Politecnico di Milano

Scuola di Ingegneria Industriale e dell'Informazione

Corso di Laurea Magistrale in Ingegneria Matematica

Bivariate Multilevel Models for the Analysis of Reading and Maths Pupils' Achievements

Relatore:
Prof. Anna Paganoni

Co-Relatore:
Dott. Francesca Ieva

Tesi di Laurea Magistrale di:
Chiara Masci
Matr. 801798

Anno Accademico 2013-2014

Contents

List of Figures

List of Tables

Abstract

This work aims to catch the differences in educational attainments between students and across classes and schools they are grouped by, in the context of Italian educational system. The purpose is to identify a relationship between pupils' maths and reading test scores and the characteristics of students themselves, stratifying for classes, schools and geographical area. 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 multilevel linear mixed models, in which the outcome variable is bivariate: reading and maths achievements. By means of these models, it is possible to estimate statistically significant "school and class effects" after adjusting for pupil's characteristics, i.e. the positive/negative impact of attending a specific school or class on student's test score, and to identify which are the characteristics of the students that more influence their performances, both in mathematics and reading. The results show that big discrepancies elapse between the three geographial macro-areas (Northern, Central and Southern Italy), where school/class effects and relevant student's features are very heterogeneous.

KEYWORDS: Pupils' achievement; Multilevel models; Bivariate models; School and class effect; Value-added.

Sommario

L'obiettivo di questo lavoro è di cogliere le differenze tra i rendimenti scolastici degli studenti e tra le scuole e le classi in cui essi sono raggruppati, nel contesto del sistema educativo Italiano. Lo scopo è di identificare una relazione tra i risultati dei test di italiano e matematica degli studenti e le caratteristiche di questi ultimi, raggruppati per classi, scuole e aree geografiche. Il dataset in questione contiene informazioni dettagliate su più di 500,000 bambini al primo anno di scuola media, nell'anno scolastico 2012/2013, fornite dall'Istituto Nazionale per la Valutazione del Sistema Educativo di Istruzione e di Formazione (INVALSI). L'innovazione di questo lavoro risiede nell'uso di modelli lineari multivariati a effetti misti, nei quali la variabile risposta è bivariata: risultati dei test di matematica e italiano. Per mezzo di questi modelli, è possibile stimare "effetti scuola e classe" statisticamente significativi dopo aver aggiustato rispetto alle covariate bambino, come per esempio l'impatto positivo/negativo di frequentare una data scuola o classe sul rendimento dello studente, e identificare quali sono le caratteristiche degli alunni che più influenzano la loro performance, in matematica e in italiano. I risultati mostrano che ci sono grandi discrepanze tra le tre macro-aree geografiche (Nord, Centro e Sud Italia), che sono caratterizzate da effetti scuola/classe e caratteristiche rilevanti degli studenti molto eterogenei.

Introduzione

L'analisi delle differenze dei rendimenti scolastici tra gruppi di studenti e tra scuole sta diventando, negli ultimi anni, sempre più interessante. A tal proposito vengono fatti numerosi studi per testare e migliorare il sistema educativo e per capire quali variabili lo definiscono (vedi [7],[12],[30],[27]). Negli stati più industrializzati, esistono istituzioni che, sottoponendo gli studenti a questionari comuni e raccogliendo informazioni sulle scuole e sulle classi, cercano di testare i rendimenti degli studenti e di capire quali sono gli aspetti che più influenzano la loro prestazione. Il Programma per la Valutazione Internazionale dell'Allievo (PISA) è stato promosso nel 2000 dall'Organizzazione per la Cooperazione e lo Sviluppo Economico (OCSE) per analizzare il livello di istruzione dei ragazzi negli stati più industrializzati. Lo scopo è quello di confrontare i risultati dei test, per identificare quali sono gli stati con i migliori rendimenti e quali sono le variabili, le caratteristiche e gli aspetti delle istituzioni scolastiche che permettono loro di avere tali risultati. Tipicamente, i test coinvolgono tra i 4,500 e i 10,000 studenti in ogni stato.

In Italia, l'Istituto Nazionale per la Valutazione del Sistema educativo e dell'Istruzione (INVALSI), fondato nel 2007, valuta gli studenti nelle loro prestazioni in matematica e in italiano in diversi stadi: alla fine del secondo e del quinto anno di scuola elementare (circa a 7 e 10 anni rispettivamente), alla fine del primo e del terzo anno di scuola media (11 e 13 anni) e alla fine del secondo anno di scuola superiore (15 anni).

Agli studenti viene chiesto di rispondere a domande, uguali per tutti, sia aperte che a risposta multipla, che testano le loro conoscenze in matematica e italiano. Questo è un modo di valutare conoscenze e metodi di ragionamento che i ragazzi avrebbero dovuto imparare nel loro percorso scolastico. Inoltre, viene chiesto di rispondere a domande circa loro stessi, la loro famiglia, il livello di istruzione dei genitori e la loro situazione socio-economica (vedi [4],[6],[17],[1]). Abbiamo a disposizione due dataset separati, il primo contenente i risultati di matematica e il secondo quelli di italiano, seguiti dalle informazioni sugli studenti, sulle classi e sulle scuole. Questi due dataset sono stati forniti in due momenti diversi: prima abbiamo ricevuto quello di matematica (già precedentemente studiato) e poi quello di italiano.

Gli obiettivi sono (i) esaminare la relazione tra le caratteristiche degli studenti, come profilo, situazione socio-culturale, risorse culturali, e i loro rendimenti scolastici, (ii) chiarire se ci sono differenze educative tra le scuole e tra le tre macro-aree geografiche dell'Italia (Nord, Cento e Sud) e (iii) scoprire come l'effetto della scuola è più/meno accentuato per certi profili di studenti. Gli strumenti statistici utili per svolgere questi tipi di studi sono specialmente modelli lineari a effetti misti, univariati e bivariati.

Il primo passo è creare un dataset congiunto in cui per ogni studente abbiamo i risultati in entrambe le materie. In questo modo, possiamo fare considerazioni e confronti, implementando modelli sullo stesso insieme di studenti.

Sono già stati fatti studi sui rendimenti di matematica, applicando modelli lineari a effetti misti univariati, per analizzare come la variabile risposta

(risultati di matematica) dipende dalle covariate e quali sono gli effetti della classe e della scuola sui rendimenti degli studenti (vedi [2]). Sono emerse grandi differenze tra Nord, Centro e Sud Italia, suggerendo il bisogno di implementare tre modelli separati per descrivere questi fenomeni completamente differenti. Quindi, la prima parte del lavoro sarà dedicata allo studio dei risultati di italiano, così da poter poi confrontare i risultati delle due materie. Visto che, comunque, c'è una forte correlazione tra le due variabili risposta, il fulcro del lavoro sarà studiare modelli lineari a effetti misti multivariati, nei quali la risposta è bivariata: risultati di matematica e italiano. Indagheremo poi se questo nuovo approccio apporta del valore-aggiunto ai modelli e spiega la relazione che intercorre tra gli effetti scuola/classe delle due materie.

Il lavoro è organizzato come segue: la Sezione 2 presenta il dataset; nelle Sezioni 3 e 4 si fanno studi sui risultati di italiano rispettivamente in Italia e nelle tre macro-aree con modelli lineari a due e tre livelli; nella Sezione 5 si introducono i modelli lineari a effetti misti bivariati per i risultati di italiano e matematica e nella Sezione 7 si implementano modelli univariati e bivariati a due livelli in cui gli studenti sono raggruppati solo per classi. Tutte le analisi sono state fatte usando il software statistico R (vedi [22]), tranne i modelli lineari a effetti misti bivariati, che sono stati implementati usando il software AsReml (vedi [11]).

1 Introduction

Nowadays, the analysis of the differences in educational attainments between groups of students and across schools is becoming increasingly interesting. Studies on this subject are made in order to test and improve the educational system and to understand which variables determine it (see [7],[12],[30],[27]). In the most industrialized countries, exist institutions that, referring students to common questionnaires and collecting information about schools and classes, aim to test pupils' achievement and to understand which are the aspects that more influence the performances. 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 2000 in order to analyze the educational level of the teenagers in the main industrialized countries. The purpose is to compare the results of the tests, in order to detect which are the countries with best and worst performances and which are the variables, the characteristics and the aspects of their scholastic institutions that permit them to have such results. Tipically, the tests involve between 4,500 and 10,000 students in each country.

In Italy, the Italian Institute for the Evaluation of Educational System (INVALSI), founded in 2007, assesses students in their reading and mathematics abilities at different stages: at the end of the second and fifth year of primary school (about at age 7 and 10, respectively), at the end of the first and third year of lower secondary school (age 11 and 13) and at the end of the second year of upper secondary school (age 15).

Students are requested to answer questions, the same for everyone, 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 (see [4],[6],[17],[1]). Our resources are two separate set of data, the former containing the mathematics achievements and the latter the reading ones, followed by the information about students, classes and schools in Italy. We obtained the two dataset in two different moments: firstly, we received the mathematics one (that have already been explored) and secondarily the reading one.

The aims are (i) to examine the relationship between pupils' characteristics, such as profile, socio-cultural background, household, cultural resources, and pupil's achievement, (ii) to detect if educational differences elapse between different schools and between the three geographical macro-areas of Italy (Northern, Central and Southern) and (iii) to discover how the school effect is stronger/weaker for specific types of students' profile. The statistical tools requested to make this kind of studies are especially multilevel linear mixed models, both univariates and bivariates.

The first step is to create a joined dataset in which for each student we have both his\her achievement in mathematics and reading. In this way, we can

make considerations and comparisons, fitting the models on the same sample of students.

Studies have already been made on the mathematics achievements, applying univariate multilevel linear mixed models, to analyze how the outcome variable (mathematics achievement) depends on the covariates and which are the school and the class impacts on student's achievements (see [2]). Big differences elapsed between North, Center and South of Italy, emphasizing the need to have three different models to explain the completely different phenomena. Therefore, the first part of the work will be dedicated to the study of the reading achievements, so that, we can start comparing the results of the two subjects. However, since a strong correlation exists between the two outcome variables, the cornerstone of the work is the study of bivariate linear mixed models in which the outcome consists in a bivariate answer: mathematics and reading scores. We will detect if this new approach may bring some value-added to the models and explain the relationship between the school/class-effects of the two subjects.

The work is organized as follows: Section 2 presents the dataset; in Section 3 and 4 we make studies on the reading achievement respectively in Italy and across macro-areas, by means of twoand three-level linear mixed models; in Section 5 we introduce the bivariate multilevel linear mixed models for mathamatis and reading achievements; Section 6 is dedicated to the analysis of the school effects and in Section 7 we focus the attention on models, both univariate and bivariate, in which pupils are nested in classes.

All the analysis are made using the statistical software R (see [22]), except the bivariate multilevel linear mixed models that are implemented using the software AsReml (see [11]).

2 Background and Motivations

We have two initial 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 is built from the mathematics test and the latter from the reading one. For each student, we have his/her achievements both in reading and mathematics tests. The information are nested in different levels: pupils are nested within classes that are nested within schools. We have information for each of these levels. Part of the variables are the same in the two dataset and were yet studied for the mathematics achievements, but the "reading dataset" contains new variables that bring other information. The "reading dataset" contains information about 510,933 students and the "mathematics one" about 509,371. As introduced before, we create a "complete dataset", collecting 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. A merge of the two dataset is possible thanks to the anonymous student ID that is known for each pupil and that allows us to distinguish and individuate students. We obtain a new dataset containing 507,229 students, for whom both the achievements in maths and reading are known, and 50 variables, loosing, fortunately, very few individuals.

2.1 The Dataset

Several information are provided at pupil, class and school level and they create the set of covariates. When considering characteristic referred to the single student, 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), variables on his/her school performances (school score of reading and mathematics, written and oral). 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 (The Organisation for Economic Co-operation and Development)-PISA framework, 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 to have mean zero and variance one. The minimum and maximum observed values in the Invalsi dataset are -3.11 and 2.67. 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 av-

erage 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, 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, i.e., the score in the Mathematics and Reading standardized test administered by Invalsi) are expressed as "cheating-corrected" scores (CMS and CRS): 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. Among data, there are also the scores in the Maths and Reading tests at grade 5 of the previous year, 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.

Unfortunately, there are lots of missing data in the score at grade 5, both in mathematics and reading achievements. This kind of data can be 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.

Thare is also a different way to treat the missing data, instead of delete them. It's possible to impute missing data using different methods: the simplest case is to substitute some statistically meaningful data available; more complex is the method of Multiple Imputation; lastly, there are iterative methods (EM) that allow to obtain estimates for the parameter of interest (see [5],[23],[24]).

Hereafter, all the analysis are made on this reduced dataset with 221,529 students. 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
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	[0:12]	Written reading score	9.41	2.74
Student	[0:12]	Oral reading score	6.80	1.13
Student	[0:12]	Written maths score	9.48	2.75
Student	[0:12]	Oral maths score	6.88	1.35
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	[0:100]	MS-Math Score	48	17.45
Outcome	[0:100]	CMS-Math Score corrected for Cheating	47.4	17.67
Outcome	[0:100]	RS-Reading Score	67	14.58
Outcome	[0:100]	CRS-Reading Score corrected for Cheating	65	14.65
Outcome	[0:100]	CMS5-5^{th} year Primary school math score	70.5	16.30
Outcome	[0:100]	CRS5-5^{th} year Primary school reading score	74.5	13.50

Table 1: variables of the database

3 Reading Achievements in Italy

As introduced before, we start analyzing the correlation between the reading achievements of students and the available information about pupils, classes and schools. The variable of interest of our analysis is the Reading Score (corrected for cheating, namely CRS) of students attending the first year of junior secondary school. The purpose is to detect which student's characteristics have a positive and which have a negative influence on the achievements and to estimate the school impacts on students' achievement, so that, how much attending a particulary school has a positive or a negative effect. The way to model this correlation is given by the multilevel linear mixed models (see [20],[9],[8]), that allow us, among others, to decompose the total variability in parts that vary between schools, classes and pupils. Univariate multilevel linear mixed models are implemented using the package *nlme* in R (see [21]).

