



SAPIENZA
UNIVERSITÀ DI ROMA

ISSN 2385-2755
DiSSE Working papers
[online]

WORKING PAPERS SERIES
DIPARTIMENTO DI
SCIENZE SOCIALI ED ECONOMICHE

**Individual and Context Variables
Effects on School Performances. A
Multilevel Analysis on Italian
UpperSecondary Students. Outputs in
the Italian Case**

Adamo Lo Cicero



N. 16/2023

SAPIENZA - UNIVERSITY OF ROME

P.le Aldo Moro n.5 – 00185 Roma T(+39) 0649910563

CF80209930587 – P.IVA 02133771002

Individual and Context Variables Effects on School Performances. A Multilevel Analysis on Italian Upper Secondary Students.

Adamo Lo Cicero

Sapienza University of Rome

adamo.locicero@uniroma1.it

May 14, 2023

Doctoral School in Social and Economics Sciences, Sapienza

Abstract

This paper analyses inequalities in education and, in particular, the factors of social origin and school context that may influence the educational performance of Italian upper secondary school students. The study is based on data collected by OECD-PISA in 2018 and refers to performance differences in Reading. The main objective is to identify the determinants and mechanisms underlying the differences in these performances, assessing not only the impacts due to individual or ascriptive features such as gender and family backgrounds, but also the role played by contextual variables along students' careers (socio-economic and cultural contexts, school-tracks, and school composition). These are two different sets of influence, reflecting the hierarchical nature of the data. For this reason, the analyses have been carried out in two different steps. Firstly, student' individual and contextual features are analysed not considering the hierarchical nature of the data. Secondly, a multilevel methodology was adopted to identify differences among schools and within each school. Results stemming from the analyses are part of a broader research project on students' expectations and life courses.

JEL codes: C00, C42, I21

Keywords: Education, Inequality, Choices, Multilevel Analysis

1. Introduction

The Italian education system and, especially the school system has been the subject of many reforms during the last twenty years. Reforms intended to optimize public costs and make curricula more innovative and updated. A desire to encourage openness towards collaborations with external institutions and companies, encouraging capital and investments attraction has been at centerstage. The principle of school autonomy, the containment of public expenditure, and a growing international pressure to improve results have driven Italian schools toward the tricky balance between quality and performance. Despite these crucial reforms, some of which yet completed, cross-countries analyses at European level reveal the persistence of an Italian under-performance, substantial inequalities among 15-year-old students as well as remarkable geographical divides in learning outcomes (INVALSI, 2021).

The European Commission has repeatedly stressed the strategic relevance of skills acquisition (EU Think Tank, 2021) as a key-factor for citizens active role in society and in the labour market. Several factors are at play when inequalities arouse among

student's achievements. This paper especially focuses on both individual or ascriptive variables (gender and family backgrounds), and context (socio-economic and cultural contexts, school-tracks and composition, geographical divides). The former has been under the traditional attention of a growing number of research and theories (Benadusi, et al., 2008; Triventi, 2014; Giancola, et al., 2019), while the latter are a relatively fresh topic of inquire. After reviewing the relevant literature on the topic, the paper illustrates methodologies, data, and the research rationale.

2. Theoretical Framework

In the social field, many scholars have investigated the functioning of educational systems to erode educational inequalities among students. This analysis will first move on to the theoretical paradigms that have revealed how complex and evolving the relationships intertwining social background, educational careers, and educational aspirations are. Then, will be considered the hierarchical nature of the data. The first type includes models that can be identified with the social reproduction paradigm. The refined interpretative model outlined by Bourdieu (1972) stresses advantages and disadvantages directly and indirectly transmitted by parents to children as distinctive cultural factors; family cultural backgrounds can integrate the school culture and ensure better chances of success in educational paths. Even the students' attitudes and motivations towards school and educational choices in a broader sense are clear expression of the social class. With the values and culture transmitted, the family environment socializes children both to a specific social status and to a given social trajectory. Family contexts and corresponding realities in which students live and learn are going to also shape the ways they will approach the world in the future. Qualifying cultural elements such as aesthetic models, tastes, and a whole way of life the broadest sense (Bourdieu & Passeron, 1972) performs as a selection in accessing educational opportunities. According to Bourdieu, this is just one of the three stages of the process that feeds educational inequalities. Another key stage regards the probabilities of removal or endurance of students in schools, via failures, and delays. Finally, a third step relies on segregation: based on previous steps students' image and expect their future path as written on social class origins (Giancola, 2009). Bourdieu explanation for the reproduction of inequalities via educational trajectories relies on a cyclical process: schools system translates social inequalities into educational disparities and converts them back into social positions (Bourdieu & Passeron, 1972). From this point of view, while education systems seem to be neutral since they assess and certify what students learn mediated by what students possess as socio-cultural capital, education systems in fact advantage the ones belonging to the dominant codified culture and sharing values and attitudes rewardable for educational success. Furthermore, education systems tend to legitimize differences among social classes and the continuation of the existing sociocultural order. Among the so-called basic skills, reading literacy constitutes a primal and key dimension by which students' social, contextual, and familial backgrounds get translated into codified culture via educational learning and vetting. This is the reason why the analyses of educational inequalities need to investigate the determinants of students' performances in reading. The choice of this domain

