192. Elsevier B.V., 2011.

[8] Geraint Johnes, Jill Johnes, and Emmanuel Thanassoulis. An analysis of costs in institutions of higher education in england. *Studies in Higher Education*, 33(5):527–549, 2008.

[9] Chiara Masci. Bivariate multilevel models for the analysis of reading and maths pupils' achievements. Tesi di laurea magistrale, Politecnico di Milano, 2013/2014.

[10] Jari Oksanen, F. Guillaume Blanchet, Roeland Kindt, Pierre Legendre, Peter R. Minchin, R. B. O'Hara, Gavin L. Simpson, Peter Solymos, M. Henry H. Stevens, and Helene Wagner. *vegan: Community Ecology Package*, 2013. R package version 2.0-10.

[11] Ingram Olkin. *Contributions to probability and statistics: essays in honor of Harold Hotelling.* Stanford University Press, 1960.

[12] J.J. C. Pinheiro and D. M. Bates. *Mixed-effects Models in S and S-PLUS.*

[13] Jose Pinheiro, Douglas Bates, Saikat DebRoy, Deepayan Sarkar, and R Core Team. *nlme: Linear and Nonlinear Mixed Effects Models*, 2014. R package version 3.1-117.

[14] R Core Team. *R: A Language and Environment for Statistical Computing.* R Foundation for Statistical Computing, Vienna, Austria, 2014.

[15] John D Spurrier. On the null distribution of the kruskal–wallis statistic. *Nonparametric Statistics*, 15(6):685–691, 2003.

[16] Robert Tibshirani. Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society. Series B (Methodological)*, 58(1):267–288, 1996.

[17] Anna Vignoles, Rosalind Levacic, James Walker, Stephen Machin, and David Reynolds. *The relationship between resource allocation and pupil attainment: a review.* Centre for the Economics of Education, London School of Economics and Political Science, 2000.

[18] J Douglas Willms and Stephen W Raudenbush. A longitudinal hierarchical linear model for estimating school effects and their stability. *Journal of educational measurement*, 26(3):209–232, 1989.

MOX Technical Reports, last issues

Dipartimento di Matematica
Politecnico di Milano, Via Bonardi 9 - 20133 Milano (Italy)

17/2015 Nestola, M.G.C.; Faggiano, E.; Vergara, C.; Lancellotti, R.M.; Ippolito, S,; Filippi, S.; Quartero
Computational comparison of aortic root stresses in presence of stentless and stented aortic valve bio-prostheses

16/2015 Fumagalli, I.; Manzoni, A.; Parolini, N.; Verani, M.
Reduced basis approximation and a posteriori error estimates for parametrized elliptic eigenvalue problems

15/2015 Taffetani, M.; de Falco, C.; Penta, R.; Ambrosi, D.; Ciarletta, P.
Biomechanical modelling in nanomedicine: multiscale approaches and future challenges

14/2015 Canuto, C.; Nochetto, R.H.; Stevenson R,; Verani, M.
Convergence and Optimality of hp-AFEM

13/2015 Bartezzaghi, A.; Dedè, L.; Quarteroni, A.;
Isogeometric Analysis of High Order Partial Differential Equations on Surfaces

12/2015 Antonietti, P. F.; Beirao da Veiga, L.; Scacchi, S.; Verani, M.
A C^1 virtual element method for the Cahn-Hilliard equation with polygonal meshes

11/2015 Antonietti, P. F.; Marcati, C.; Mazzieri, I.; Quarteroni, A.
High order discontinuous Galerkin methods on simplicial elements for the elastodynamics equation

10/2015 Antonietti, P. F.; Grasselli, M.; Stangalino, S.; Verani, M.
Discontinuous Galerkin approximation of linear parabolic problems with dynamic boundary conditions

06/2015 Perotto, S.; Zilio, A.
Space-time adaptive hierarchical model reduction for parabolic equations

09/2015 Ghiglietti, A.; Ieva, F.; Paganoni, A.M.; Aletti, G.
On linear regression models in infinite dimensional spaces with scalar response