ISSN 2385-2755
DiSSE Working papers
[online]

SAPIENZA
UNIVERSITÀ DI ROMA

WORKING PAPERS SERIES
DIPARTIMENTO DI SCIENZE SOCIALI ED ECONOMICHE

School Performance Gaps in Italian Regions: Estimating the Impact of Individual, School, and Territorial Factors. INVALSI 2021/22 Data Analysis

Adamo Locicero

N. 13/2024

SAPIENZA - UNIVERSITY OF ROME
P.le Aldo Moro n.5 – 00185 Roma T(+39) 0649910563
CF80209930587 – P.IVA 02133771002

School Performance Gaps in Italian Regions: Estimating the Impact of Individual, School, and Territorial Factors. INVALSI 2021/22 Data Analysis.

Adamo Lo Cicero
Sapienza University of Rome
adamo.locicero@uniroma1.it
October, 2, 2024
Doctoral School in Social and Sciences and Economics, Sapienza

Abstract

The introduction of the INVALSI standardized assessment system in Italy has revealed significant variations in upper secondary school students learning achievements across different regions. Several factors may play a role in these inequalities, and there is a theoretical debate about their different way to affect performance. The aim of the study is to investigate the territorial differentiations on student's reading performance in Italy, by analysing the impact of three families of factors - individual, school, and territorial - that act and interact at different levels. From a methodological perspective, the study uses an adapted version of the 'Coleman boat' model, originally introduced by James Coleman (1990:702). The model is specifically designed to describe the macro-micro-macro transition in the Italian educational context considered. The analyses use a specially created dataset: information on the characteristics of students and schools comes from the INVALSI 2021/22 dataset; whereas regional socio-economic information comes from the Noi.Italia (i.Stat). Because of the hierarchical structure of the dataset, the multilevel regression model technique is used besides the multiple linear regression model. This technique helps to analyse the link between the individual and contextual factors and how they affect student performance at different level, either directly or indirectly. It is hypothesized that contextual territorial factors have an impact along with individual and school variables. The results confirm the relevance of considering multiple factors at different levels when analysing the gap in reading performance among Italian students. At the micro level, differences between geographical areas primarily stem from factors such as family socio-economic background and school track preferences. School composition's importance becomes clear when advancing to a higher level in the analysis models. However, these factors alone do not fully clarify these inequalities. Analysis of third-level factors reveals the significant role played by the characteristics of the territorial context. Results stemming from these analyses are part of a broader research project on students' expectations and life trajectories.

1. Introduction

Surveys conducted at both international and national levels reveal significant territorial disparities in the learning outcomes of Italian students. Specifically, there is evidence of a disadvantage of competencies in reading, math, and science in regions of southern Italy. Of all, reading competencies represents a fundamental pillar of any educational effort, ensuring essential skills for adequate participation in modern societies (OECD, 2019a). This study investigates the territorial differentiations on reading performance by analysing the effects of three families of factors: a) differences in student and family characteristics (such as socio-economic and cultural status, gender, migration background); b) the actual differences between schools (i.e.: typology, social composition, location, track); c) the regional disparities in the areas where the students and their families live (i.e., GDP, employment rate, educational poverty). These three families of factors (individual, educational, territorial) combine differently across Italy, resulting in site-specific 'configurations'. The variability is because of events linked to public policies, the economic system, and family agency. Even in regions that are not typically seen as favourable, such as southern Italy, there are factors that have the potential to improve student achievement. The study identified and examined the mechanisms between these factors and analysed their impact on reading results at different action levels (Coleman, 1988). After a brief description of the relevant literature (section 2), an illustration of the research questions, data, statistical methods, and variables considered are presented in the paper (section 3). In the same paragraph, a description of the used methodologies and analysis models is provided. Next, the study's results are described in section 4 and discussed in the last section (section 5) in terms of policy implications and further research developments.