is mainly remarkable since it refers to capabilities that allows students to confront written information in one or more texts for specific purposes and the future citizen to adapt his/her daily life to fast changing social settings (Perfetti, Landi, & Oakhill, 2005). Reading is not the mere ability to read a text, but rather the ability to understand and integrate what is written with pre-existing knowledge (INVALSI, 2019). Reading ability is therefore pivotal for, recalling students' possession of the inherited cultural capital and of linguistic skills that Bourdieu has identified as scholastically profitable (Bourdieu & Passeron, 1972:164). Conversely, Raymond Boudon has stressed a more individual and specific set of competencies aligning with his theoretical paradigm based on the Rational Action Theory (RTA) (Boudon, 1974; Mare, 1980; Mare, 1981; Jackson, 2013). Conceived as a convergent but alternative explanation to the bourdieusian theory RTA has emerged over the last decades, pointing out students' (and their families') conscious actions on commonsensical grounds, able to choose and opt according to their own rational principles. Individuals' family characteristics manifest in the process of choices and strategies according to patterns of class-bound conformity. Disadvantaged groups operate rationally in school decision-making adapting aspirations, expectations and will to their criteria of classification of reality based on an assessment of their social situation. Put it simply, students from lower socioeconomic backgrounds do not translate their academic performance to the same level of ambition as their privileged counterparts through their educational choices. According to this view, the actors involved in educational trajectory are driven not by exogenous pressures forcing them to choose a certain option, but rather by rational calculations of the available educational alternatives. These calculations reflect both the structural constraints of the socio-economic class they belong to and the evaluation of their previous experience in terms of educational performances. Boudon does not underestimate the weight of family backgrounds; he in fact develops a dialectical dimension between social mechanisms and individual expectations, in a preclusive of any determinism fashion (Besozzi, 2006, 175). Boudon's positional theory argues that students make different choices «according to their position in the stratification system» (Boudon 1974, 36). This position is further elaborated in two dimensions: sociocultural identity and economic rationale. The former means that students make decisions that are shaped and constrained by their family characteristics and identity. The latter is often argued to be the “rational” choice, which calculates the economic costs and benefits of a degree given extant resources and maximum long-term returns. It should be noted that the cost-benefits calculation is influenced by both cultural inheritance (constraints/opportunities) and individual experience of delays/advancements, failures/successes. Boudon's conclusion is that when educational system expands, inequalities in educational opportunities may well gradually decrease, but that does not necessarily translate into upgrading social statuses: disadvantaged families may consider the investments needed for children to attain high education levels of to be too much onerous compared to the returns of education on the labour market. Just as education inflates because of its expansion, likewise opportunities for social mobility shrink. This can apply to almost all advanced Western societies, including Italy, where the positive effects of the ever-increasing democratization of the educational system have been partly offset by the structural dimensions, such as the differentiation of school paths and the students' stratification between schools according to