3.1 Two-level Linear Mixed Model

The first model proposed is a two-level school effectiveness model in which we consider the variables at student level (level 1) with a random effect on schools (level 2). We detect how the answer variable, the students' reading achievement, is correlated with the characteristics of students and which is the value-added that the school gives to that achievement. Therefore, we use a two-level linear mixed model in which pupil i, i = 1,..., n_{lj}; $n = \sum_{l,j} n_{lj}$ (first level) is nested within school j (second level), j = 1,..., J:

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

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

where

y_{ij} is the reading test achievement of student i within school j;

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

$\beta = (\beta_0, ..., \beta_K)$ is the (K+1) dimensional vector parameters to be estimated;

b_j is the random effect of the j-th school and it's assumed to be Gaussian distributed and independent to any predictor variables that are included in the model;

ϵ_{ij} is the zero mean Gaussian error.

The histogram of the answer variable, the reading test achievement corrected for cheating (CRS), is reported in Figure 1.

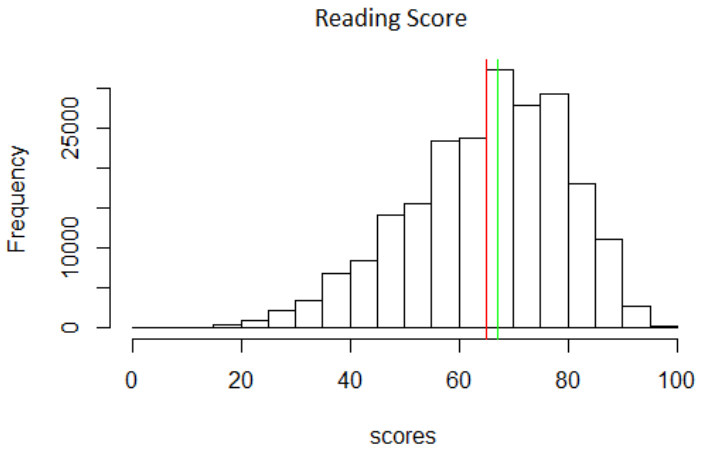

Figure 1: Histogram of Corrected Reading Score of pupils in the Invalsi database. The red line refers to the mean, the green one to the median.

Before analyzing the results of the model, it's interesting to see if there are some evident differences between groups of students. In Figure 2 are reported boxplots of CRS, stratified by some student level variables.

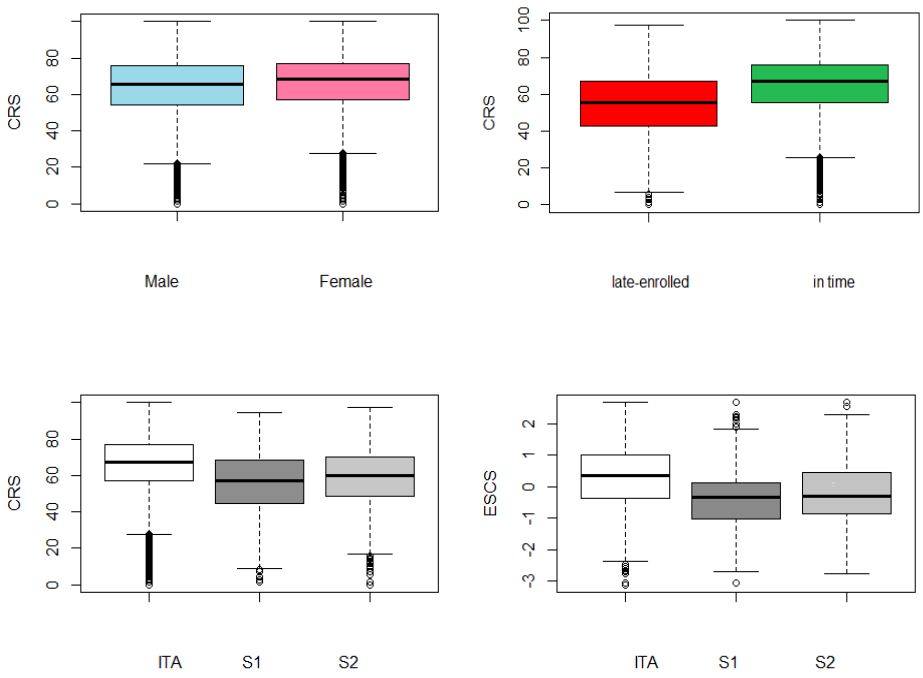

Figure 2: Boxplots of CRS stratified by gender, being late enrolled and being first or second generation immigrant student. The last stratification is also adopted for ESCS.

From the first boxplot, we can deduce a slight prevalence of the outcome of females over those of males, that means that females have better results than males, contrarily to what is obtained in mathematics. Since we can not test the normality of the data because the dimensions are too high, we use the Wilcoxon test for the difference between the medians, that is the non-parametric equivalent of the t-test, obtaining a p-value less that $2.2e-16$. From the second one, we see that late-enrolled students, i.e. student that must repeat one grade, or students admitted at school one year later if immigrants, have worst results than "in time" students (p-value of Wilcoxon test less than $2.2e-16$). Regarding the foreign students, from the last two boxplots, it's clear that the 1^{st} and 2^{nd}generations of immigrants have worst performances than the Italians (p-values of Wilcoxon test less than $2.2e-16$), but this is also strictly connected with the gap between their respective ESCS indices; there is a significant positive correlation between the CRS and the ESCS (coefficient 0.27). Usually, immigrants have a low ESCS index. Except the first boxplot (where in mathematics males have better results than females), these results are the same of the ones obtained for the mathematics attainment: immigrants and late-enrolled students have worst

performances and the ESCS index seems positively relevant.

The estimates of model (1) are reported in Table 2.

Fixed Effect	Estimate	Standard Error
Intercept	$24.333 * **$	0.183
Female	$2.091 * **$	0.051
1^{st} generation immigrant	$-3.449 * **$	0.142
2^{nd} generation immigrant	$-3.201 * **$	0.122
South	$-4.616 * **$	0.174
Center	$-1.165 * **$	0.215
Early-enrolled student	$-0.699 * **$	0.204
Late-enrolled student	$-3.372 * **$	0.171
ESCS	$1.986 * **$	0.028
Not living with both parents	$-1.008 * **$	0.078
Student with siblings	$-0.579 * **$	0.070
written reading score	0.001	0.002
oral reading score	$0.024 * **$	0.002
CRS5	$0.552 * **$	0.001
Random Effect		
σ_b	4.383	
σ_ϵ	11.448	
VPC	12.7%	
Size		
Number of observations	$221,529$	
Number of groups	$3,920$	

Table 2: ML estimates (with standard errors) for model (1), fitted to the dataset. Asteriscs 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.

All the variables seem to be pretty significant, except the written and oral reading score, that have respectively very high p-value and correlation coefficients very small (respect to the values that the variable assumes); this suggests that there is no a strong link between the usual academic attainment of the student at school and his/her Invalsi test's result.

Being a female increases the mean result of reading of 2 points, as we expected from the boxplot of Figure 2. Also, as we could expect, being a 1^{st} or 2^{nd} generation of immigrants involves a lowering of the mean result ($\sim$ -3),

since being strangers involves a worst understanding of the Italian language. Being a late-enrolled student reduces the mean result of 3.3 points. The positive coefficient (0.55) between the actual Invalsi score and the Invalsi score of the 5^{th} year of primary school (CRS5) suggests that students that had good results at the primary school, continue to have good result also in the junior secondary school. The positive ESCS coefficient (1.98) tells us that students with a high ESCS, that means good parents' occupation and educational titles and a substantial amount of "cultural resources" at home, have better results that students with a lower ESCS. Lastly, the difference between geographical macro-areas is interesting: respect to the reference variable (being at North), attending a school in the Center of Italy decreases the mean result of 1 point and, especially, attending a school in the South decreases the mean result of more than 4 points.

The positive/negative correlations between the output and the covariates are quite the same than in mathematics achievement, except, as we introduced before, the correlation between sex and score: females have better performances in reading skills, while males are better at maths.

In this model, the total variability varies between schools and between pupils. The Variance Partition Coefficient (VPC) captured by Random Effects is obtained as the proportion of random effects variance over the total variation

$$\frac{\sigma_b^2}{\sigma_b^2 + \sigma_\epsilon^2} \tag{3}$$

In model (1), 12.7% of the total variance is explained by the variance of random effects, that is, the variance between different schools. This suggests that the educational level is not the same in all the schools and a consistent part of the total variance of performances is explained by attending different schools.

3.1.1 Variables at School level

Now, we would like to understand how the information at school level (number of students, percentage of female, immigrants..., private schools etc.) is correlated with the coefficients 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 use these variables to model the factors affecting the estimated random effects, through a linear model:

$$\hat{b}_j = \gamma_0 + \sum_{l=1}^{L} \gamma_l z_{lj} + \eta_j \tag{4}$$

$$\eta_j \sim N(0, \sigma_\eta^2) \tag{5}$$

where

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

$\hat{b}_j$ is the estimated randon effect of the j-th school of model (1);

z_{lj} is the value of the l-th predictor variable at school's level;

$\gamma = (\gamma_0, ..., \gamma_L)$ is the (L+1) dimensional vector of parameters;

η_j is the zero mean gaussian error.

The histogram of the estimated Random Effect coefficients $\hat{b}_j$ is reported in Figure 3.

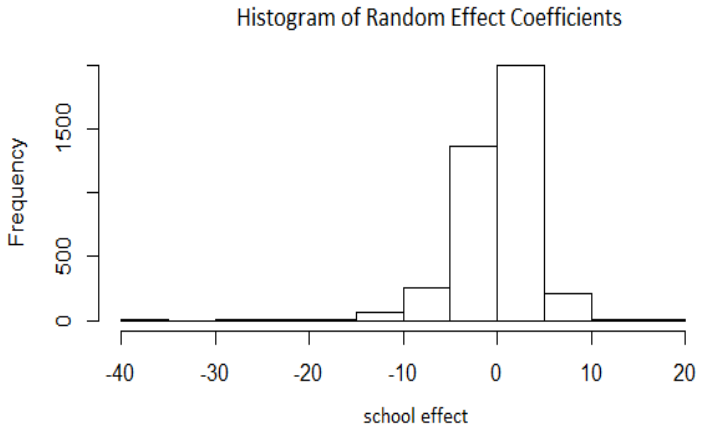

Figure 3: Histogram of the estimated Random Effect coefficients.

The ML estimates of model (4) are reported in Table 3.

Model coefficients	Estimate
Intercept	$-2.336 * *$
Mean ESCS	$-0.282.$
Female percentage	$0.028 * *$
1^{st} generation immigrant percentage	$0.046 * *$
2^{nd} generation immigrant percentage	$0.042*$
Early-enrolled student percentage	-0.004
Late-enrolled student percentage	-0.005
Number of classes	0.015
Number of students	0.000
Average number of students per class	0.030
Private school	$-2.199 * **$
"Istituto Comprensivo"	0.201

Table 3: ML estimates for model (4). Asteriscs 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.

The only variables that seem to have a significant influence on the "school effect" are the index of Private or Public school and the percentages of females and immigrants. Anyway, the very low R^2s of the regression (about 4 %), suggests that a lot of variability remains unexplained considering the measurable variables only. Moreover, the design matrices present a high correlation among their columns. In order to solve this issue, we fit a Lasso regression model (see [26]) to the random effects estimates of model (1):

$$\lambda = argmin_{\gamma}\left(\gamma_0 + \sum_{l=1}^{L}\gamma_l z_{lj}\right)^2 \tag{6}$$

subject to

$$\sum_{l}|\gamma_l| \leq \gamma \tag{7}$$

The Lasso regression is a variable selection algorithm and it produces some coefficients that are exactly 0, that is, it's a way to choose which are the most important variables in order to refit the model using only these variables and avoiding wrong estimates, given by the correlation between all the variables. Table 4 shows the results of the model selected by Lasso regression, that is implemented in the R package *lars* (see [13]).

Lasso Model coefficients	Estimates
Intercept	$-1.286219 * *$
Female percentage	$0.024781 * *$
1^{st} generation immigrant percentage	$0.038944 * *$
2^{nd} generation immigrant percentage	$0.045787 * *$
Private school	$-2.865086 * **$

Table 4: ML estimates for model (4), fitted on the reduced space of variables selected by Lasso. Asteriscs 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.

From the variable selection algorithm results that the covariates that seem more relevant are those that describe the composition of student body (percentages of females and immigrants) and the index of private school. In particular, the percentages of females and immigrants are positively associated with the school effect, instead of being a "Private school", that reduces the mean value-added of the school of 2.8 points, suggesting that public schools are generally better than the private ones.