2. Theoretical Framework

According to OECD-PISA surveys results, Italian schools show lower performance in upper secondary learning outcomes compared with the European average (OECD, 2019b). Geographical differences linked to these inequalities are clear in the surveys conducted over time by the INVALSI National Assessment Service (SNV). Multiple factors (individual, educational, and territorial factors) may contribute to the differences in students' performance. There is a theoretical debate about how these factors and related mechanisms act while maintaining inequalities. To simplify theoretical analysis, each of these factors is now considered individually. It's important to note that none of them act alone. The observed effects on student performance are the outcome of their interconnected influence on each other. A first theoretical perspective relates the differences in school performance to the social background of the students. In the literature, researchers have approached the analysis of educational attainment inequalities by focusing on the primary effect of social background: how the attributes of students' social background directly shape their performance (Boudon, 1974; Jackson, 2013). Alongside the analysis of secondary effects: students' choice of school and their willingness to continue their studies as an indirect effect of the socio-economic and cultural context in which they grew up (Bourdieu and Passeron, 1977; Coleman, 1988; Breen and Goldthorpe, 1997). Besides these effects, some authors (Scheinder, 2014; Esser, 2016) suggest examining the tertiary effects of 'social class' on educational outcomes. These effects can be considered as an indirect representation of the teaching practices and composition of the school. Several studies show that in Italy social background plays a crucial role in educational attainment (Giancola and Salmieri 2020; Pensiero et al., 2019). There are studies that analyse the effects of diverse family backgrounds (Bernardi and Triventi, 2020). According to the Authors, students from more privileged socio-economic and cultural backgrounds pay less for their failures and are better equipped to recover from them. This is referred to as the 'compensatory advantage' (Boudon, 1974; Bernardi, 2014). Social capital and parental involvement in children's education have been extensively studied internationally, with an emphasis on their role in the early years (McNEAL Jr, 1999). Similar research has been conducted in Italy as well (Argentin and Pavolin 2013). The study suggests that the quality of time parents devote to their children's homework is a crucial factor alongside the quantity. According to Coleman (1988), the family (micro-context) and other social groups and institutional contexts (meso- and macro-contexts) shape individual aspects of social background. Moving on to the next family of factors: schools (meso context). Italian schools are required to comply with the general standards outlined by the State for teaching and learning objectives, curricula, and school regulations. As an example, the Italian education system is primarily funded by public funds and governed by formal procedures to ensure a fair allocation of resources among schools, granting them only limited organizational autonomy. The differences in performance between schools are indeed quite surprising. The situation is far more intricate. In the last twenty years, the Italian education system, particularly the school system, has undergone multiple reforms. Similar to many European school systems, it has experienced changes aiming for increased autonomy, resulting in schools engaging in competition as they strive to maintain a balance between efficiency and quality. According to certain Authors this scenario places schools in a hybrid position, subject to external control through centrally mandated policies (such as predetermined learning levels and funding), while the responsibility for accountability rests with school leaders (Van Zanten, 2005). The responsibility for delivering results and performance lies with school managers, even in the absence of complete authority over their organizations. The teacher recruitment policy provides a simple example of this. Prior to gaining a permanent position, which requires inclusion on the merit list through a public competition, teachers dedicate years to applying for the two-year Provincial Supply List (GPS). Teachers face factors of job insecurity, complications arising from frequent transfers to diverse regions, and a lack of rewarding professional trajectories. These elements play a crucial role in distinguishing schools and the quality of teaching (Argentin, 2018; Barbieri et al., 2011). Equally crucial are factors such as personalized educational approach and positive peer interdependence (Johnson and Johnson 2009). Several international and national studies show that including differentiated tracks in education systems plays a significant role in sustaining social inequities (Tarabini et al., 2022; Giancola and Salmieri, 2022). Further research highlights the pedagogical importance of heterogeneous classes (Wilkinson and Fung, 2002) and the indirect impact of school composition on enhanced school performance (Hattie, 2002; Argentin and Pavolini, 2020). Considering non-school factors that directly affect education, it is worth noting the international financial crises that have resulted in decreased public expenditures. For instance, Italy reduced education spending by 5% in the two years after the 2008 financial crisis (OECD, 2013:3), thus limiting resources for teachers and students. The Covid-19 pandemic and the subsequent economic and financial crisis, whose effects are still being experienced, are recent examples (World Development Report, 2022). The impact of these factors extends beyond public spending to include those of companies, small entrepreneurs, and families (ILO, 2020). Consequently, the macro-context in which Italian schools are situated plays a crucial role, although they receive public funding. These factors can have either beneficial or detrimental effects on fostering optimal conditions for schools and teachers to perform. They influence both families and the educational performance of students. Consider the scenario of a more affluent region with a gross domestic product (GDP) surpassing the national average, enhanced transportation infrastructure (e.g., high-speed rail), and improved housing conditions. More qualified, and experienced teachers may prefer this area over less developed areas. Schools would have a sufficient quantity and quality of teachers, granting them a competitive edge over schools in different regions. A higher GDP region leads to greater family wealth, resulting in a higher social composition in terms of average income within schools. Public policies implemented over the years have had a significant influence on many of these aspects. So far, policies targeted at territorial development and the reduction of inequalities have proven ineffective in reducing disparities. The combination of the three families of factors varies among Italian regions, and sometimes even within regions. In summary, it is essential to consider the territorial factors in students' performance gaps, empirically estimating the role of the three families of factors (students, schools, territory). The resulting information is a first step in understanding the actions that can be implemented to mitigate existing territorial school gaps.

3. Hypothesis, Data & Methods

Surveys conducted over time by the INVALSI reveal significant territorial disparities in reading performance of Italian students. Disparities in student performance can be linked to the effects of three distinct families of factors: individual, scholastic, and territorial. It is hypothesized that these factors, depending on the territorial area, act with different mechanisms and produce both direct and indirect effects on students' reading performance. The following graphic representation illustrates how the factors within the action system examined interact at different levels (Fig. 1).

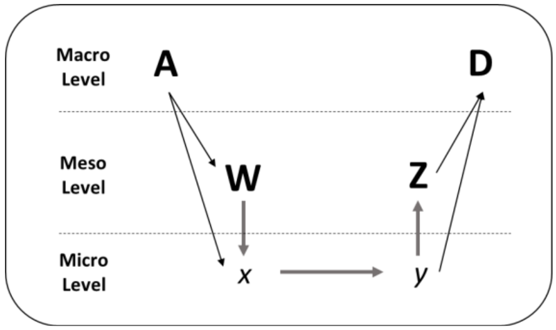

FIGURE 1. *Chaining effect between variables*
Author's adaptation of Coleman's boat model (Coleman, 1988)

The diagram represents a system with three levels of abstraction: the action system at the bottom, consisting of individual actors and their characteristics; in the middle mechanisms operating within the school system; finally, at the top, the macro-phenomenon, and collective outcomes. The aim of this study, in line with the precedent hypothesis, is to investigate the territorial differentiations on students' reading performance in Italy, by analysing the impact of three families of factors - individual, school, and territorial - that act and interact at different levels. From a methodological perspective, it is essential to:

1. Investigate the impact of first-level variables (individual_level) and second-level variables (school_level) on reading performance.
2. Examine the impact of higher level (third level) variables because of territorial differences between regions.
3. Describe the mechanisms that impact reading performance and perpetuate social inequalities.