their social composition. Research has shown that the students' educational trajectories are differentiated along ascribed variables (Brint , 1999; Ballarino & Checchi, 2006; Schizzerotto & Barone, 2006). The continuing expansion of the educational system generates a hierarchical differentiation that makes students from lower social classes increasingly at odds to higher levels. Other scholars, including Coleman (1988), have stressed out the dimension of social capital. This perspective considers social capital and deriving expectations, as part of a complex structure of relations among social actors. Coleman argues that social capital does not reside in individuals, nor in the physical place of production. It rather stems out from to the structure of relations between two or more persons (Coleman, 1990, p. 302). Thus, family expectations are consequently seen as being part of the social capital which in turns is made of *adaptive* skills acquired via socialization. In this context, socialization is an important resource for the individual. Coleman's theoretical contribution has the virtue of expanding the bourdieusian notion of social capital in relation to the concept of social network: something that is not naturally given, but rather implemented by means of family and individual strategies (Lin, 1999). Family expectations have long been regarded as both implicit rules that guide the behaviour of family members (Tilly, 1977, p. 5-19) and implicit or habitus principles (Bourdieu, 1983). Coleman's contribution in connecting expectations with social capital allows to re-establish intra-group interactions (micro approach) into social action at large (meso and macro approaches). Again, according to Coleman, expectations are a resource and therefore motivation for action. Finally, when looking at family relationships, parents' expectations of children's educational achievements can be seen as a *measurable* expression of the strength of family ties. Individual aspects of social background cannot be derived solely by the family of origin (micro context) but are shaped by other social groups and institutional contexts (macro context). Coleman (1990) perspective is therefore very helpful to enhance the second part of the analysis of mechanisms at play on two different levels (macro and micro) and the ways they concatenate each other, influencing individuals' actions and shaping the collective forms of students' backgrounds (micro-macro). Several studies have stated that social inequalities clearly impact on Italian students' learning outcomes and highlight single schools' social composition influence on learning outcomes (Benadusi, Fornari, & Giancola, 2010). Relevant geographical segregation between schools with different social backgrounds as well as also school choice patterns paly a remarkable role the importance of the role played by in the Italian case (Giancola & Salmieri, 2022).

3. Hypothesis, Data & Methods

At the methodological level, the aim of the paper is to perform a secondary analysis on the 2018 OECD-PISA data referring to Italian secondary school students, to:

- First, highlight the impacts of individual and ascribed factors on student performance;
- Secondly, to estimate the effects due to different school path choices;
- Finally, show the impacts of contextual factors on school performance;

Based on the points above, the study wants to describe the mechanisms in action that produce the differences in performance and the maintenance of social inequalities.

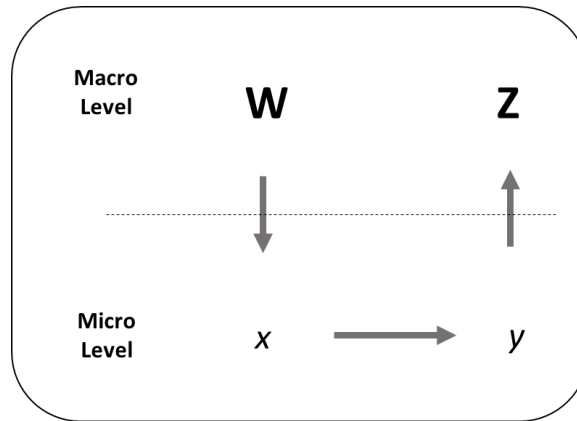


FIGURE 1. *Chaining effect between variables*
Author's adaptation of Coleman's boat model

The diagram represents a system with two levels of abstraction: the action system at the bottom, consisting of individual actors and their characteristics; at the top, the macro-phenomenon and collective outcomes. The arrows connecting the levels represent the direction of explanation. The diagram highlights three relationships: macro-micro, micro-micro, micro-macro. The macro variable refers to the school composition (W) which is hypothesised will influence the micro level variable "y" (school performance) and its effect is mediated by the student's individual and family variables "x" such as the socio-economic status index. The set of effects (direct and mediated) concatenate with each other producing an impact on the set of performance in the school context represented by the macro variable Z. In summary, can be seen a causal chain that explains through which micro-variables an association between the macro-level variables W and Z is realised (Coleman, 1990). The analysis aims at confirming the existence of a broad set of interlinked micro-mechanisms at different levels that concur to determine the school performances of Italian students. The assumptions made above were checked with data from the dataset of the 2018 OECD-PISA survey. In Italy, the survey was conducted in collaboration with the National Institute for the Evaluation of the Educational System (INVALSI). A total of 26,361 Italian secondary school classes (grade 10), with a total of 543,296 pupils, took part in the survey of the learning levels of Italian students in the classes surveyed in 2018. Although the INVALSI tests are census-based, i.e., they are taken by all students in the classes surveyed, for the OECD-PISA analyses, a sample is drawn from the total number of participating schools and classes using a two-stage method: in the first stage, schools are sampled and in the second stage, as a rule, two whole classes for each school selected in the previous stage. The 2018 sample consists of 48,664 secondary students. It is important to note that the data

collection design chosen by OECD-PISA (multistage sample) produces a hierarchical (or nested) structure: students; classes; schools; geographical area.