3.2 Three-level Linear Mixed Model

Reverting to model (1), we observe that 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. In other words, part of the school effect is actually driven by the differences between classes of the same school and so, exploring the variance between classes (within school) can add explanatory power to our model. Therefore, we use a three-level linear mixed 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; $L = \sum_k L_j$(second level) that is in school j, j = 1,..., J:

$$y_{ilj} = \beta_0 + \sum_{k=1}^{K} \beta_k x_{kilj} + b_j + u_{lj} + \epsilon_{ilj} \tag{8}$$

$$b_j \sim N(0, \sigma^2_{School}), u_{lj} \sim N(0, \sigma^2_{Class}), \epsilon_{ilj} \sim N(0, \sigma^2_{\epsilon}) \tag{9}$$

where

y_{ilj} is the CRS of the student i, in the class l, in the school j;

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

$\beta = (\beta_0, ..., \beta_K)$ is the (K+1)-dimensional vector of parameter;

b_j is the random effect for the j-th school;

u_{lj} is the random effect for the l-th class in the j-th school;

ϵ_{ilj} is the zero mean gaussian error.

The estimates of model (8) are reported in table 5.

Fixed Effect	Estimate	Standard Error
Intercept	$22.885 * **$	0.190
Female	$2.110 * **$	0.048
1^{st} generation immigrant	$-3.494 * **$	0.128
2^{nd} generation immigrant	$-3.247 * **$	0.109
South	$-4.743 * **$	0.165
Center	$-1.202 * **$	0.201
Early-enrolled student	$-0.735 * **$	0.183
Late-enrolled student	$-3.390 * **$	0.154
ESCS	$1.986 * **$	0.025
Not living with both parents	$-0.967 * **$	0.070
Student with siblings	$-0.605 * **$	0.062
written reading score	0.002	0.002
oral reading score	$0.032 * **$	0.002
CRS5	$0.572 * **$	0.001
Random Effect		
σ_{School}	3.114	
σ_{Class}	5.343	
σ_{ϵ}	10.494	
VPC_{Class}	19.2%	
Size		
Number of observations	221,529	
Number of groups (School)	16,246	
Number of groups (Class)	3,920	

Table 5: ML estimates (with standard error) for model (8), fitted to the dataset. Asteriscs 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.

The coefficients of the variables at pupils level are very similar to the ones estimated in model (1). What is interesting is the proportion of explained variability: in model (1), 12.7 % of the total variability was explained at school level, here, 19.2 % of the variability is explained at class level and 6.5 % at school level. This high proportion of total variability present between classes may be due to the direct influence of the peers and the presence of good/bad

teachers. Lastly, the variance of the error σ_ϵ is a bit smaller respect to model (1), 10.494 respect to 11.448. Anyway, we must take into account that this variability between classes is nested within schools, so that, it's different from the variability between classes that we would obtain in a two-level linear mixed model with pupils nested only within classes.

3.2.1 Variables at Class Level

In Section 2.2.1, we analyzed how the estimated random effects $\hat{b}_j$ depend on the school level variables. Now it may be interesting to investigate which are the variables at class level that more influence the random effect $\hat{u}_{lj}$, where the class l, $l = 1, ...L_j$ and $L = \sum_k L_j$ is in school j=1,..J. As the variables at school level, these variables 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. The model is:

$$\hat{u}_{lj} = \alpha_0 + \sum_{k=1}^{K} \alpha_k w_{ljk} + \eta_{lj} \tag{10}$$

$$\eta_{lj} \sim N(0, \sigma_\eta^2) \tag{11}$$

where

$\hat{u}_{lj}$ is the estimated random effect of the l-th class in the j-th school from the model (8);

$\alpha = (\alpha_0, ..., \alpha_K)$ is the (k+1)-dimensional vector of parameters;

w_{ljk} is the value of the of the k-th predictor variable at class level;

η_{nj} is the zero mean gaussian error.

The histogram of the random effects at class level is reported in Figure 4.

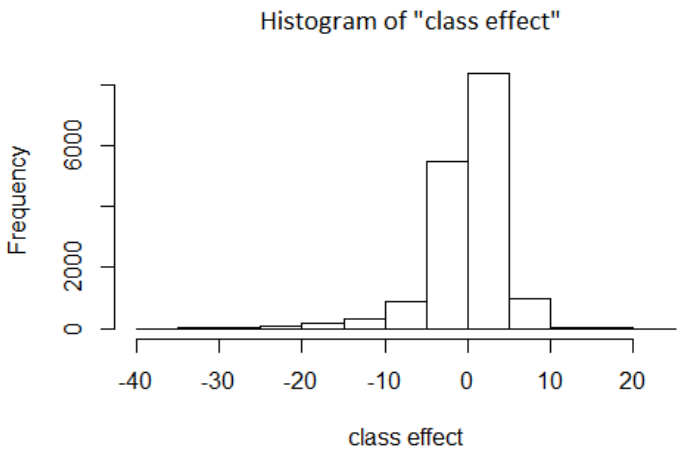

Figure 4: Histogram of random effects at class level.

Also in this case, with a lasso regression method, we select the main variables to use in the model.

Lasso Model coefficients	Estimates
Intercept	$-1.625 * **$
Mean ESCS	$-0.477 * **$
1^{st} generation immigrant percentage	$0.037 * **$
Late-enrolled students percentage	$0.018 * *$
Number of students	$0.061 * **$

Table 6: ML estimates for model (10), fitted on the reduced space of variables selected by Lasso. Asteriscs 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.

Again, the R^2s is very low (~ 0.03) but the mean ESCS and the number of students in the class seem to be the most important variables.

4 Reading Attainments across Macro-areas

In the analysis of the mathematics achievements emerged that the differences between the geographical three macro-areas (North, Center and South) were so deep that we could study three different models in which different variables were important and "school effect" was actually very heterogeneous, showing differences in the context of Italian educational system. We try now to understand what kind of differences elapses between macro-areas in the reading achievements. In both the multilevel linear mixed models seen before (models (1) and (8)), the variables related to the areas are relevant and the coefficients show a negative influence of attending a school in the Center and especially in the South respect to the North. The number of schools and students in the dataset are not equally distributed in the three area: there are 115,368 students in 1,800 schools in the North, 39,847 students in 688 schools in the Center and 66,314 students in 1,432 schools in the South (we lost lots of data in the South deleting the missing values in CS5).

An idea of the CRS distribution in the three macro-area is obtained from the boxplots in Figure 5.

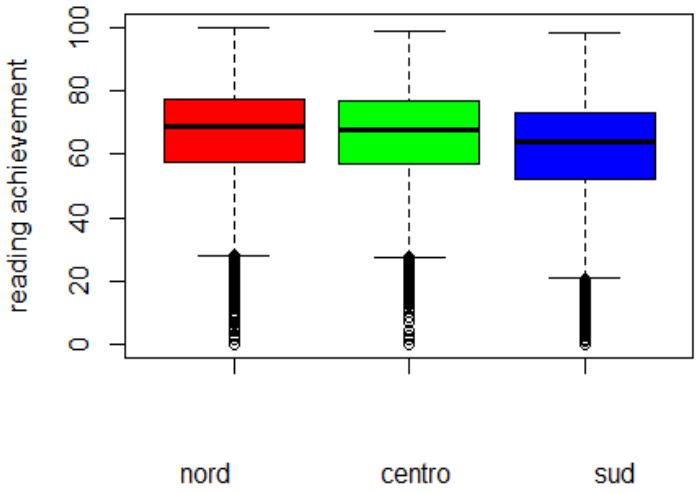

Figure 5: Boxplots of CRS stratified by geographical macro-areas

Looking at the boxplots, emerge that there is no a significant difference between the CRS in the North and Center, but the CRS of the South is visibly lower respect to the other ones and has a greater variance. Particularly, the mean of these two populations are different (p-value of the Wilcoxon test less than $2.2e-16$), 66.54 in the North and 62.07 in the South, and the variance of the CRS in the South is higher than the one in the North. This last difference is tested by a Levene's test (see [16]) based on Kruskal-Wallis (see [25]), from

the R package *lawstat* (see [31]), that is an inferential statistic used to assess the equality of variances for a variable calculated for two or more groups (p-value of the Levene's test for comparing variances less than $2.2e-16$). We use non-parametric tests because data are not normally distributed.

4.1 Two-level Linear Mixed Model in the three Macro-areas

To analyze if there are different correlations between CRS and the covariates and different school effects, we fit model (1) for each of the three geographical macro-areas:

$$y_{ij}^{(R)} = \beta_0^{(R)} + \sum_{k=1}^{K} \beta_k{}^{(R)} x_{kij}{}^{(R)} + b_j{}^{(R)} + \epsilon_{ij}{}^{(R)} \tag{12}$$

where R = {Northern, Central, Southern}
Table 7 shows the estimates of the three models.

Fixed Effects	North	Center	South
Intercept	$18.70 * **$	$25.10 * **$	$27.44 * **$
Female	$2.12 * **$	$1.86 * **$	$2.16 * **$
1^{st} generation immig	$-3.45 * **$	$-3.34 * **$	$-1.47*$
2^{nd} generation immig	$-3.37 * **$	$-2.94 * **$	$-1.02*$
Early-enrolled student	$-1.84 * **$	-0.72	-0.35
Late-enrolled student	$-3.17 * **$	$-2.58 * **$	$-4.64 * **$
ESCS	$1.55 * **$	$2.00 * **$	$2.58 * **$
not living with both parents	$-0.92 * **$	$-1.33 * **$	$-0.94 * **$
Student with siblings	$-0.50 * **$	$-0.56 * **$	$-0.66 * **$
written reading score	0.00	0.00	$-0.01.$
oral reading score	$0.01 * **$	$0.02 * **$	$0.07 * **$
CRS5	$0.63 * **$	$0.52 * **$	$0.44 * **$
Random Effects			
σ_{School}	3.81	4.18	4.83
σ_{ϵ}	10.59	11.59	12.57
VPC	11.5%	11.5%	12.8%
Size			
Number of observations	115,368	39,847	66,314
Number of groups (School)	1,800	688	1,432

Table 7: ML estimates for model (12) fitted to data of Northern, Central and Southern area. Asteriscs 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.

First of all, we note that the VPCs in the three areas are almost the same, that is the variability between schools explains almost the same part of the total variability in the three areas. This is different from the model fitted for mathematics data, in which the variability between schools was much stronger in the South than in the North (about 20% in the South respect to 10% in the North) (see [2]). The ESCS positively influences the attainments in all three cases, but it weighs more in the South (coefficient 2.58) than in the Center and in the North (2.0 and 1.55 respectively), suggesting that the family situation and the socio-cultural background is more relevant in the South. Also, being immigrants in the North weighs more than in the South (~ -3.3 vs -1.2) and this is probably due to the fact that in the South there are less immigrants than in the North. The coefficients of the CRS5 decreases from North to South, suggesting than in the North there is more continuity in school performances that in the South. All the other coefficients of fixed effects are very similar in

the three macro-areas. Lastly, the highest variance of the error is in the South, where the data are more dispersed.

4.1.1 Variables at School Level

Fitting three lasso regression models, we can individuate which are the variables that weigh more at school level in the three geographical macro-areas, that is which are the main characteristics of the schools that exert a positive/negative effect on students' achievement. The linear model is:

$$\hat{b}_j^{(R)} = \gamma_0^{(R)} + \sum_{l=1}^{L} \gamma_l{}^{(R)} z_{lj}{}^{(R)} + \eta_j{}^{(R)} \tag{13}$$

where $\hat{b}_j^{(R)}$ are the school random effects of area R, estimated by models (12) and the variables z_{lj} are selected by the Lasso regression. The boxplots of the $\hat{b}_j$ are reported in Figure 6.

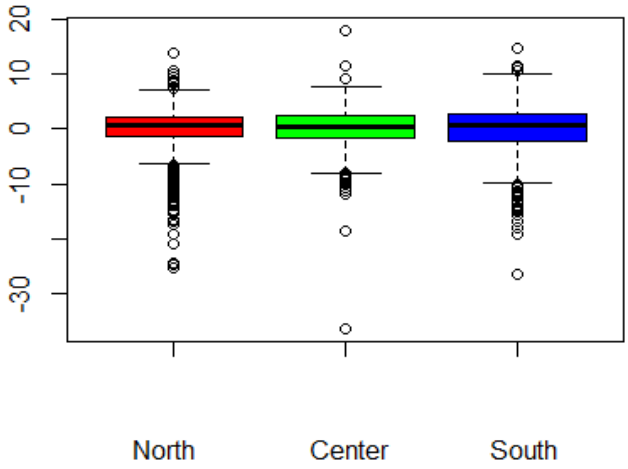

Figure 6: Boxplots of the $\hat{b}_j$ estimated in the three macro-areas.

The variability of $\hat{b}_j$ in the South is higher than in the rest of Italy (p-value $5.476e-16$ of the Levene's test of the three populations).

Lasso Model coefficients	North	Center	South
Intercept	0.156	−0.380	−0.733
Mean ESCS	−1.726 ∗ ∗	−0.396	1.035 ∗ ∗
Female percentage			0.031.
1^{st} generation imm perc			0.050
2^{nd} generation imm perc		0.153 ∗ ∗	
Early-enrolled student perc		−0.203∗	−0.091 ∗ ∗
Late-enrolled student perc	0.025		−0.073∗
Number of classes			
Number of students	0.002.		
Average num of stud per class			
Private school	−1.423 ∗ ∗	−2.455 ∗ ∗	

Table 8: ML estimates of model (13) fitted to data of Northern, Central and Southern area schools. Asteriscs 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.