The study relies on the secondary analysis of data from a dataset specifically constructed to meet the research objectives. The starting point was the INVALSI 2021/22 dataset, which contains data on schools and students in the 10th grade. Furthermore, the Noi.Italia dataset of I.Stat has been used to incorporate territorial-level contextual information for each region in Italy. Unlike the scenario where a single data source is employed (Lo Cicero, 2023), the constructed data set allows analysis of the impact on reading performance of three variables families: "**student variables**" individual and family characteristics of the student (student's socio-economic and cultural background, gender, number of repeats, country of origin); "**school-level variables**" the characteristics of schools (type of secondary school, the aggregate social composition at school level); "**territorial variables**" the characteristics of the region in which the school is located (GDP, unemployment rate, NEETs, family poverty). Because of the hierarchical structure of the data, the analyses were conducted using both multiple linear regression and multilevel regression techniques (Bottoni, 2022). The latter technique allows the impact of variables at all three levels, namely individual, school, and territorial. Table 1 contains a detailed description of the variables used.

Frame of reference	Database label	Description
Student variables	Gender	Gender (Female; Male)
	ESCS (student)	Social, economic, and cultural status index (student)
	IMMIG	Migration background (Native, First generation, Second generation)
	Repetition	Grade repetition (Yes, No)
School-level variables	Reading	Average scores of Reading test (WLE, average=200)
	School_track	Type of secondary school (High School, Technical and Professional)
	School_id	School identification code
	ESCS (school)	Average social, economic and cultural status index by school
	GEO_area	Geographical area (North_east, North_west, Center, South, South_Islands)
Territorial variables	GDP_reg	Regional Gross Domestic Product (value added created through the production of goods and services)
	NEET	Not in Education Employment or Training (%) 15-29-year-old
	Un_rate_reg	Unemployment rate for the region (%) 15-74-year-old
	Youth_Un_rate_reg	Youth unemployment rate for the region (%) 15-24-year-old
	25-64_less_sec_edu	25-64-year-old with lower secondary education or less (%)

TABLE 1. *List of indicators (or variables) and indices used in this paper (some have been the subject of subsequent treatments)*
Source: Dataset built by Author using dataset from: INVALSI 2021/22; Noi.Italia (I.Stat).

The factors included in this analysis affect students' performance on different levels. On a lower level, individual student factors come into play (level 1). In addition, there are school factors (level 2) and regional factors (level 3). At level 2, the analyses take into consideration the influence of school composition (even in challenging territorial conditions) and the data are grouped by school identifier variable. At level 3 the analyses take into consideration socio-economic and cultural factors (job market, unemployment rates, educational disadvantage, local GDP). Level 3 grouping variable is regional/provincial identifier as shown in Figure 2.

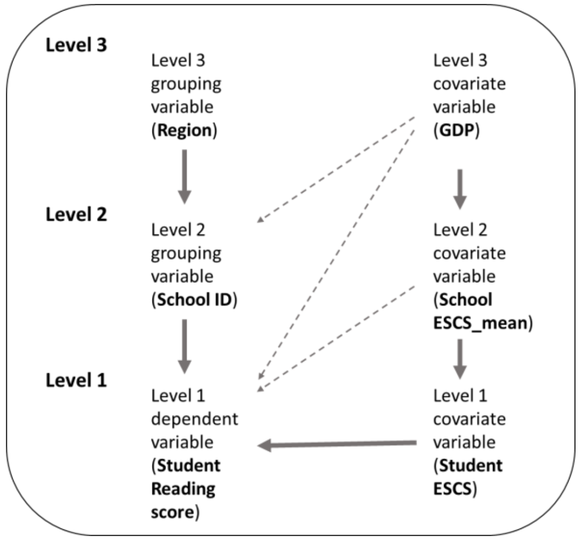

FIGURE 2*. Influence of variables on student's performance at different level*

I employed a set of statistical models, each specifically designed to address specific research goals. In the first stage of the analysis, I constructed four analytical models using multiple linear regression. This technique allows us to evaluate the effects of individual factors on student performance, without considering the hierarchical structure of data. The models required transforming several variables into dummy variables, namely: Gender, IMMIG, Repetition, School_track, and GEO_area. For the second stage, I analysed the data, considering the hierarchical nature of the dataset. In this scenario, students are within larger 2nd level contexts like schools, which are also part of a 3rd level context such as regions/provinces. The estimate is calculated for each large unit, considering the characteristics of the groupings at each level. The construction of the analysis models followed a step-by-step approach (Raudenbush and Bryk, 2002) to facilitate clear and intuitive presentation and discussion of the results. Three multilevel models were constructed to investigate the effects of the three families of factors on the reading performance of upper secondary students. The first step involved constructing the null model (M_0), which provided the basis for assessing the variance in the results at the student, school, and region/province levels.

[3.1]
$$M_0 = Null\ Model$$
$$y_{ijk} = \gamma_{000} + \mu_{00k} + r_{0jk} + \epsilon_{ijk}$$

Where school performance y_{ijk} is expressed as differences between regions μ_{00k} (difference between the average school score for each region k and the grand average of schools in all regions), plus the difference within the same region r_{0jk} (difference between the average school score for each school j and the grand average of schools in all schools in the region), plus the first-level residual ϵ_{ijk}, which corresponds to the within-school variation of the ith student compared to the school average. The intercept γ_{000} corresponds to the score of a randomly selected student from one of the k-schools. This model allows estimating the proportion of

variability associated with each level (within-schools, between-schools, and between-regions/provinces).
The subsequent multilevel models $M_{1\ and} M_2$ include not only the variables of the first level (student level) but also those of the subsequent levels (school and territorial level). Model M_1 includes school-related factors as second-level variables. There are many school factors that can influence students' academic performance, but this study focuses on school composition in terms of average social background. In the M_2 model (see formula 3.2), level 3 factors are introduced to summarise the territorial aspects.