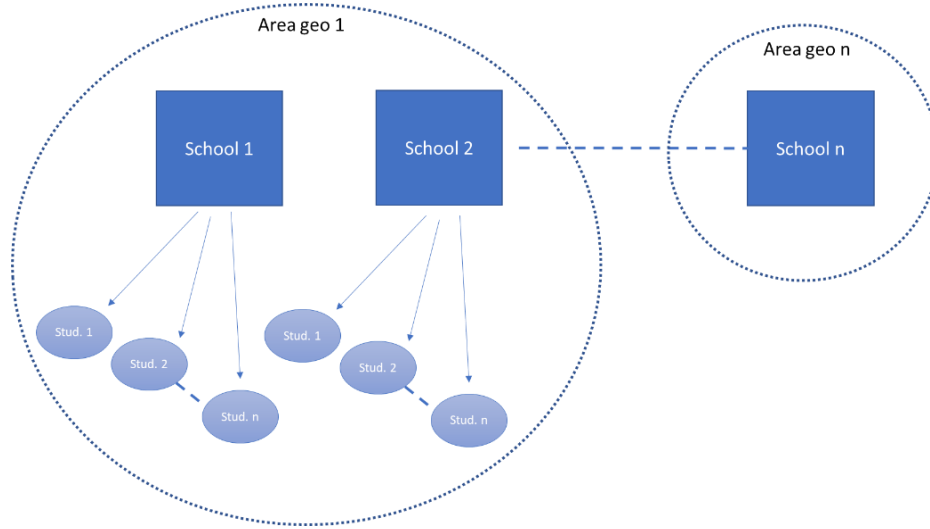


FIGURE 2. Diagram representing the hierarchical structure of the data subject to analysis

Bearing this hierarchy in mind, in the second part of the analysis, not only the individual and ascribed characteristics of individual students were investigated, but also the contextual characteristics referable to the educational institutions in which they are enrolled. This is a non-neutral aspect for the analyses, since selection processes, environmental context or self-selection allow to hypothesise that students attending the same school institution may have more pronounced similarities than those observed on pupils from other schools. Based on the above, the choice of an analysis model that can take the hierarchical structure of the data into account is of crucial importance. In fact, most classical statistical models assume independence between observations, whereas in reality individual observations embedded in hierarchical frameworks are not completely independent and ignoring this aspect produces a 'miraculous' multiplication of the number of units (Snijders & Bosker, 2012) and underestimated statistical models (Aitkin & Longford, 1986). Consequently, to obtain statistically valid analyses, this study uses a multi-level methodology to consider the hierarchical structure of the information and express the relationships between the individual and group level (Goldstein, 1987; Snijders & Bosker, 2012). Multilevel analysis can be described as an approach to analysing such data with complex patterns of variability, with a focus on the nested sources of this variability (cf. pupils in schools) and includes both statistical techniques and methodology for their correct application. In the standard linear model, the parameters generally have a fixed value estimated from the sample (a fixed effect). Generally, this is represented by the following formula [2.1] where it is assumed that the β -value is fixed:

$$[2.1] \quad Y_{ij} = \beta_{0j} + \varepsilon_{ij}$$

There are many school factors that can influence students' academic performance, but this study focuses on school composition in terms of average social origin. For this purpose, has been constructed a new variable ESCS_mean, as the average value of the level of social origin for each school. The construction started from the information on the social origin of each student (ESCS) and the school identifier to which they belong. Then has been calculated the average ESCS value for each student in the same institution. The final objective is to show what are the context effects at school level (2nd level) on students' school performance (1st level).

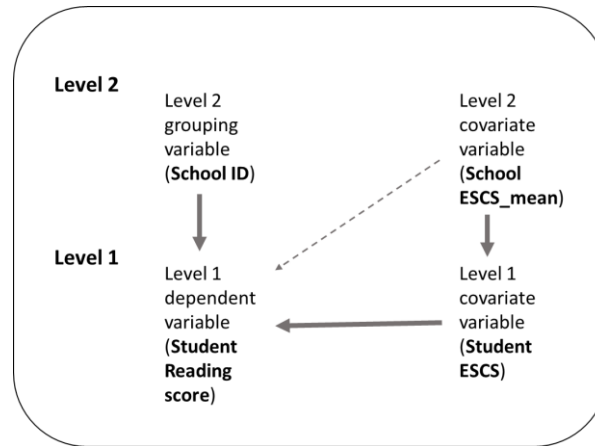


FIGURE 3. Influence of variables on student's performance at different level

The aim is to contribute to a more precise understanding and reconstruction of the complex links between students' social origins and educational inequalities.