The composition of student body (female, immigrants and early/late-enrolled students percentages) seems more relevant in the South than in the North, where, instead, weigh more managerial and structural features of the school (number of students in the school and private/public school). In particular, being a private school decreases the school mean value-added of -1.42 points in the North. Interesting is the mean school ESCS: it's relevant in all the three areas, but in the North the greater the medium school ESCS, the lower the value-added of the school. In the South is the opposite: schools with a high mean school ESCS give a high value-added.

4.2 Three-level Linear Mixed Model in the three Macro-areas

Lastly, for each area, we fit a three-level linear mixed model to analyze differences at class level. For each geographical area R:

$$y_{ilj}^{(R)} = \beta_0{}^{(R)} + \sum_{k=1}^{K} \beta_k{}^{(R)} x_{kilj}{}^{(R)} + b_j{}^{(R)} + u_{lj}{}^{(R)} + \epsilon_{ilj}{}^{(R)} \tag{14}$$

$$b_j^{(R)} \sim N(0, \sigma^{2(R)}_{School}), u_{lj}^{(R)} \sim N(0, \sigma^{2(R)}_{Class}), \epsilon_{ilj}^{(R)} \sim N(0, \sigma^{2(R)}_{\epsilon}) \tag{15}$$

Fixed Effects	North	Center	South
Intercept	$17.46 * **$	$23.28 * **$	$26.01 * **$
Female	$2.15 * **$	$1.86 * **$	$2.17 * **$
1^{st} generation immigr	$-3.48 * **$	$-3.27 * **$	$-1.59*$
2^{nd} generation immigr	$-3.38 * **$	$-2.98 * **$	$-1.18.$
Early-enrolled student	$-1.85 * **$	$-0.93.$	-0.31
Late-enrolled student	$-3.20 * **$	$-2.76 * **$	$-4.46 * **$
ESCS	$1.593707 * **$	$2.02 * **$	$2.51 * **$
not living with both parents	$-0.87 * **$	$-1.26 * **$	$-0.94 * **$
Student with siblings	$-0.54 * **$	$-0.58 * **$	$-0.65 * **$
written reading score	0.00	0.00	-0.00
oral reading score	$0.01 * **$	$0.02 * **$	$0.07 * **$
CRS5	$0.64 * **$	$0.55 * **$	$0.46 * **$
Random Effects			
σ_{School}	2.33	2.96	3.67
σ_{Class}	5.00	5.37	5.77
σ_{ϵ}	9.68	10.63	11.55
VPC_{Class}	20.1%	19.1%	18.4%
VPC_{School}	4.4%	5.8%	7.5%
Size			
Number of observations	115,368	39,847	66,314
Number of groups (Classes)	7,754	3,066	5,426
Number of groups (School)	1,800	688	1,432

Table 9: ML estimates of model (10) fitted to data of Northern, Central and Southern area schools. Asteriscs 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.

The estimates of the coefficients are similar to the ones obtained from the two-level mixed model (12), in Table 7. Here, part of total variability is explained by the variability between classes. While the percentage of variability explained at class level is higher in the North than in the South, the percentage explained at school level is higher in the South than in the North. So that, there is more difference between classes in the North than in the South but more difference between schools in the South than in the North. Anyway, as we did for model (8), we must take into account that this variability between classes is nested within schools, so that, it's different from the variability between classes that we would obtain in a two-level linear mixed model with pupils nested only

within classes.

4.2.1 Variables at Class Level

From model (14) we estimate the coefficients of random effects at class level, to fit linear models using variables at class level, in each of the three geographical macro-areas. The model is:

$$\hat{u}_{lj}^{(R)} = \alpha_0^{(R)} + \sum_{k=1}^{K} \alpha_k^{(R)} w_{ljk}^{(R)} + \eta_{lj}^{(R)} \tag{16}$$

$$\eta_{lj}^{(R)} \sim N(0, \sigma_\eta^{2(R)}) \tag{17}$$

The boxplots with the estimated random effects at class level are reported in Figure 7. Like all the random effects at different levels, the higher variance is in the South and the lower is in the North (p-value less than $2.2e-16$ of the Levene's test).

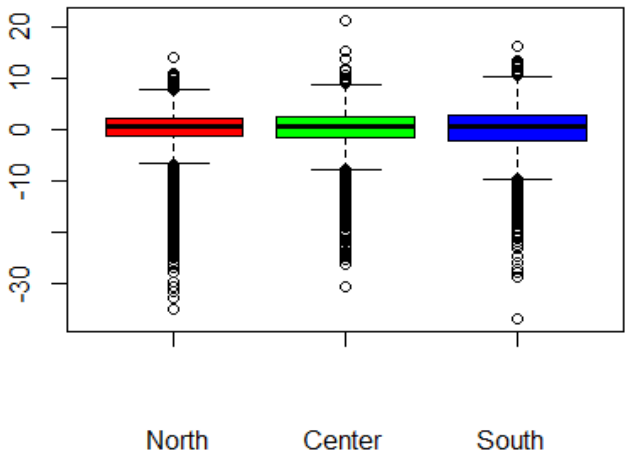

Figure 7: Boxplots of the estimated Random Effects at school level in the three macro-areas.

The coefficients selected by the Lasso regression model are reported in Table 10. The only variable relevant in all the three macro-areas is the medium ESCS of the class: in the North (coefficient -1.19) classes with a high mean ESCS give a negative contribution at the result, instead of in the South (coefficient 0.32), where classes with a high mean ESCS give a positive value-added. As

we have seen before, the percentage of immigrants is irrelevant in the South, where, however, class sizes are important, contrary to the North.

Lasso Model coefficients	North	Center	South
Intercept	$-1.70 * **$	$-0.54 * **$	$-1.03 * *$
Mean ESCS	$-1.19 * **$	$-0.59 * *$	$0.32 * *$
Female percentage			
1^{st} generation imm perc	$0.32 * *$	0.02	
2^{nd} generation imm perc		$0.03 * *$	
Early-enrolled student perc			$-0.02*$
Late-enrolled student perc	$0.03 * **$	$0.05 * **$	
Disable percentage			-0.00
Number of students	$0.06 * **$		$0.053 * **$
Tempo Pieno			

Table 10: ML estimates of model (16) fitted to data of Northern, Central and Southern area. Asteriscs 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.

5 Bivariate Multilevel Linear Mixed Models

As we can expect, there is a positive correlation between the performances of students in the two topics, CMS and CRS. By a test of correlation, we obtain a coefficient of correlation of 0.59 with a high significance (pval $< 2.2e-16$). Also, Figure 8 shows that correlation.

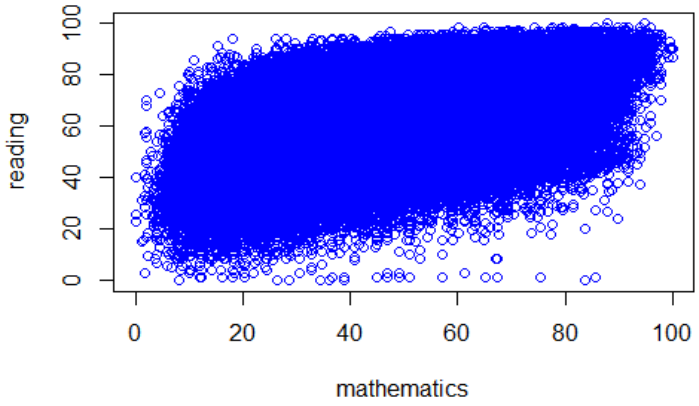

Figure 8: Reading vs mathematics achievement.

Therefore, it can be interesting to fit bivariate multilevel linear mixed models where the outcome variable is a matrix with both the achievements in mathematics and reading. This can be useful to explore the correlation between the two random effects and to extract information from the interaction between the two fields. All the bivariate multilevel linear mixed models are implemented using the software AsReml (see [11]).

5.1 Bivariate Two-level Linear Mixed Model

Let's take the model where pupil i, i = 1,..., n_{lj}; $n = \sum_{l,j} n_{lj}$ (first level) is in class l, l = 1,..., L_j; $L = \sum_k L_j$(second level) that is in school j, j = 1,..., J:

$$\overrightarrow{y}_{ij} = \overrightarrow{\beta}_0 + \sum_{k=1}^{K} \overrightarrow{\beta}_k x_{kij} + \overrightarrow{b}_j + \overrightarrow{\epsilon}_{ij} \tag{18}$$

where

$\overrightarrow{y} = \begin{pmatrix} y_{mat} \\ y_{read} \end{pmatrix}$ is the bivariate outocome with mathamatics and reading achievements;

$\overrightarrow{\beta} = \begin{pmatrix} \beta_{mat} \\ \beta_{read} \end{pmatrix}$ is the bivariate coefficient of the student's variables;

$\overrightarrow{b} = \begin{pmatrix} b_{mat} \\ b_{read} \end{pmatrix} \sim N(\overrightarrow{0}, \Sigma)$ is the matrix of the two random effects (mathematics and reading) at school level;

$\overrightarrow{\epsilon} = \begin{pmatrix} \epsilon_{mat} \\ \epsilon_{read} \end{pmatrix} \sim N(\overrightarrow{0}, W)$ is the error.

Figure 9 shows the histograms of the outcome variables, CRS and CMS.

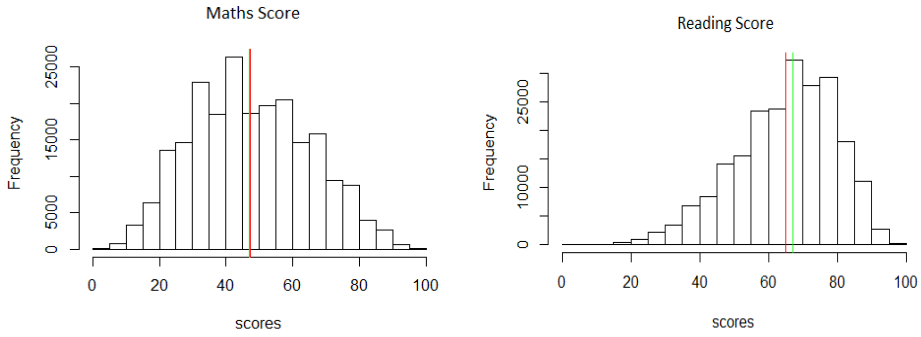

Figure 9: Histogram of Corrected Reading and Mathematics Score of pupils in the Invalsi database. The red lines refer to the mean, the green ones to the median.

Using the software ASReml, we can fit this bivariate model and obtain the new estimates of the coefficients. Table 11 shows the results. Note that we managed to use the CRS5/CMS5 as a regressor only for the reading/mathamatics achievement respectively, because the reading achievement doesn't depend on the mathematics score at grade 5 and the maths 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	−0.6447
CS5	0.505	0.4763
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)	3,900	

Table 11: ML estimates of model (18) fitted to the entire dataset.

We can now compare the estimates of the coefficients of the two topics. As we anticipated before, the coefficients of the variable "female" are almost opposites: being a female has a good effect in reading and a bad one in mathematics. Being immigrants has a negative effect in both the fields, but especially in reading, suggesting that the main difficulty for immigrants students is the language. Being a student in the South of Italy has a worst effect in mathematics than in reading, while anyway has a negative effect in both the topics. The ESCS and the score at grade 5 are positively correlated with the achievements and have similar coefficients in both the fields.

Looking at the variance/covariance matrix of the random effects, is clear that the variability of the mathematics random effects is much higher than the reading one (23.04 vs 13.08). The two effects are correlated with coefficient 0.307. Figure 10 shows that different variability.

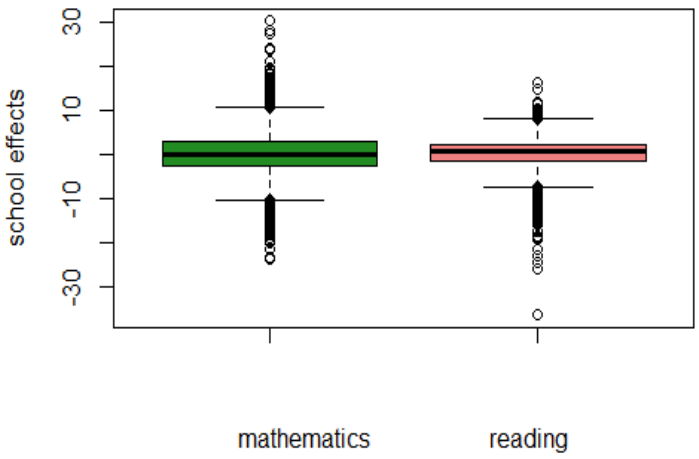

Figure 10: Estimated school effects $\hat{b}_j$ in mathematics and reading

To test this difference in variability, being both the populations not normal distributed (p-values of the Shapiro test less than $2.2e-16$), we implement a non-parametric Levene's test and we obtain a very low p-value (less than $2.2e-16$), proving that the variances of the random effects of the two topics are different. If we compare the random effects of the school estimates by the bivariate model with the ones estimated by the two univariate models (mathematics and reading), we can observe that they are almost the same, with correlation's coefficients of about 0.98. Anyway, we notice that the variability of the bivariate radom effects is smaller that the univariate's one, both in reading and mathematics (Figure 11). Again, we prove that with Kruskal-Wallis tests, testing the difference of variances for both the topics and obatining p-values of 0.0019 for mathamatics and 0.0011 for reading.