[3.2] $$y_{ijk} = \gamma_{000} + \gamma_{001} Region_k + \gamma_{010} ESCS\_mean_{jk} + \gamma_{020} School\_id_{jk} + \gamma_{030} School\_track_{jk} + \gamma_{100} ESCS_{ijk} + \gamma_{200} Gender_{ijk} + \gamma_{300} IMMIG_{ijk} + \gamma_{400} Repetition_{ijk} + \epsilon_{ijk} + r_{0jk} + \mu_{00k}$$

3.1. Descriptive Analysis

This section provides a non-hierarchical analysis of individual and contextual characteristics of students. The impact of independent variables on the dependent variable (reading score) was estimated through linear regression modelling. The estimation process consisted of sequentially estimating four different models. Each model introduced one or more variables, with the last model being the complete model that includes all the independent variables considered (Table 2). The first model has an R^2 of 0.257. It includes individual and school choice variables. Gender (female vs. male) has a positive effect on performance. The effect of the family background (ESCS students) is more pronounced. At the individual level, this is consistent with the theoretical approaches to social and cultural reproduction (Bourdieu and Passeron, 1977). School choice partly absorbed the family background effect, confirming the importance of choosing high schools (Classical and Scientific) over other paths and in line with previous research (Giancola and Salmieri, 2022). The detrimental impact of repetition experiences and having a migrant background is unsurprising, although the second generation exhibits better performance.

	Mod1	**Mod2**	**Mod3**	**Mod4**
R^2	.257	.280	.321	.327
Gender_F	.080	.077	.072	.071
1st_gen._Student	-.060	-.064	-.076	-.076
2nd_gen._Student	-.035	-.041	-.058	-.058
Repetition	-.072	-.071	-.089	-.086
ESCS_(student)	.114	.050	.077	.045
High_sch_Class_Scien	.545	.335	.559	.441
High_sch_other_Lyc	.320	.196	.317	.251
Technical_school	.247	.174	.231	.192
ESCS_(class mean)		.155		.109
ESCS_(school mean)		.102		.034
Geo_5_North_West			.140	.142
Geo_5_North_East			.138	.134
Geo_5_South			-.067	-.036
Geo_5_South_Islands			-.111	-.092

TABLE 2. *Determinants of competence in Reading. Model on Italian upper secondary school students. Standardised Beta coefficients (N=456110). All coefficients in the table are significant for p < .000.*

Introducing background variables at class and school level in model 2 results in a reduction of the impact of school tracking. The impact of all regressors diminishes, highlighting the specific effect of school composition. Incorporating geographical macro areas in model 3 allows us to check the findings from model 1. R^2 increases (0.321), while a decrease in the effect of gender and an increase in the effect of migrant background and repetition can be observed. Including school background (class and school) in model 4 leads to a decrease the impact of school tracking and partially narrow the gap between the northern and southern regions. The analysis in this section serves as a reference for the multi-level analyses that will follow.

4. Multilevel Analysis

4.1. Variance Component Model (null model M_0)

The first multilevel model (or mix-model) has been constructed starting from equation [3.1]. It is used for comparison with the predictor-included model.

Criteria	Value
-2 Restricted Log Likelihood	4717296.048
Akaike's Information Criterion (AIC)	4717304.048
Hurvich and Tsai's Criterion (AICC)	4717304.048
Bozdogan's criterion (CAIC)	4717352.372
Schwarz's Bayesian Criterion (BIC)	4717348.372

TABLE 3. *Information criteria (null model)*

Origin	Gl numerator	Gl denominator	F	Sign.
Intercept	1	21.298	10406.821	.000

Dependent variable: Reading score WLE

TABLE 4. *Type III test of fixed effects (null model)*

The fixed effects estimate includes regression parameters. The null model only estimates the intercept.

Parameter	Estimate	Standard error	gl	t	Sign.	Confidence interval 95% Lower bound	Confidence interval 95% Upper bound
Intercept	195.664	1.918	21.298	102.014	.000	191.679	199.650

Dependent variable: Reading score WLE

TABLE 5. *Estimates of fixed effects (null model)*

The intercept (γ_{000} =195.664) represents the grand-mean reading performance across regions/provinces. It shows the predicted regional/provincial average for reading performance, and by extension, the expected score for any randomly selected Level 1 student.

Parameter	Estimate	Standard error	Z of Wald	Sign.	Confidence interval 95%	
					Lower bound	Upper bound
Residual	1063.620	2.181	487.747	.000	1059.354	1067.902
Intercept [subject = Reg_ID] Variance	70.845	23.436	3.023	.003	37.045	135.485
Intercept [subject = School_ID * Reg_ID] Variance	401.455	9.987	40.198	.000	382.351	421.514

Dependent variable: Reading score WLE

TABLE 6. *Estimates of the covariance parameters (null model)*

Table 6 contains the estimated within-group (level 1) and between-group variances (level 2 and level 3). The within-group variance in the test scores is $\sigma^2_{\epsilon_{ijk}}$ 1063.620 while, the level 2 between-group variance (reflecting the variation in the intercepts due to school level factors) is $\sigma^2_{r_{0jk}}$ 401.455. The value of regional variance (level 3) is $\sigma^2_{\mu_{00k}}$ 70.845.

The values predict a higher variability of results between schools than within the same school (evidence of clustering of level 1 units within level 2 clusters). As expected, there is a decrease in variability at the regional level when analysing the values. The first step of the multilevel analysis is now complete.

4.2. Intraclass Coefficient (ICC) Calculation

To prepare for the next steps and proceed with the construction of the saturated models, Intraclass Correlation Coefficient (ρ), or ICC, calculations are performed for each level (refer to Formula 4.1).