For the construction of multilevel analysis model, a two steps procedure is used:

1. Construction of the variance component model (null model). This multilevel model without explanatory variables is useful for us to have a baseline on the variability explained for subsequent analyses once the explanatory variables (predictors) are included;
2. Calculation of the intraclass coefficient (ICC) of the percentage of total variability explained by clustering in schools;
3. Construction of the saturated model (with the inclusion of predictors) that combines the determinants identified at the micro level (students) and at the macro level (schools).

Let begin by describing how the variance component model (null model) is constructed. The null model is used in the preliminary stages of a multilevel analysis to assess whether significant differences between individuals belonging to different groups make it appropriate to apply a hierarchical model. In this case, the null model is used to estimate the effect of grouping students into schools on their performance.

[2.2]

$$Y_{ij} = \gamma_{00} + \mu_{0j} + \varepsilon_{ij}$$

Variables used: School_ID; Reading performance (see Table 1 for reference).
 Before proceeding with the second step and thus with the construction of the saturated model, the Interclass Correlation Coefficient (ρ), or ICC was calculated, which we can consider as a descriptive statistic used to find correlations within a single class of data (in this case schools).

$$[2.3] \quad \rho \frac{\sigma_B^2}{(\sigma_B^2 + \sigma_w^2)}$$

Formula [2.3] contains the variance values between (σ_B^2) the groups and the variance values within (σ_w^2) the groups.

As a second step of multilevel analysis, a model with predictors has been built. This new mixed model is used to estimate two coefficients:

- a) The variation of the intercepts μ_{0j}
- b) And the variation of the slopes of ESCS μ_{1j}

The aim is always to see how the reading performance of Italian 15-year-old students is influenced by individual factors, family background and school and geographical context.

$$[2.4] \quad Y_{ij} = \gamma_{00} + \mu_{0j} + \gamma_{01} \text{school} + \gamma_{02} \text{Gender} + \gamma_{03} \text{IMMIG} \\ + \gamma_{04} \text{AreaGeo} + \gamma_{05} \text{Repetition} + \gamma_{06} \text{ESCS}_{mean\ ij} \\ + \mu_{0j} + \mu_{1j} \text{ESCS}_{ij} + \varepsilon_{ij}$$

This is the final multilevel model [2.4] which will be used below to estimate the impact on student performance of individual, social, geographical, and contextual factors.

Variables used: School_ID; Reading performance; Gender; IMMIG; Repetition; AreaGeo; ESCS; ESCS_mean (see Table 1 for reference).

Below is a summary table of the OECD-PISA indicators and indices that were taken into consideration for the analyses produced in this study (see Table 1).

Frame of reference	Database label	Description
Schooling	Reading	Average scores of tests in Reading (standardized; range values: 0 to 1,000; average: 500)
	TFStudyProg	Type of secondary school (High School, Technical and Professional)
	School_id	School identification code
	AREA_GEO	Geographical area (North_east, North_west, Center, South, South_Islands)
	ESCS_mean	Average social, economic and cultural status index by school
Socio-biographical	TFGender	Gender (F=1; M=2)

ESCS	Social, economic, and cultural status index (student)
IMMIG	Migration background (Native, First generation, Second generation)
Repetition	Grade repetition (Yes, No)

TABLE 1. List of indicators (or variables) and indices used in this paper (some have been the subject of subsequent treatments)

Source: OECD-PISA (PISA 2018 Technical Report, 2018)

The 2018 'Reading' score of Italian students is the main dependent variable. Since there is theoretically no minimum or maximum in the PISA scores, the results are processed to conform to a normal distribution. The overall metric for the reading scale is based on an average for all participating OECD countries set at 500, with a standard deviation of 100 (OECD, 2019, p. 43). The 'ESCS' (Economic, Social and Cultural Status) index is among the independent variables central to the study. The index is the outcome of an aggregation of elementary indices such as: HISEI (highest parental occupation); PARED (highest parental education); HOMEPOS (home possessions). The three dimensions were reduced by applying the PCA technique (principal component analysis). For the purposes of the analysis, some variables were recoded to make them more balanced and suitable for further analysis and modelling. Table 2 shows the recoding information.