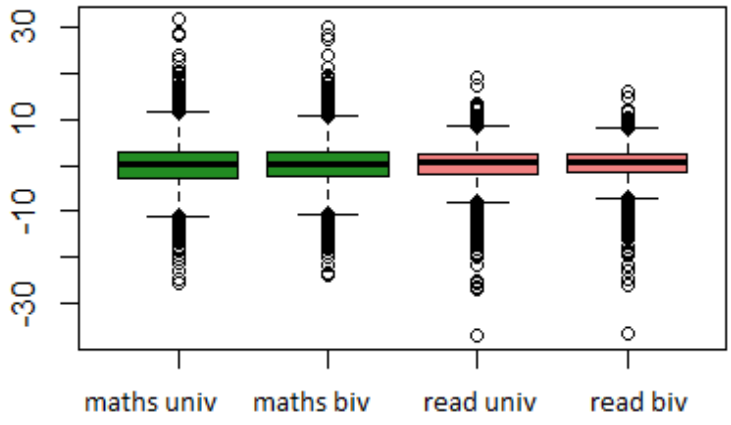

Figure 11: School effects $\hat{b}_j$ in mathematics and reading estimated by both the univariates and bivariate models.

5.2 Bivariate Two-level Linear Mixed Models among Macro-areas

Let's fit now the model (18) for each of the three macro-areas:

$$\overrightarrow{y}_{ij}^{(R)} = \overrightarrow{\beta}_0^{(R)} + \sum_{k=1}^{K} \overrightarrow{\beta}_k^{(R)} x_{kij} + \overrightarrow{b}_j^{(R)} + \overrightarrow{\epsilon}_{ij}^{(R)} \tag{19}$$

The estimates of the model are reported in Table 12.

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 fixed 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 12: ML estimates of model (19) fitted for each of the three macro-areas.

Looking at the estimates of the three models, we observe that, in general, the coefficients of variables immigrants and late/eartly-enrolled students of the South are closer to the 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 bigger 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. 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 variance 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 those of reading (43.6 vs 15.7). The correlation's coefficients 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. The matrices of errors don't seem to be significatively different.

Figure 12 reports the plots of the random effects of reading and mathematics estimated by the two univariate models first and then by the bivariate model. Also here, it's clear that the variability of the points in the univariate case is higher than the bivariate one and as we could expect from the variance/covariance matrices estimaed in table 12, the variance of the South is much higher than the North and the Center ones.

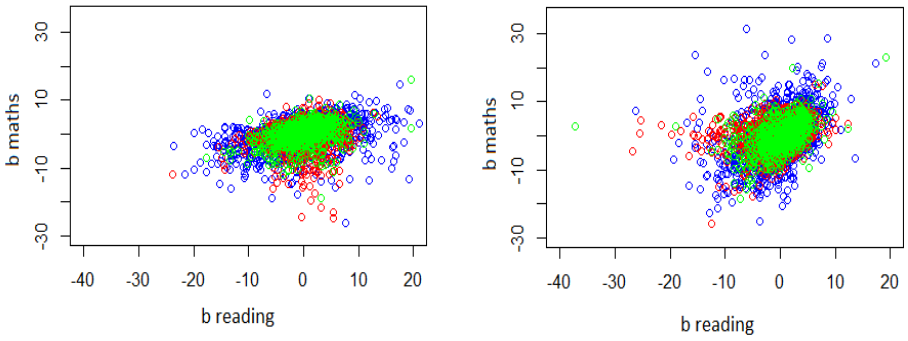

Figure 12: The first image represents the random effects estimated by the bivariate model by the two univariate models, while the second one represents the random effects estimated by the two univariate models. Colours identify the three macro-areas: blue for the South, red for the North and green for the Center.

5.2.1 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 some distance-based test for homogeneity of multivariate dispersions.

Traditional likelihood-based tests for homogeneity of variance–covariance matrices are extremely sensitive to departures from the assumption of multivariate normality, but there are tests that were found to be quite robust to departures from normality (see [18],[14]). In a univariate context, an example

is the Levene's test for homogeneity, that is essentially an analysis of variance (ANOVA) done on deviations from group means. In a multivariate context, Van Valen (see [29]) proposed a multivariate analogue to Levene's test as an ANOVA on the Euclidean distances of individual observations to their group centroid, defined as the point that minimizes the sum of squared distances to points within the group. O' Brien and Manly suggested that this approach could be made more robust by replacing centroids with multivariate median as the median for each variable within each group.

Following the ideas of Van Valen, O' Brien and Manly, who used Euclidean distances, a dissimilarity-based multivariate generalization of Levene's test is proposed. The suggested test statistics are ANOVA F-statistic comparing distances to centroids or spatial medians.

Let x_{ij} be the vector that denotes the point for the j-th observation in the i-th group in the multivariate space of p variables. Furthermore, let $\Delta(.,.)$ denote the Euclidean distance between two points. The centroid vector c_i for group i is defined as the point that minimizes the sum of squared distances to points within that group. One multivariate analogue to levene's test is to perform ANOVA on the Euclidean distances from individual points within a group to their group centroid, $z_{ij}^c = \Delta(x_{ij}, c_i)$.

A P-value for the F-statistic calculated on distances to centroids may be obtained using the traditional F-distribution.

A more robust version of Levene's test, suggested by Brown and Forsythe is to analyze deviations from medians instead. One multivariate analogue of this would be to calculate ANOVA on distances from the spatial median, $z_{ij}^m = \Delta(x_{ij}, m_i)$.

The approach may be extended to any distance or dissimilarity measure of choice through the use of principal coordinates. Let D = [d] be a square symmetric matrix of distances calculated between every pair of observations, l = 1,...,N and l' = 1,...,N. In the case of the Euclidean distance measure,

$$d_{ll'} = \Delta(x_l, x_{l'}) = \sqrt{\sum_{k=1}^{p}(x_{lk} - x_{l'k})^2} \tag{20}$$

To obtain principal coordinates, first let matrix A = $[a_{ll'}]$, where $a_{ll'} = -\frac{1}{2}d_{ll'}^2$. Centering this matrix in the manner of Gower (1966) gives $G = [g_{ll'}] = [a_{ll'} - \overline{a}_l - \overline{a}_{l'} + \overline{a}..]$, where $\overline{a}_l$ is the mean for row l, $\overline{a}_l$ is the mean for column l', and $\overline{a}..$ is the overall mean of the values in matrix A. Next, spectral decomposition of the G matrix yields

$$G = \sum_{l=1}^{N} \lambda_l q_l q_{l'}^T \tag{21}$$

,

where $\lambda_1, ...\lambda_N$ are the ordered eigenvalues of G and $q_1, ...q_N$ are the corresponding orthonormal eigenvectors. Principal coordinate axes (column vectors) are then obtained by scaling each axis q_l by the square root of its corresponding

eigenvalue, $u_l = \lambda^{\frac{1}{2}} q_l$. Now, unless the dissimilarities are indeed distances (Euclidean embeddable), matrix G may not be nonnegative definite and so some eigenvalues may be negative. The axes of matrix Q can be split into two sets, $Q = [q_1, ...q_r|q_{r+1}, ..., q_N]$, such that the first r eigenvectors correspond to the positive eigenvalues and the last (N-r) to the negative ones. For eigenvectors corresponding to nonnegative eigenvalues, l = 1,...,r, we denote scaled axes as $u_l^+ = (\lambda_l)^{\frac{1}{2}} q_l$. For eigenvectors l = r + 1,...,N corresponding to negative eigenvalues, we may scale by the square root of the absolute value of λ_l and subsequently multiply by $(-1)^{12}$, recognizing that these correspond to axes in imaginary space, i.e., $(-1)^{12} u_l^- = (|\lambda_l|)^{12} q_l$. Now, the original dissimilarity between two points x_{ij} and $x_{i'j'}$ can be recovered in the principal coordinate space using Euclidean distances, as

$$d_{ij,i'j'} = \sqrt{\Delta^2(u_{ij}^+, u_{i'j'}^+) - \Delta^2(u_{ij}^-, u_{i'j'}^-)} \tag{22}$$

. Furthermore, we can calculate a centroid for each of the i = 1, . . . , g groups in each of the real and imaginary spaces as c_i^+ i and c_i^- , respectively, in the usual way. Then, the distance (or dissimilarity) from the ij-th point to its centroid in the full principal coordinate space is

$$z_{ij}^c = \sqrt{\Delta^2(u_{ij}^+, c_i^+) - \Delta^2(u_{ij}^-, c_i^-)} \tag{23}$$

, where we will consider only positive square root. The test for homogeneity of dispersions then simply consists of doing univariate one-way ANOVA on the z's (see [3]).

Applying this last method and using the R package *vegan* (see [19]), 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 dispersed. 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 or random effects in the three marco-areas. Figure 13 shows the euclidean distances of the points from the centroids of each macro-area and what is created are the convex hulls in the Euclidean plane, that are the smallest convex set that contain the points. The biggest one is referred to the South.

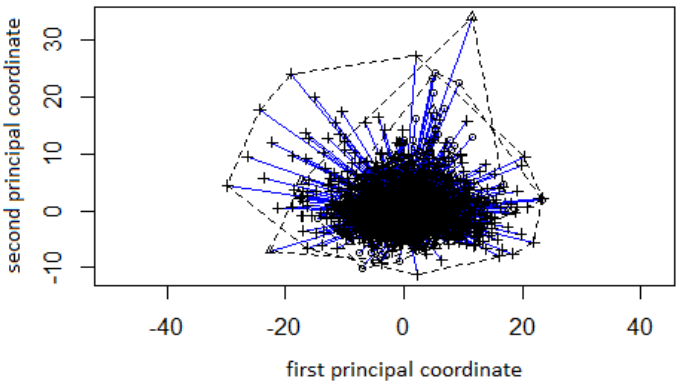

Figure 13: Euclidean distances between the points and the centroid of each macro-area in the principal component space.

If we try to repeat this study on the variance/covariance matrices of the errors, we can not use the entire set of data, with 221,529 residuals because the dimensions are too high for the software R. Therefore, we apply the method to different subsample of 3,000 data for each macro-area and we always notice the same trend of the random effects' matrices: the distributions and the variances of the residuals of the three macro-area 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 the variability that remains unexplained.

5.2.2 Variables at School Level among Macro-areas

Let's fit now three bivariate linear models in which the outcome variables are the school effects $\overrightarrow{\hat{b}_j}$ estimated by models (19) for each macro-area and the covariates are the variables at school level.

$$\overrightarrow{\hat{b}_j}^{(R)} = \overrightarrow{\gamma_0}^{(R)} + \sum_{k=1}^{K} \overrightarrow{\gamma_k}^{(R)} z_{jk}^{(R)} + \overrightarrow{\eta_j}^{(R)} \tag{24}$$

Estimates of model (20) are reported in Table 13.

Lasso Model coefficients	North mat	Center mat	South mat
Intercept	$-4.032**$	$-4.965**$	-2.679
Mean ESCS	0.262	0.815	$1.458***$
Female percentage	$0.034**$	$0.060**$	$0.072**$
1^{st} generation imm perc	$-0.040*$	0.053	$0.135*$
2^{nd} generation imm perc	-0.007	$0.152***$	0.037
Early-enrolled student perc	-0.097	$-0.229*$	$-0.116*$
Late-enrolled student perc	$-0.062**$	-0.063	$-0.272***$
Number of classes	$0.466*$	-0.031	-0.723
Number of students	$-0.017.$	0.007	$0.035.$
Average num of stud per class	$0.122*$	0.030	0.022
Private school	$-0.797.$	$-2.119*$	1.379
IC	0.270	0.608	0.518
R^2	0.11	0.07	0.06
Lasso Model coefficients	North read	Center read	South read
Intercept	-1.23	$-3.968*$	-1.257
Mean ESCS	$-1.810***$	-0.381	$0.752**$
Female percentage	0.005	0.021	$0.039*$
1^{st} generation imm perc	-0.002	0.027	$0.064.$
2^{nd} generation imm perc	0.017	0.129	-0.014
Early-enrolled student perc	0.025	$-0.225***$	$-0.101**$
Late-enrolled student perc	0.025	$0.041*$	$-0.080**$
Number of classes	0.105	0.125	-0.301
Number of students	-0.002	-0.003	0.015
Average num of stud per class	0.044	0.082	-0.009
Private school	$-1.212*$	$1.580.$	1.146
IC	0.047	0.345	0.316
R^2	0.02	0.07	0.02

Table 13: ML estimates of model (20) fitted to data of Northern, Central and Southern area. Asteriscs 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.

First of all, we note that the R^2s of the models of mathematics are a bit higher that those of the reading ones, suggesting that the regressors predict the outcome variable in a better way in mathamatics than in reading. Again, the medium ESCS of the school is very relevant in the South, in both the topics, with a positive influnce, while in the North has a negative weight. Generally, the composition of the school's peers, such that female, $1^{st}/2^{nd}$ generation immi-

grants, early/late enrolled students percentage, weights more in the South. At last, when it is relevant (North and Center), being a private school has always a negative influence.

6 Analysis of the School Effects

6.1 Exploratory Analysis

Having for each school the values-added given both for reading and mathamatics estimated by model (18), we may detect if there is a particular geographical distribution of positive/negative school effects or if there are some areas where the effects in reading and mathamatics are particularly "coherent", such that, schools with a positive/negative value-added in reading have also positive/negative effects in mathematics. Nevertheless, the data are really uniformly distributed between the three macro-areas, each area has the same percentage of very good/bad effects in both the topics, as we could expect from Figure 12, where, except few cases, the points are compact and don't emerge groups with different behaviors. Therefore, it's almost impossible to individuate clusters in this way.

Anyway, we can create a "total effect" of the school, mixing both the effects $\hat{b}_j$ of reading and mathamatics. The first variable created is the absolute value of the difference between the school effect of reading and mathematics:

$$diff = |\hat{b}_{mat} - \hat{b}_{read}|$$

that measures how much similar are the contributions that the school gives in both the fields and if the difference is small, it means that the school is "coherent" in its value-added in the two topics.