[4.1] $$\rho_{(level\ 1)} = \frac{\sigma^2_{\epsilon_{ijk}}}{\sigma^2_{\epsilon_{ijk}} + \sigma^2_{\mu_{00k}} + \sigma^2_{r_{0jk}}}$$

Formula [4.1] contains the variance values of student's level regressors ($\sigma^2_{\epsilon_{ijk}}$) and the variance values between school level ($\sigma^2_{r_{0jk}}$) and regional level ($\sigma^2_{\mu_{00k}}$). ICC value for level 1 ($\sigma^2_{\epsilon_{ijk}}$) is ρ =0.69 indicating the possibility of continuing with the analysis at the next levels.

ICC value for level 2 is obtained replacing the value ($\sigma^2_{\epsilon_{ijk}}$) with ($\sigma^2_{r_{0jk}}$) in the numerator. The value ρ =0.26 shows the possibility of continuing with the analysis at the next level.

Similarly, ICC value for level 3 is obtained replacing the value ($\sigma^2_{\epsilon_{ijk}}$) with ($\sigma^2_{r_{0jk}}$) in the numerator. The value observed here is quite low, at ρ =0.046. It is necessary to evaluate the possibility of proceeding with the analysis at this level. The ICC is calculated to determine if there is clustering in the data. According to Heck et al. (2014), values below .05 are typically considered as indicating a lower level of clustering. The implications of this in terms of analysis in the constructed dataset will be discussed at a later point (Section 4.4).

4.3. Model with level 2 predictor insertion (M_1)

In the second step of multilevel analysis, a two-level predictor model is used to estimate how individual, social, geographic, and contextual level 2 factors affects student performance. The model was constructed using the following parameters: **Fixed Effects**: all independent variables except ESCS_school; **Random Effects**: Covariance Type = Unstructured, Include intercept, Factors = ESCS_school; Subject grouping = School_ID; **Method of estimation**: Maximum Likelihood; **Statistics**: Parameter estimates for fixed effects; Tests for covariance parameters; Covariance of random effects.

Criteria	Value
-2 Restricted Log Likelihood	4418926.048
Akaike's Information Criterion (AIC)	4418960.048
Hurvich and Tsai's Criterion (AICC)	4418960.049
Bozdogan's criterion (CAIC)	4419164.566
Schwarz's Bayesian Criterion (BIC)	4419147.566

TABLE 7. *Information criteria (M_1)*

Parameter	Estimate	Standard error	gl	t	Sign.	Confidence interval 95%	
						Lower bound	Upper bound
Intercept	170.250	.458	3998.660	371.863	.000	169.352	171.147
Gender_F	4.985	.101	455357.138	49.338	.000	4.787	5.183
High_sch_Clas_Scien	43.143	.260	81431.452	165.771	.000	42.632	43.653
High_sch_other_Lyc	26.941	.266	74643.700	101.275	.000	26.420	27.462
Technical_school	17.498	.220	96936.331	79.640	.000	17.068	17.929
Repetition	-8.284	.138	455240.654	-60.119	.000	-8.554	-8.014
1st gen. Student	-12.655	.212	453735.581	-59.689	.000	-13.071	-12.240
2nd gen. Student	-9.171	.196	453779.450	-46.694	.000	-9.556	-8.786
ESCS (student)	2.395	.051	454543.576	46.764	.000	2.294	2.495
Geo_5_NorthW	12.720	.560	2911.698	22.693	.000	11.621	13.819
Geo_5_NorthE	13.812	.613	2794.101	22.518	.000	12.609	15.015
Geo_5_South	-4.746	.570	2894.941	-8.320	.000	-5.864	-3.627
Geo_5_South_Islands	-10.348	.598	2831.670	-17.294	.000	-11.522	-9.175

Dependent variable: Reading score WLE

TABLE 8. *Estimates of fixed effects (M_1)*

The estimates from this multilevel model confirm the relevance of students' school choice, as seen on page 7 in Mod4 of the linear regression model. It is estimated that students in high schools, in particular in the 'Classical' and 'Scientific' tracks, score 25.6 points higher in reading than those in technical schools and up to 43.1 points higher than those in vocational schools. The presence of such different outcomes across school tracks emphasizes how internal structures within educational systems can contribute to social disparities (Emmerich and Hormel, 2021). Gender continues to be important for achieving better results. Repetitions have a relatively limited impact. In relation to the migrant background, second-generation students diminish the disadvantage when compared to first-generation students (although the estimate remains negative). The individual ESCS index has a limited impact, which is partly absorbed by the school ESCS (Table 9). The effects of students' social origin come into play directly at the time of school choices and then are absorbed in the chosen school track (Boudon, 1974; Bourdieu and Passeron, 1977). Finally, the impact of geographical area variables is mitigated, because of the school's composition (Benadusi et al., 2010a; Lo Cicero, 2023). There is still a notable gap in estimates between the northern and southern regions. The performance difference between individuals in the South_Island regions and those in the North_East is estimated to be 24 points, corroborating findings from previous studies (Benadusi et al., 2010b).

						Confidence interval 95%	
Parameter		**Estimate**	**Standard error**	**Z of Wald**	**Sign.**	**Lower bound**	**Upper bound**
Residual		925.830	1.948	475.295	.000	922.020	929.656
Intercept + ESCS_school [subject = School_ID]	UN (1,1)	88.994	3.474	25.619	.000	82.440	96.070
	UN (2,1)	23.633	4.517	5.232	.000	14.779	32.487
	UN (2,2)	114.889	14.557	7.892	.000	89.625	147.275

Dependent variable: Reading score WLE

TABLE 9. *Estimates of the covariance parameters* (M_1)

The estimated value of the school-level intercept (UN=88.994) shows significant variation, showing high school-level variability in reading scores. There is a large variation (UN=114.889) in the slopes observed for the ESCS at the school level. This refers to a substantial variation in performance among students in different schools. The estimated covariance of 23.633 between the slopes and intercepts at the school level shows the impact of students' socio-economic status within the same school, proving homogeneity within the school.

4.4. Multivariate linear regression models with level 3 predictor insertion

In the third step, I attempted to apply a three-level model (M_2) with level 3 predictors to estimate the impact of spatial factors on student performance. The model was constructed using equation [3.2]. After multiple iterations of processing, I concluded that the proposed level 3 multilevel model did not adequately fit the dataset's data structure. Once I established that the use of a multilevel model was not practical, I constructed alternative models using the multiple linear regression technique,

in accordance with previous studies referenced in the literature (Benadusi et al., 2010b; Argentin et al., 2017).
There have been 12 models built in total. The first model (Table 10) includes only first and second-level regressors. Tracking and school composition factors clearly influenced performance. These results show how the initial influence of social origin (Bourdieu and Passeron, 1977) is absorbed by the second-level factors introduced, revealing secondary effect mechanisms in action (Boudon, 1974; Collins, 2000) and pointing to a chain effect between them (Giancola and Salmieri, 2020). This model will be used as a reference to analyse models 2 to 12 that account for territorial factors, with a focus on understanding further mechanisms in action.

	Mod 1	**Mod 2**	**Mod 3**	**Mod 4**	**Mod 5**	**Mod 6**
R^2	.275	.323	.317	.317	.312	.315
Gender_F	.079	.071	.073	.072	.073	.072
1st gen. Student	-.065	-.076	-.077	-.077	-.076	-.076
2nd gen. Student	-.040	-.059	-.059	-.059	-.057	-.056
Grade_repetition	-.072	-.090	-.089	-.089	-.088	-.088
ESCS (student)	.075	.062	.062	.061	.064	.063
School_track Lic. Scient. e Class.	.388	.485	.492	.493	.471	.481
School_track other Licei	.216	.267	.272	.273	.260	.266
School_track Ist. Tecnico	.193	.205	.209	.210	.205	.208
ESCS (school mean)	.194	.090	.078	.078	.098	.096
NEET_reg		-.013	.119			-.303
Un_rate_reg		-.299	-.340	-.127		
Youth_Un_rate_reg		.049		-.099		
25-64_less_sec_edu		.158			.062	.111
GDP_reg		.102			.256	

TABLE 10. *Determinants of reading competence at levels 1, 2 and 3. Model (1-6) on Italian upper secondary school students (N=456110). Standardised Beta coefficients - All coefficients in the table are significant at p < .000.*

Model 2 is the saturated model and includes all territorial regressors (level 3) besides those in model 1. The R^2 value (.323) increases significantly in this model. The negative impact of the regional unemployment rate is clear. As an indicator of the employability of the respective territories, it is likely to be more discriminating than the average-ESCS and the school-ESCS indices. An explanation for the significance of this factor is that higher rates are associated with a greater probability of students having family members who are unemployed, whether they are parents or siblings. On one hand, this diminishes the economic prospects of the family and constrains the resources for education. Conversely, it acts as a deterrent to school participation. A portion of this mechanism's impact is also manifested in the NEET variable, albeit partially absorbed by the former. However, this does not apply to youth unemployment, which shows a positive trend. Two rationales account for this. The initial consideration is the age range of the sample, which comprises individuals aged 15 to 24, a time when many young people are still engaged in their educational endeavours. Conversely, the detrimental effect of unemployment is absorbed by the previous two variables, as stated in model 4. Both GDP and individuals with less than a secondary education have a positive impact. The latter group being linked to a growing necessity for higher educational qualifications (Collins, 2011) and a family desire to improve educational attainment. It is important to mention that the impact of school tracking becomes much more pronounced in this model (compared to model 1), along with the effects of migration background and the extent of repeating grades. To illustrate the mechanisms linking these factors and their relevance, the five contextual variables were rotated two at a time at the regional level (mod 3 to 12). The results reveal variable's collinearity and significant joint role in explaining the variance of the scores (Tables 10 and 11). Collinearity among the third level variables is supported by the consistent R^2 values observed in the 3 to 12 models.

	Mod 7	Mod 8	Mod 9	Mod 10	Mod 11	Mod 12
R^2	.318	.321	.316	.315	.317	.317
Gender_F	.071	.071	.072	.073	.073	.072
1st gen. Student	-.076	-.076	-.077	-.078	-.078	-.078
2nd gen. Student	-.057	-.058	-.058	-.059	-.060	-.059
Grade_repetition	-.088	-.089	-.088	-.090	-.090	-.090
ESCS (student)	.063	.062	.062	.061	.061	.061
School_track Lic. Scient. e Class.	.483	.485	.488	.491	.494	.493
School_track other Licei	.268	.268	.271	.272	.274	.273
School_track Ist. Tecnico	.206	.206	.208	.211	.211	.209
ESCS (school mean)	.093	.092	.084	.076	.074	.075
NEET_reg			.063	-.121		
Un_rate_reg		-.328			-.174	
Youth_Un_rate_reg	-.272		-.282			-.161
25-64_less_sec_edu	.066	.127				
GDP_reg				.108	.057	.073

TABLE 11. *Determinants of reading competence at levels 1, 2 and 3. Model (7-12) on Italian upper secondary school students (N=456110). Standardised Beta coefficients - All coefficients in the table are significant at p < .000.*

Regional unemployment is alternatively paired with the NEET and youth unemployment variables in models 3 and 4. In both instances, there is a shift in sign for the two variables that are associated with young people. There is sign of collinearity among the three. Model 6 provides additional evidence, showing that the negative effect of the NEET variable increases when the other unemployment variables are not present. Model 5 shows a stronger influence of GDP when the counterbalancing effects of unemployment factors from model 2 are not present. It highlights the importance of understanding wealth distribution instead of only considering average regional wealth levels. Model 7 exhibits similar outcome as Model 6, but with the youth unemployment variable. The effect of regional unemployment in Model 8 strengthens because of the absence of other youth-related variables. The variable related to education below secondary level exhibits a stable effect. The sign of the NEET variable in Model 9 changes, but this is because NEET individuals are a subset of the unemployed population, which becomes stronger in this model. Models 10, 11, and 12 provide further evidence supporting the previous analysis showing the operating of mechanisms.