The second variable is the mean of the two school effects:

$$b_{mean} = \frac{\hat{b}_{mat} + \hat{b}_{read}}{2}$$

that measures the "global" value-added of the school. If b_{mean} is high/low, it means that the school weighs positively/negatively in the performances of both the fields.

Figure 14 shows the boxlots of the two variables among macro-areas.

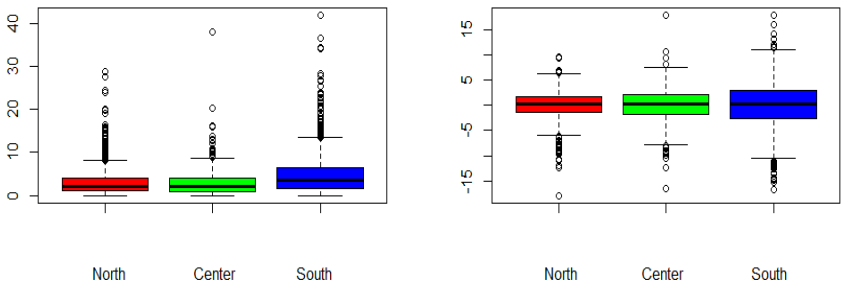

Figure 14: The first image reports the boxplots of the variable $diff$; the second image reports the boxplots of the variable b_{mean}.

From the first image, emerges that the South is the less "coherent" in the contribution in mathematics and reading: the mean of the difference is 4.70 against 2.99 of the North (p-value less than 2.2e-16 of the Wilcoxon-test for the difference of the two medians) and furthermore the South is also the most variable: the variance of the $diff$ in the North and in the Center are about 8.3 while the variance in the South is 20.40 (p-value less than 2.2e-16 of the Levene's test). In the second image, the means of the b_{mean} are the same, what is different among macro-areas are the variances: the variance at North is 6.90, at Center is 10.52 and at South is 19.27 (p-value of the Kruskal-Wallis based Levene's test less than 2.2e-16).

6.2 Depths of the School Effects

A different approach is to define the depth measures of multivariate data (see [28]). They are a generalization of the unidimensional quantiles in a multidimensional space. A data depth is a tool that allows to measure the "centrality" of a set of multivariate data respect to the reference data. More precisely, for a distribution P in $\mathbb{R}^d$, the correspondent data depth is a function $D(P; x)$ that allow to share the space $\mathbb{R}^d$ in different concentric regions. Taken a particular region, a point x may be internal or estarnal. The analysis of data depth is collocated in the non-parametric statistic and its strength is that it is easily formulated and directly applicable to the multivariate case. Introducing the data depth, we can define an order relation, indicated by symbol $\preceq$, in a set X of $\mathbb{R}^d$, that respect the reflexive, antisymmetric, transitive and total properties. This order relation has the *center* as maximum element. Therefore, the center is the maximum element of the order relation ($\preceq$), inducted by the data depth $D(P; x)$. Similarly, we define the *center* as the point that has maximum depth measure. By this definition, we deduce that the points "close" to the center have a high depth measure, while points "far" from the center have low depth measure. If a datum is surrounded by many other data, it will have a high level of depth, if it's in the "periphery" of the data it will have a low level of depth.

In a multivariate context, given a set of n points P, the *location depth* of a point u is the minimum number of points contained in any half-plane passing through u. To compute the set of all depth contours in the plane, observe the following relation between depth of a point and its dual line (Figure 15). Given a point $u \in \mathbb{R}^2$, let $l_u = D(u)$ be the dual line of u. Let $\mathrm{L} = \bigcup_i D(p_i)$ be the set of dual lines to all the points in P. Then any line passing through u, say a line l, in the primal plane corresponds to a point $D(l)$ in the dual plane such that $D(l)$ lies on the line l_u. Furthermore, due to the order preserving properties of the dual, the number of points of P lying above the line l in the primal plane corresponds to the number of lines in L vertically below the point $D(l)$ in the dual plane, i.e. the level of the point $D(l)$. Therefore, to compute the depth of point u in the primal plane, we can look at the dual line l_u, and find the point on l_u with the minimum level (see [15]).

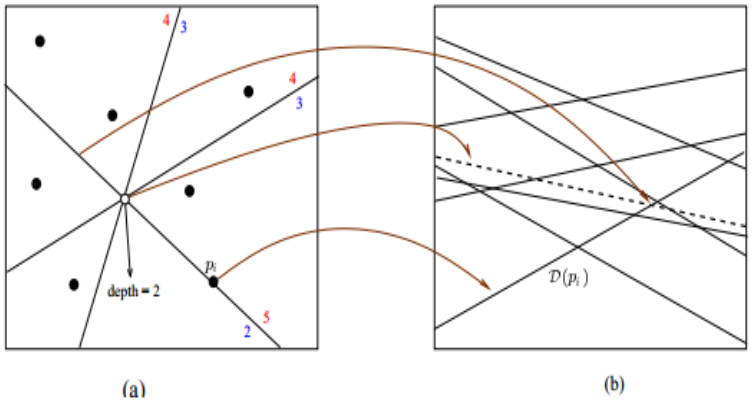

Figure 15: A set of points in the plane (a) and its set of dual lines (b)

Using the R package *depth* (see [10]), we calculate the depth measures for our bivariate data of school effects of Italy that go from 0 to 0.5. The depth of each point (school) is computed respect to the entire set of data . Figure 16 shows these depths, that are grouped in the three macro-areas.

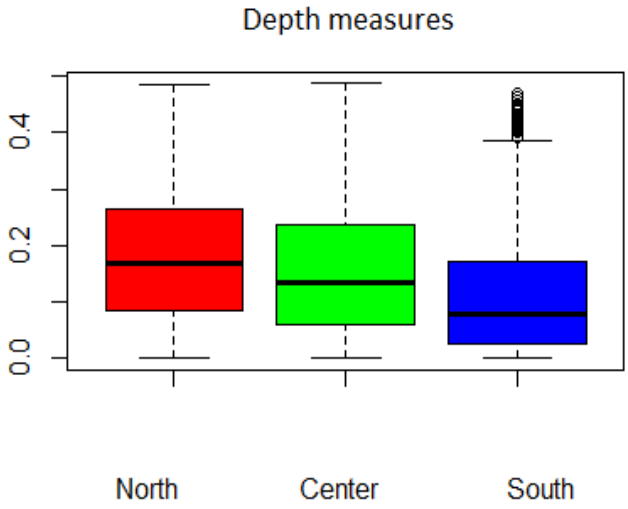

Figure 16: Depth levels of the school effects divided in the three macro-areas.

It's clear that the depths decrease from North to South, the means are differ-

ent (0.18 in the North, 0.16 in the Center and 0.11 in the South). Particularly, looking at the North and South, the Wilcoxon-test confirm the difference between the medians with a p-value less than $2.2e-16$ and also the variances result different (p-value $1.302e-07$ of the Levene's test). This result confirm the considerations of Figure 12, where we deduced that the Southern data are more diserse than the others. Looking at Figure 17, the depth levels of the Northern data are distributed almost uniformly between 0 and 0.4, while the Southern ones are mostly concentrated between 0 and 0.1. This distribution of the data suggests that, in the South, the school effects are more scattered than in the other areas.

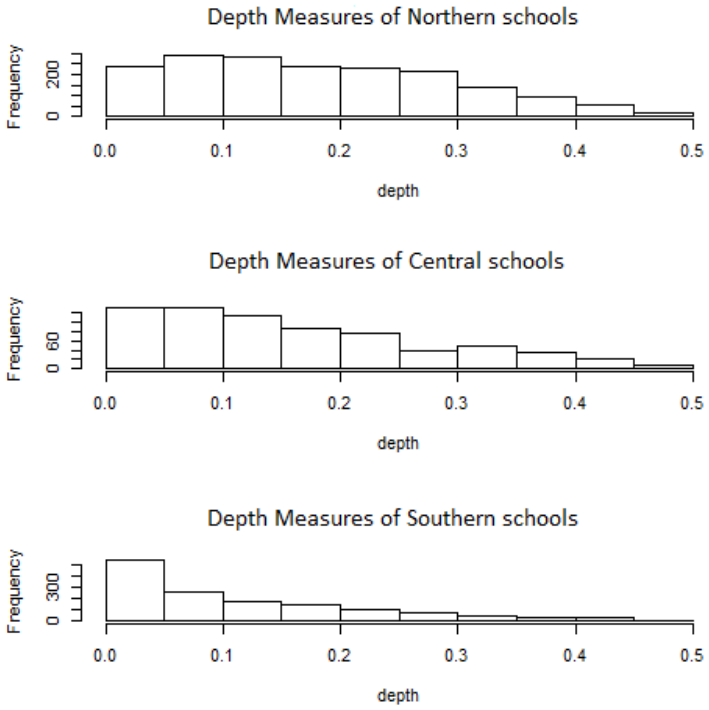

Figure 17: Histograms of depth levels in the three macro-areas.

7 Grouping for Classes

Until now, we focused the analysis on models in which pupils were nested in schools or, in some cases, in classes nested in their turn in schools. This was because our interest was, among the others, to identify differences between schools and schools' value-added. Now, we would like to observe such differences between classes. In this section, we implement two-level models in which pupils are nested only in classes.

7.1 Two-level Linear Mixed Model for Reading Achievement

First of all, we repeat model (1), changing the level of grouping: in model (1), pupils were nested in schools, here pupils are nested in classes. Therefore, pupil $i = 1...n_l$, $n = \sum_l n_l$ (first level) is nested in class $l = 1...L$ (second level). The model is:

$$y_{il} = \beta_0 + \sum_{k=1}^{K} \beta_k x_{kil} + u_l + \epsilon_{il} \tag{25}$$

$$u_l \sim N(0, {\sigma_u}^2), \epsilon_{il} \sim N(0, {\sigma_\epsilon}^2) \tag{26}$$

where

y_{il} is the reading test achievement of student i within class l;

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

$\beta = (\beta_0, ..., \beta_K)$ is the (K+1) dimensional vector parameters to be estimated;

u_l is the random effect of the l-th class and it's assumed to be Gaussian distributed and independent to any predictor variables that are included in the model;

ϵ_{il} is the zero mean Gaussian error.

The estimates of model (21) are reported in Table 14.

Fixed Effect	Estimate	Standard Error
Intercept	$23.332 * **$	0.174
Female	$2.113 * **$	0.045
1^{st} generation immigrant	$-3.498 * **$	0.128
2^{nd} generation immigrant	$-3.253 * **$	0.109
South	$-4.789 * **$	0.124
Center	$-1.249 * **$	0.147
Early-enrolled student	$-0.777 * **$	0.183
Late-enrolled student	$-3.413 * **$	0.154
ESCS	$1.986 * **$	0.025
Not living with both parents	$-0.974 * **$	0.070
Student with siblings	$-0.613 * **$	0.062
written reading score	0.002	0.002
oral reading score	$0.024 * **$	0.002
CRS5	$0.569 * **$	0.001
Random Effect		
σ_b	6.101	
σ_ϵ	10.497	
VPC	25.2%	
Size		
Number of observations	$221,529$	
Number of groups	$16,246$	

Table 14: ML estimates (with standard errors) for model (21), fitted to the dataset. Asteriscs 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.

The estimates are very similar to the ones estimated in both the models (1) and (8). What is interesting is that the variance of the error (10.497) is almost the same of the one of model (8) (10.494), where pupils where nested in classes, nested in schools and that the VPC in this model, that is the proportion of variation captured by the variation between classes (25.2%), is almost equal to the sum of the two VPCs in model (8), where 19.2% of the total variability was explained by the variance between classes and 6.5% by the variance between schools. Therefore, the proportion of variation captured in the two models is the same.

We fit now model (21) for each of the three macro-areas:

$$y_{il}^{(R)} = \beta_0^{(R)} + \sum_{k=1}^{K} \beta_k^{(R)} x_{kil} + u_l^{(R)} + \epsilon_{il}^{(R)} \tag{27}$$

Comparing this model with model (14), we see that the estimates of the two models are the same (we don't write them again) in the three macro-areas and the standard deviation of the errors too (9.68 in the North, 10.63 in the Center and 11.55 in the South). In Table 15, we see that, again, the sum of the two VPCs of each macro-area of model (14) is almost equal to the VPC of model (23). So that, the two models explain, in each macro-area, the same proportion of variability, suggesting that the proportion of variance that was captured by the difference between schools in model (14), now is totally captured by the differences between classes.

Model (14)	North	Center	South
σ_ϵ	9.68	10.63	11.55
VPC_{class}	20.1%	19.1%	18.4%
VPC_{school}	4.4%	5.8%	7.5%
Model (23)	**North**	**Center**	**South**
σ_ϵ	9.68	10.63	11.55
VPC_{class}	24%	24.7%	25.3%

Table 15: Comparison of standar deviation of errors and VPCs between models (14) and (23).