5. Conclusions

The aim of the study was to investigate the territorial differentiations on students' reading performance in Italy, by analysing the impact of three families of factors - individual, school, and territorial - that act and interact at different levels. The study used advanced multivariate analysis techniques to estimate the impact of these factors, revealing the role of different mechanisms in affecting school performance. Initially, the analyses focused on estimating the influence of first level (individual) and second level (school) factors on student performance using a multilevel regression model. The influential factors affecting school performance in Italy, such as students' family background and school track choices, have been extensively studied and confirmed by various research (Pensiero et al., 2019; Bernardi and Triventi, 2020; Giancola and Salmieri, 2020). When considering the hierarchical structure of the data, introducing school composition factors results in a reduction in the influence of all individual factors on student performance (Hattie, 2002; Wilkinson and Fung, 2002). It is noteworthy to mention that the influence of the student's family background is significantly reduced, as it is partially absorbed by the variable of school composition. Among the secondary effects, school track choices also see its impact reduced in absolute terms but continue to be a crucial factor in students' reading performance (Giancola and Salmieri, 2022). Despite relate to aspects like family background and students' motivations (Jackson, 2013; Schneider, 2014), their relevance stems from the way the Italian school system is organized (Hesser, 2016). This is a simple demonstration of how internal structures within educational systems can contribute to social inequalities (Emmerich and Hormel, 2021). The next step involved integrating territorial variables and constructing a three-level multilevel regression model. After several attempts, the dataset's structure did not fit the proposed level 3 multilevel model, resulting in unsuccessful data processing. In line with previous studies mentioned in the literature (Benadusi et al., 2010b; Argentin et al., 2017), twelve multiple linear regression models were constructed. The models shed light on the action of mechanisms generated by third-level factors. More specifically, the negative effects of high unemployment rates, besides an increased impact of school tracking factors (Collins, 2000; Giancola and Salmieri, 2020). Although the variable was collinear with the youth unemployment and NEETs variables. Performance has been positively affected by the GDP, which is also associated with the reduction of the impact of school composition. Nevertheless, the analysis of the models has revealed the importance of not only considering the average regional wealth level, but also the distribution of wealth among individuals. The positive impact of low regional educational levels is clear, as it is closely tied to the growing demand for educational credentials and a family's aspiration to improve educational attainment (Collins, 2011). The result of the analyses validates the coexistence of mechanisms that interact at various levels and show their role in generating both primary and secondary effects on reading outcomes and the perpetuation of educational inequalities. Based on the findings of the analyses conducted on the factors affecting school performance, it is worthwhile to propose a set of policy recommendations, encompassing both internal and external aspects of the education system. An initial course of action that should be undertaken relates to the decision-making process concerning the upper secondary school track. For instance, the timing of the choice could be changed to occur two years later, intending to encourage students to make more informed choices. Another suggestion is to reduce the homogeneity of classes/schools in terms of socioeconomic background of the student body. Last, there are aspects beyond the school system that can affect employment rates and unemployment levels, such as introducing policies to foster job creation and reduce joblessness. This is unquestionably a challenging issue, particularly for young individuals, considering the economic implications of the pandemic.

The present study ought to be considered as a prototype analysis and will be supplemented with further comparisons using the new data collected by INVALSI in the academic year 2022/23. The purpose is to investigate the modifications in educational inequalities caused by the potential repercussions of the pandemic and the resultant socio-economic crisis.

References

Argentin, G., Pavolini, E. (2013), *Studiare a casa e a scuola. I supporti familiari all'apprendimento e la disuguaglianza sociale*, paper presented at the VI Annual Conference Espanet Italia, Università della Calabria, 15-18 September

Argentin, G., Barbieri, G., Falzetti, P., Pavolini, E., Ricci, R. (2017), «I divari territoriali nelle competenze degli studenti italiani: tra fattori di contesto e ruolo delle istituzioni scolastiche», *Il Mulino - Rivisteweb, 1*, pp. 7-28, (Ed. gennaio-aprile)

Argentin, G. (2018), *Gli insegnanti nella scuola italiana - Ricerche e prospettive di intervento*, Bologna, Il Mulino

Argentin, G., Pavolini, E. (2020), «How Schools Directly Contribute to the Reproduction of Social Inequalities. Evidence of Tertiary Effects, Taken from Italian Research», *Il Mulino - Rivisteweb, 1*, pp. 149-176, (Ed. gennaio-aprile)

Barbieri, G., Rossetti, C., Sestito, P. (2011), «The determinants of teacher mobility: Evidence using Italian teachers' transfer applications», *Economics of Education Review*, 30, 6, pp. 1430-1444

Benadusi, L., Fornari, R., Giancola, O. (2010a), «La questione dell'equità scolastica in Italia», *Fondazione Giovanni Agnelli, Working Paper* (26)

Benadusi, L., Fornari, R., Giancola, O. (2010b), «Così vicine, così lontane. La questione dell'equità scolastica nelle regioni italiane», *Scuola Democratica*, 1, pp. 52-79

Bernardi, F. (2014), «Compensatory advantage as a mechanism of educational inequality: A regression discontinuity based on month of birth», *Sociology of Education*, 87, 2, pp. 74-88

Bernardi, F., Triventi, M. (2020), «Compensatory advantage in educational transitions: Trivial or substantial? A simulated scenario analysis», (1st ed. 2018), *Acta Sociologica*, *63*(1), pp. 40–62

Bottoni, G. (2022), *Introduzione all'analisi multilivello - Teoria e pratica dei modelli di regressione multilivello*, Milano, Franco Angeli

Boudon, R. (1974), *Education, opportunity and social inequality,* New York, Wiley

Bourdieu, P., Passeron, J.-C. (1977), *Reproduction in education, society, and culture.* Beverly Hills, CA, Sage

Breen, R., Goldthorpe, J. H. (1997), «Explaining Educational Differentials: Towards A Formal Rational Action Theory», *Rationality and Society*, *9*(3), pp. 275–305

Coleman, J. S. (1988), «Social Capital in the Creation of Human Capital. *The American Journal of Sociology, 94,* pp. S95-S120

Coleman, J. S. (1990), *Foundations of Social Theory.* Cambridge: Harvard University Press

Collins, R. 2000, *Situational Stratification: A Micro-Macro Theory of Inequality*, in Sociological Theory, v. 18, n. 1, 17-43

Collins, R. (2011), «Credential Inflation and the Future of Universities», *Italian Journal of Sociology of Education, 3*(2), pp. 228-251

Emmerich, M., Hormel, U. (2021), «Unequal Inclusion: The Production of Social Differences in Education Systems», *Social Inclusion*, Vol 9, No 3

Esser, H. (2016), «The model of ability tracking – Theoretical expectations and empirical findings on how educational systems impact on educational success and inequality», in H.-P. Blossfeld, S. Buchholz, J. Skopek and M. Triventi (Ed.), *Models of secondary education and social inequality: An international comparison,* Cheltenham-Northampton, Edward Elgar

Giancola, O., Salmieri, L. (2020), «Family Background, School-Track and Macro-Area: the Complex Chains of Education Inequalities in Italy. Working Papers 4/20», *Sapienza University of Rome*, DISSE

Giancola, O., Salmieri, L. (2022), «Chain Effects in Diachronic Perspective. Social Inequalities and School-Tracks-Choices Affecting Educational Outcomes in Italy», *Scuola democratica* (2), pp. 385-409, (Ed. maggio-agosto)

Hattie, J. A.C. (2002), «Classroom composition and peer effects», *International Journal of Educational Research*, Volume 37, Issue 5, pp. 449-481

Heck, R.H., Thomas, S.L., Tabata, L.N. (2013), *Multilevel and Longitudinal Modeling with IBM SPSS* (2nd ed.), Routledge

ILO (2020), *ILO Monitor: COVID-19 and the world of work. Second edition - Updated estimates and analysis,* Retrieved from https://www.ilo.org/wcmsp5/groups /public/---dgreports/---dcomm/documents/briefingnote/ wcms_740877.pdf

INVALSI (2022), *Rilevazioni nazionali sugli apprendimenti 2021-22,* Retrieved from https://invalsi-areaprove.cineca.it/docs/2022/rilevazioni_nazionali/rapporto /Sintesi_Prove_INVALSI_2022.pdf

Jackson, M. (2013), *Determined to succeed? - Performance versus Choice in Educational Attainment*. Stanford, California: Stanford University Press

Lo Cicero, A. (2023), «Individual and Context Variables Effects on School Performances. A Multilevel Analysis on Italian Upper Secondary Students. Outputs in the Italian Case. Working Papers 16/23», *Sapienza University of Rome*, DISSE

McNeal, R. B., Jr. (1999), «Parental involvement as social capital: Differential effectiveness on science achievement, truancy, and dropping out», *Social Forces, 78*(1), pp. 117–144

OECD (2013), *What is the impact of the economic crisis on public education spending? Education Indicators in Focus*. Retrieved from https://www.oecd.org/education/skills-beyond-school/EDIF%202013--N%C2%B018%20(eng).pdf

OECD (2019a), *PISA 2018 Assessment and Analytical Framework*. Retrieved from https://read.oecd.org/10.1787/b25efab8-en?format=pdf

OECD (2019b), *Featured country-specific overviews – Italy*. Retrieved from https://www.oecd.org/pisa/publications/PISA2018_CN_ITA.pdf

Pensiero, N., Giancola, O., Barone, C. (2019), «Socioeconomic Inequality and Student Outcomes in Italy», In L. Volante, S. V. Schnepf, J. Jerrim, D. A. Klinger, *Socioeconomic Inequality and Student Outcomes - Cross-National Trends, Policies, and Practices*. Singapore: Springer Nature Singapore Pte Ltd

Raudenbush, S. W., Bryk, A. S. (2002), *Hierarchical Linear Models. Applications and Data Analysis Methods* (2nd ed.), Thousand Oaks, CA: Sage Publications

Schneider, T. (2014), «Reviewed Work: Determined to Succeed? Performance versus Choice in Educational Attainment», *European Sociological Review*, 30, 83, pp. 410-412

Tarabini, A., Curran, M., Castejón, A. (2022), «Ability as legitimation of tracking: Teachers' representations of students in vocational and academic tracks», *British Educational Research Journal*, 48, pp. 1049–1064

Van Zanten, A. (2005), «New Modes of Reproducing Social Inequality in Education: The Changing Role of Parents, Teachers, Schools and Educational Policies», *European Educational Research Journal*, 4(3), pp. 155–169

World Development Report (2022), «Finance for an Equitable Recovery», *World Bank Group Flagship Report,* Retrieved from https://www.worldbank.org/en /publication /wdr2022

Wilkinson, I. A.G., Fung, I. Y.Y. (2002), «Small-group composition and peer effects», *International Journal of Educational Research*, Volume 37, Issue 5, pp. 425-447