7.2 Bivariate Linear Mixed Model

Let's now fit a bivariate model with outcome variables CRS and CMS, grouping pupils for classes. Pupil i, i = 1,..., n_l, $n = \sum_l n_l$ (first level) is in class l, l = 1,...,L (second level):

$$\overrightarrow{y}_{il} = \overrightarrow{\beta}_0 + \sum_{k=1}^{K} \overrightarrow{\beta}_k x_{kil} + \overrightarrow{u}_l + \overrightarrow{\epsilon}_{il} \tag{28}$$

where

$\overrightarrow{y} = \begin{pmatrix} y_{mat} \\ y_{read} \end{pmatrix}$ is the bivariate outocome with mathamatics and reading achievements;

$\overrightarrow{\beta} = \begin{pmatrix} \beta_{mat} \\ \beta_{read} \end{pmatrix}$ are the bivariate coefficients of the student's variables;

$\overrightarrow{u} = \begin{pmatrix} u_{mat} \\ u_{read} \end{pmatrix} \sim N(\overrightarrow{0}, \Sigma)$ is the matrix of the two random effects (mathematics and reading) at class level;

$\overrightarrow{\epsilon} = \begin{pmatrix} \epsilon_{mat} \\ \epsilon_{read} \end{pmatrix} \sim N(\overrightarrow{0}, W)$ is the error.

The estimates of the coefficients are again very similar to the ones obtained in model (18) (see Table 11). We focus the attention on the variance/covariance matrix of random effects:

$$\begin{pmatrix} 34.9 & 4.46 \\ 4.46 & 26.5 \end{pmatrix}$$

The variances of random effects at class level are really big respect to the ones at school level, obtained in model (18): 34.9 vs 23.04 in mathematics and 26.5 vs 13.08 in reading, for the variances. This suggests that the variability of the value-added of the class is much higher than the value-added of the school, so that, attending a particular class can influence the mean result more than attending a particular school. Figure 18 shows this difference between the variabilities in maths and reading (both the p-values of the Levene's test are less than $2.2e-16$). Lastly, the correlation between the random effects of classes of the two topics is 0.147, less than the one of the random effects of school (0.3), suggesting that the contributes of the class in the two topics are not particularly correlated.

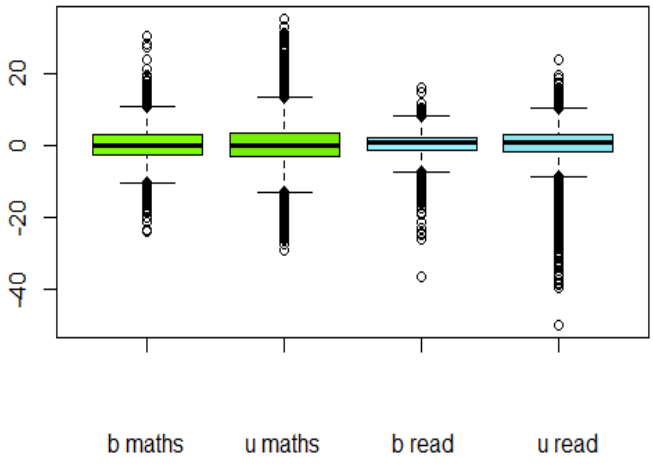

Figure 18: Boxplots of the random effects in maths and reading, at school and class levels.

Let's fit now model (23) for each of the three macro-areas:

$$\overrightarrow{y}_{il}^{(R)} = \overrightarrow{\beta}_0^{(R)} + \sum_{k=1}^{K} \overrightarrow{\beta}_k^{(R)} x_{kil} + \overrightarrow{u}_l^{(R)} + \overrightarrow{\epsilon}_{il}^{(R)} \tag{29}$$

Again, the estimates of the coefficients are very similar to the ones obtained in models (19) (see Table 12). The variance/covariance matrices of random effects are reported in Table 16.

North	Center	South
$\begin{pmatrix} 13.9 & 1.05 \\ 1.05 & 22.5 \end{pmatrix}$	$\begin{pmatrix} 24.6 & 3.86 \\ 3.86 & 26.3 \end{pmatrix}$	$\begin{pmatrix} 73.9 & 8.67 \\ 8.67 & 30.7 \end{pmatrix}$

Table 16: Variance/Covariance matrices of random effects in the three macro-areas.

The three matrices are quite different: while in the North, the variance of the class effect in reading is much higher than the one in maths (22.5 vs 13.9), in the South it is the opposite (30.7 vs 73.9) and in the Center the variances are similar (26.3 vs 24.6). In the South, the effects in the two topics are more correlated than in the North (coefficients of correlation 0.182 vs 0.059). In order to test the differences between the three matrices, we apply again the method presented in 5.2.1: the means of the Euclidean distances between points and centroid within each group are 4.670 in the North, 5.695 in the Center and 8.438 in the South, showing that, as we obtained in other analysis, the points of the South are more dispersed. The test ANOVA gives a p-value less than $2.2e-16$.

Comparing these three matrices with the variance/covariance matrices of the school effects obtained in models (19), we see that the variances of the random effects at class level in both the topics are almost the double of the ones at school level in all the three macro-areas. Lastly, computing the coefficients of correlation, we notice that the class effects are less correlated than the school effects in the two topics, in all the three macro-areas (0.059, 0.152 and 0.182 vs 0.62, 0.56 and 0.36). Such a low correlation between the class effects in the two topics can be explained by the fact that the positive/negative value-added of the class may depends on the "good"/ "bad" teachers, that are different from reading to mathematics.

7.2.1 Variables at Class Level

We try now to understand how the variables at class level are correlated with the value-added $\overrightarrow{u_l}$ of the classes. Let's start fitting a model with, as outcome variable, the estimates of $\overrightarrow{u_l}$ of model (24), where we consider data of all Italy:

$$\overrightarrow{\hat{u}_l} = \alpha_0 + \sum_{k=1}^{K} \alpha_k w_{lk} + \eta_l \tag{30}$$

where
$\overrightarrow{\hat{u}_j} = \begin{pmatrix} \hat{u}_{mat} \\ \hat{u}_{read} \end{pmatrix} \sim N(\overrightarrow{0}, \Sigma)$ is the matrix of the two random effects (mathematics and reading) at class level estimated by model (24).

Using the Lasso regression method to select the variables, we fit a model with a reduced space of variables. The results of model (26) are reported in Table 17.

Mathematics	Lasso Model coefficients	Estimates
Intercept	$-8.364590 * **$	0.680758
Mean ESCS	$0.865413 * **$	0.095082
Female percentage		
1^{st} generation immigrant percent		
2^{nd} generation immigrant percent		
Early-enrolled student percent		
Late-enrolled students percent	$-0.052170 * **$	0.007233
Disable percent		
Number of students per class	$0.083316 * **$	0.012094
Compiled percent	$0.070950 * **$	0.006610
Tempo Pieno		
R^2	0.0243	
Reading	Lasso Model coefficients	Estimates
Intercept	$-4.577520 * **$	0.545140
Mean ESCS	$-0.939660 * **$	0.081052
Female percent		
1^{st} generation immigrant percent	$0.054025 * **$	0.006247
2^{nd} generation immigrant percent	$0.033567 * **$	0.007038
Early-enrolled student percent		
Late-enrolled students percent		
Disable percent		
Number of students per class		
Compiled percent	$0.045249 * **$	0.005794
Tempo Pieno		
R^2	0.01967	

Table 17: ML estimates (with standard errors) for model (26), fitted to the dataset. Asteriscs 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.

While in mathematics the mean ESCS has a positive correlation with the class effect, in reading this correlation is negative. Furthermore, in mathematics seem to be more relevant the variables describing the size of the class, such as number of students per class (positively correlated) and the percentage of compiled tests. In reading, instead, the composition of the student body of the class, such as 1^{st} and 2^{nd} generation immigrant percentages (positively correlated) weighs more.

We fit now model (26) for each of the three macro-areas, in order to assess

if the value-added of the class is influenced by different aspects:

$$\overrightarrow{\hat{u}_l}^{(R)} = \alpha_0^{(R)} + \sum_{k=1}^{K} \alpha_k^{(R)} w_{lk}^{(R)} + \eta_l^{(R)} \tag{31}$$

with R = {North, Center, South}

Table 18 shows the estimates of model (27) fitted to the dataset of the North:

Mathematics	Lasso Model coefficients	Estimates
Intercept	$-5.204984 * **$	0.800781
Mean ESCS	$0.226303*$	0.095990
Female percent		
1^{st} generation immigrant percent		
2^{nd} generation immigrant percent		
Early-enrolled student percent		
Late-enrolled students percent	$-0.026891 * **$	0.006571
Disable percent		
Number of students per class	$0.047197 * **$	0.012776
Compiled percent	$0.044688 * **$	0.007801
Tempo Pieno		
R^2	0.01034	
Reading	Lasso Model coefficients	Estimates
Intercept	$-2.095851 * **$	0.373828
Mean ESCS	$-1.813397 * **$	0.121579
Female percent		
1^{st} generation immigrant percent	$0.035354 * **$	0.008883
2^{nd} generation immigrant percent		
Early-enrolled student percent		
Late-enrolled students percent	$0.035903 * **$	0.009448
Disable percent		
Number of students per class	$0.088483 * **$	0.015980
Compiled percent		
Tempo Pieno		
R^2	0.04709	

Table 18: ML estimates (with standard errors) for model (27), fitted to the dataset of the North. Asteriscs 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.

Table 19 shows the estimates of model (27) fitted to the dataset of the Center:

Mathematics	Lasso Model coefficients	Estimates
Intercept	$-8.10860 * **$	1.20537
Mean ESCS	$0.53021 * *$	0.20537
Female percent		
1^{st} generation immigrant percent		
2^{nd} generation immigrant percent	$0.05293 * **$	0.01472
Early-enrolled student percent		
Late-enrolled students percent		
Disable percent		
Number of students per class	$0.12765 * **$	0.02351
Compiled percent	$0.05088 * **$	0.01160
Tempo Pieno		
R^2	0.02172	
Reading	Lasso Model coefficients	Estimates
Intercept	$-0.64498 * **$	0.18622
Mean ESCS	$-1.29164 * **$	0.21632
Female percent		
1^{st} generation immigrant percent	$0.03684*$	0.01617
2^{nd} generation immigrant percent	0.05851	0.01521
Early-enrolled student percent		
Late-enrolled students percent	$0.06141 * **$	0.01707
Disable percent		
Number of students per class		
Compiled percent		
Tempo Pieno		
R^2	0.04306	

Table 19: ML estimates (with standard errors) for model (27), fitted to the dataset of the Center. Asteriscs 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.

Table 20 shows the estimates of model (27) fitted to the dataset of the South:

Mathematics	Lasso Model coefficients	Estimates
Intercept	$-10.31382 * **$	1.46350
Mean ESCS	$2.21170 * *$	0.22395
Female percent		
1^{st} generation immigrant percent		
2^{nd} generation immigrant percent		
Early-enrolled student percent		
Late-enrolled students percent	$-0.13799 * **$	0.01962
Disable percent		
Number of students per class	$0.11951 * **$	0.02585
Compiled percent	$0.09116 * **$	0.01418
Tempo Pieno		
R^2	0.05807	
Reading	Lasso Model coefficients	Estimates
Intercept	$-7.455701 * **$	0.951469
Mean ESCS	0.242989.	0.146591
Female percent		
1^{st} generation immigrant percent		
2^{nd} generation immigrant percent		
Early-enrolled student percent	$-0.050022 * *$	0.015269
Late-enrolled students percent		
Disable percent		
Number of students per class	$0.078980 * **$	0.017077
Compiled percent	$0.063195 * **$	0.009319
Tempo Pieno		
R^2	0.01495	

Table 20: ML estimates (with standard errors) for model (27), fitted to the dataset of the South. Asteriscs 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.

The variables that seem to be significant in most of the models are the mean ESCS and the number of students per class: the number of students per class has always a positive coefficient, suggesting that classes with a high number of students give high value-added; the mean ESCS has positive coefficients in the South and negative or close to zero in the North, suggesting that, once again, the ESCS is more relevant in the South, where classes with a high mean ESCS give positive value-added, than in the North, where this regressor is not

so relevant. In all the models, the R^2s is very low, therefore, a big part of the models remains unexplained.

7.3 Analysis of the Class Effects

Let's make some exploratory analysis on the class effects $\overrightarrow{\hat{u}_l}$, estimated in model (24). In Figure 12, we saw that the school effects in the South were more scattered that the ones in the Center and in the North, that were similar. We repeat that kind of plot for the class effects in Figure 19.

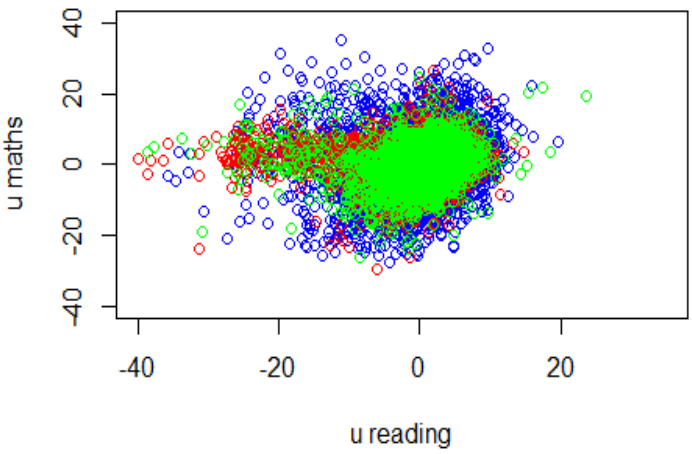

Figure 19: Random effects estimated by model (24) coloured by macro-areas: blue for the South, green for the Center and red for the North.

The trend is very similar to the one of figure 12: the points of the South are the most scattered, so that, in the South the values-added of the class are very heterogeneous and significantly influence the medium results of pupils. Again, the points of the North and the Center are similar and we can notice a queue on the left, where there are lots of classes that give a negative contribute in reading and a negligible one in maths.

We define now the depth measures for the class effects. As we did in Section 6.2, we compute the depths for each of the three macro-areas. Each depth is computed respect to the entire set of points.

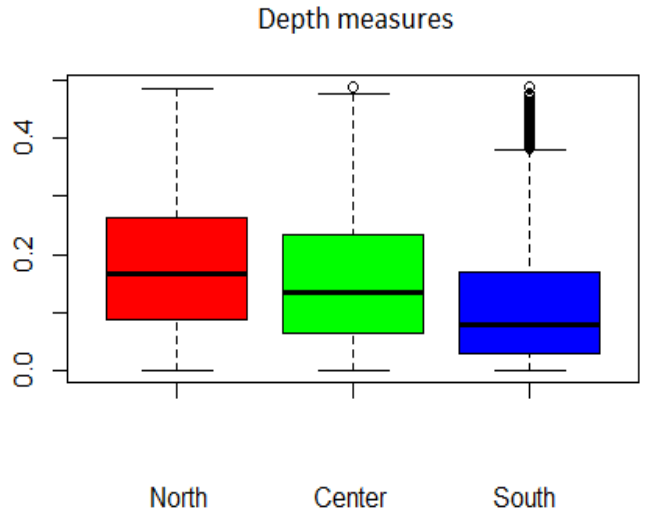

Figure 20: Depth levels of the class effects divided in the three macro-areas.

The trend is similar to the one in Figure 16. The depth levels of the South are generally lower than the other two macro-areas because, as we saw below, the points are more scattered. The mean are respectively 0.182 in the North, 0.156 in the Center and 0.114 in the South. The Kruskal-Wallis test confirm the difference between the three medians of the group, with a p-value less than 2.2e-16. Furthermore, the variance of the depth levels in the South is lower respect to the others (p-value of the Levene's test less than 2.2e-16). These aspects can be seen also from the histograms of the depth levels, shown in Figure 21.

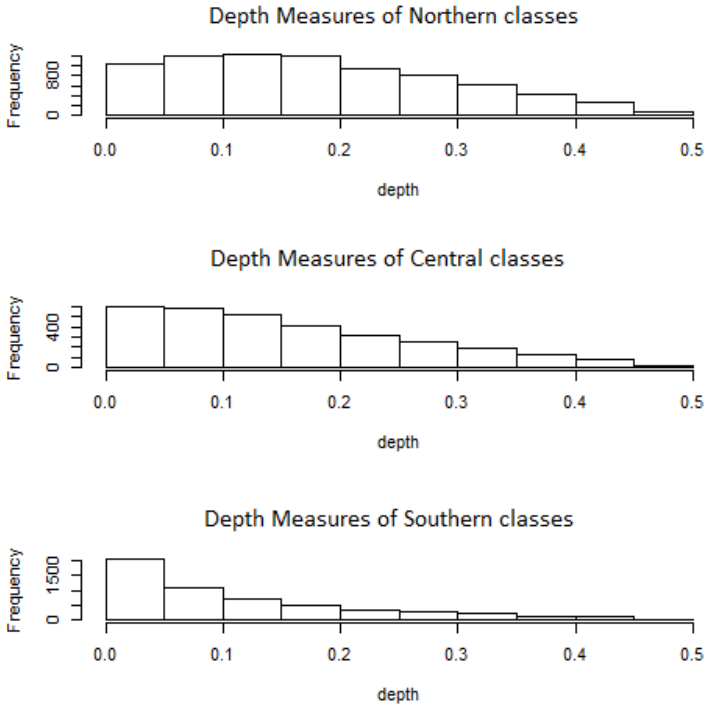

Figure 21: Histogram of depth levels in the three macro-areas.

It's immediatly clear that the variability of depth levels in the South is strictly reduced respect to the Northern one. In the South, most of the points has a depth measure between 0 and 0.1, so that, most of the points are on the "periphery" of the sample (significant value-added of the class, different from zero); the depth measures in the Center decrease uniformly and the depth measures of the North are the most uniformly distributed between 0 and 0.5.

8 Concluding Remarks

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

Initially, univariate multilevel models have been applied in order to explain how the reading achievements depend on students' characteristics. What has emerged are some recurrent dependencies between the outcome variable and regressors: females have, on average, better results than males, 1^{st} and 2^{nd} generation immigrants have more difficulties than Italian students and being early/late-enrolled student decreases the medium result, such as being a pupil not living with both parents. Furthermore, the pupil's ESCS, index of the socioeconomic status of the student, has a strong positive influence on the achievement and the CRS5 (the student's achievement at grade 5) is positively correlated with the current score, contrarily to the written and oral reading score at school, that do not seem correlated with the INVALSI test score. This kind of relations between pupils' characteristics and achievements emerged also in mathematics test (explored in previous literature), except the fact that, in mathematics, males are generally better than females. Lastly, from the model emerged that students of the Central and especially of the Southern Italy, have worst results than students of the North, showing big heterogeneities within the Country.

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 b_j and has been regressed against a school level variables with the aim of characterising the features of those schools that exert a positive/negative effect on academic performance. Particularly, the result shows that in Italy, Private schools give a negative value-added respect to the Public ones.

What is interesting is that school effects and 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 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. Even the school effects are different: 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 medium 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.

Actually, the main contribute of this work has been the bivariate approach, in which bivariate multilevel mixed models have been used in order to explore the reading and mathematics scores together. In this way, it has been possible to compare especially the school and class effects in both the topics. Regarding

the school effects, emerged that, in Italy, the effect in mathematics is much higher than the one in reading. In the macro-areas, this behaviour occurs also in the South of Italy, while in the North is less pronounced. The values-added of the South are always higher than the ones of the North. 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. This behaviour does not occur in the classes where, instead, the value-added in the two topics are less correlated, 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 can have a good teacher of maths and a bad one of reading and 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 higher than in the North in both the topics, being different across macro-areas.

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's level to everyone.

9 Code

In this section, are reported the main parts of the R and AsReml code, used to compute this study.

Univariate two-level linear mixed model in R

Function that fits the two-level linear mixed model, with the reading achievement as outcome variable, the variables at student level (fixed effects) as covariates and the random effect given by the school.

```
library(nlme)

regmixedita = lme(pu_it_no_corr  ~ FEMMINA  +S1 +S2 +Sud
                  +Centro +ANTICIPATARIO +POSTICIPATARIO
                  +ESCS  + No_genitori +Si_fratelli +voto_scritto_ita
                  + voto_orale_ita + pu_it_no_5, data = mcompleto,
                  random = ~ 1 | CODICE_SCUOLA, method ="ML")

summary(regmixedita)

# to obtain the coefficients of random effects
random.effects(regmixedita)
```

Univariate three-level linear mixed model in R

Function that fits the three-level linear mixed model, with the reading achievement as outcome variable, the variables at student's level (fixed effects) as covariates and the random effects given by the school and the class.

```
library(nlme)

regmixedita3 = lme(pu_it_no_corr  ~ FEMMINA  +S1 +S2 +Sud
                   +Centro +ANTICIPATARIO +POSTICIPATARIO
                   +ESCS  + No_genitori +Si_fratelli +voto_scritto_ita
                   + voto_orale_ita + pu_it_no_5, data = mcompleto,
                   random = ~ 1 | CODICE_SCUOLA/CODICE_CLASSE,
                   method = "ML")

summary(regmixedita3)
```

```
# to obtain the coefficients of the two random effects
random.effects(regmixedita3)[[1]]
random.effects(regmixedita3)[[2]]
```

Bivariate model in AsReml

Function that fits bivariate two-level model, with both reading and maths achievements as outcome variables, variables at student's level as covariates and random effect given by the school.

```
modelli lineari bivariati a effetti misti

# definition of the variables

pu_ma_no_corr
pu_it_no_corr
FEMMINA 2
S1        2
S2       2
ESCS
ANTICIPATARIO 2
POSTICIPATARIO 2
No_genitori  2
Si_fratelli 2
area 3
pu_5 !G 2
CODICE_SCUOLA !A

# definition of the model

biv.asd !skip 1 !WORKSPACE 1906
pu_ma_no_corr pu_it_no_cor ~ Trait Trait.FEMMINA Trait.S1
        Trait.S2 Trait.ANTICIPATARIO Trait.POSTICIPATARIO
        Trait.ESCS Trait.No_genitori, Trait.Si_fratelli Trait.area
        Trait.vect(pu_5) !r Trait.CODICE_SCUOLA

# definition of variance/covariance structures
# of error and random effects matrices

1 2 1
0 0 ID
Trait 0 US
3*0
Trait.CODICE_SCUOLA 2
Trait 0 USH !GP
CODICE_SCUOLA 0 ID
```

Depth Measures

Function that associates each point to its depth.

```
library(depth)

prof = rep(0,3920)
for (i in 1:3920){
  prof[i]= depth(as.vector(b_biva[i,]),b_biva)
}
```

Differences between Matrices

Code to compute the scattering of three bivariate populations and test the differences between them.

```
library(vegan)

jap = rbind( b_biv_nord, b_biv_centro, b_biv_sud)
dim(jap)
dis <- vegdist(jap,method="euclidean")
groups <- factor(c(rep(1,1800), rep(2,688),rep(3,1432)),
          labels = c("nord","centro","sud"))

mod <- betadisper(dis, groups,type="centroid")
summary(mod)
anova(mod)

mod2 <- betadisper(dis, groups,type="median")
summary(mod2)
anova(mod2)
```

Acknowledgments

This work is within FARB - Public Management Research: Health and Education Systems Assessment, funded by Politecnico di Milano. The author is grateful to Invalsi for having provided the original dataset, and prof. Tommaso Agasisti for the support provided during this thesis.

References

[1] T. Agasisti, G. Catalano, and G. Vittadini. *Rapporto sulla scuola in Lombardia: Strumenti di analisi e di policy.* Guerini e Associati, Febbraio 2013.

[2] 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*, January 31, 2014.

[3] M. J. Anderson. Distance-based tests for homogeneity of multivariate dispersions. *Department of Statistics, University of Auckland, Private Bag 92019, Auckland, New Zealand*, March 2006.

[4] F. Angeli. *Le competenze degli studenti quindicenni lombardi. I risutati di Pisa 2006.* Istituto Regionale Ricerca Educativa, 2009.

[5] A. N. Baraldi and C. K. Enders. An introduction to modern missing data analyses. November 2009.

[6] D. Checchi. *Stato e Mecrcato*, chapter Da dove vengono le competenze scolastiche? Il Mulino, Dicembre 2004.

[7] J. Dronkers and P. Robert. Differences in scholastic achievement of public, private government-dependent, and private independent schools. *Educational Policy*, 2008.

[8] J. Fox and S. Weisberg. An appendix to an r companion to applied regression, second edition multivariate linear models in r. July 2011.

[9] John Fox. Linear mixed models. *Appendix to An R and S-PLUS Companion to Applied Regression*, May 2002.

[10] Maxime Genest, Jean-Claude Masse, and Jean-Francois Plante. *depth: Depth functions tools for multivariate analysis*, 2012. R package version 2.0-0.

[11] 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.*

[12] 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.

[13] Trevor Hastie and Brad Efron. *lars: Least Angle Regression, Lasso and Forward Stagewise*, 2013. R package version 1.2.

[14] W. J. Krazanowski. Permutational tests for correlation matrices. *Department of Mathematical Statistics and Operational Reasearch, University of Exeter, UK*, September 1992.

[15] S. Krishnan, N. H. Mustafa, and S. Venkatasubramanian. Statistical data depth and the graphics hardware. *DIMACS Series in Discrete Mathematics and Theoretical Computer Science*, 1991.

[16] Levene. *Contributions to Probability and Statistics: Essays in Honor of Harold Hotelling*. 1960.

[17] P. Montanaro. *I divari territoriali nella preparazione degli studenti italiani: evidenze dalle indagini nazionali e internazionali*. June 2008.

[18] P. O' Brien. Robust procedures for testing equality of covariances matrices. *International Biometric Society*, October 2014.

[19] 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.

[20] J.J. C. Pinheiro and D. M. Bates. *Mixed-effects Models in S and S-PLUS*. Springer, 2000.

[21] 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.

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

[23] D. B. Rubin. *Multiple Imputation for Nonresponse in Surveys*. 1987.

[24] D. B. Rubin. *Inference and Missing Data*. Biometrika, 2003.

[25] J.D. Spurrier. On the null distribution of the kruskal-wallis statistic. *Journal of Nonparametric Statistics*, 2003.

[26] R. Tibshirani. Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society - Series B*, 1996.

[27] P. E. Todd and K. I. Wolpin. The production of cognitive achievement in children: Home, school, and racial test score gaps. *University of Pennsylvania*.

[28] J. W. Tukey. Mathematics and the picturing of data. *Proceedings of the International Congress of Mathematicians*, Vancouver, 1974.

[29] L. Van Valen. The statistics of variation. *Evolutionary Theory 4*, 1978.

[30] A. Vignoles, R. Levacic, J. Walker, S. Machin, and D. Reynolds. *The Relationship Between Resource Allocation and Pupil Attainment: A Review*. ISBN, 2000.

[31] Joseph L. Gastwirth; Yulia R. Gel <ygl@math.uwaterloo.ca>; W. L. Wallace Hui <wlwhui@uwaterloo.ca>; Vyacheslav Lyubchich <vlyubchich@uwaterloo.ca>; Weiwen Miao <miao@macalester.edu>; Kimihiro Noguchi <kinoguchi@ucdavis.edu>. *lawstat: An R package for biostatistics, public policy, and law*, 2013. R package version 2.4.1.