Original variables	Recoded variables	Method used
REPEAT	Repeat_ric	Not Repeat; Repeat
ESCS	ESCS_ric	Low; Medium; High (divided into tertiles)
TFGender	TFGender_ric	F (Female); M (Male)
IMMIG	IMMIG_ric	Native; First-Generation; Second-Generation
TFStudyProg	TFStudyProg_ric	High School; Technical; Professional
ESCS_mean		The average value of the level of social origin for each school

TABLE 2. Independent variables subject to recoding

Source: Author's elaboration

Subsequently, new dummy variables were constructed and used in the two proposed models (linear regression and multilevel analysis). Table 3 shows the complete information on the dummies included in the study.

Original variables	Dummy generated
Repeat_ric	Not_Repeat; Repeat
ESCS_ric	ESCS_Low; ESCS_Medium; ESCS_High
TFGender_ric	Gender_F (Female); Gender_M (Male)
IMMIG_ric	Bk_migr_1 (Native); Bk_migr_2 (Second-Generation); Bk_migr_3 (First-Generation)
AREA_GEO	Area_geo_North_east; Area_geo_North_west; Area_geo_Center; Area_geo_South; Area_geo_South and Islands
TFStudyProg_ric	School1_High_school; School2_Technical_school; School3_Professional_school

TABLE 3. *Independent variables transformed into dummy variables*

Source: Author's elaboration

The geographical macro-areas represented are: North-West, North-East, Centre, South and Islands. Within each macro-category we find: North-West = Liguria, Lombardia, Piemonte, Valle D'Aosta; North-East = Friuli-Venezia Giulia, Emilia-Romagna, Veneto, Trentino-Alto Adige; Centre = Lazio, Marche, Toscana, Umbria; South = Abruzzo, Campania, Molise, Puglia; South and Islands = Basilicata, Calabria, Sardegna, Sicilia.

The proposed analysis is characterized by being a prototypical analysis and constitutes the first part of a broader path that will be completed with subsequent analysis and comparisons using the new datasets produced by the OECD-PISA surveys.

4. Descriptive Analysis

In the first part of the study, the individual and contextual characteristics of the students with the data in a non-hierarchical form are analysed. The analyses conducted in this first part are:

- a) Mean comparisons
- b) Contingency tables
- c) Linear regression model

The creation of a reference snapshot allows to understand how individual and contextual variables act in relation to each other on students' performance. The first information investigated concerns the percentage of students who experienced at least one repetition during their school career. A contingency table (Table 4) is used to compare the three branches of higher education (High school, Technical and Professional school).

		Type of school rec			Total
		High School	Technical	Professional	
repeat_ric	Not repeat	94.6%	85.2%	74.7%	88.6%
	Repeat	5.4%	14.8%	25.3%	11.4%
Total		100.0%	100.0%	100.0%	100.0%
		(254,302)	(150,128)	(74,245)	

TABLE 4. Independent Contingency table between type of second grade high school (*Type_of_school_rec*) and whether or not to experience repeats (*repeat_ric*). Values in percent of column. Number of weighted valid cases = 478,675; chi-square of Pearson = 24,906.312; contingency coefficient = .222; sign. = .000

Source: Author's elaboration from PISA 2018 dataset

Analysing the table, a clear differentiation in the ratios of repeating (and non-repeating) students emerges immediately between the various Italian upper secondary school tracks. The percentage of students in the High school tracks who have experienced repetition is meagre. Compared with the Technical or Vocational school paths, there are three- and five-times higher values, respectively. It can be speculated that self-selection effects may be at work, with the best students deciding to enrol in High schools because considered better paths. Or it may be a matter of family selection, linked to their socio-economic and cultural characteristics that impose an investment in education (Boudon, 1974). It is legitimate to assume that there are other differences at play, factors that may make students in a particular type of school pathway more similar. So, will be analysed the socio-economic and performance characteristics of repeating students. In this case, given the nature of the variables, a mean comparison is used (Table 5), with both student's socio-economic and cultural status index and their reading test results.

Grade Repetition		Reading	Index of economic, social and cultural status
Did not repeat a grade	Mean	488.482	-.152
	N	442,750	441,962
	Std. Deviation	89.123	.905
Repeated a grade	Mean	401.120	-.657
	N	67,399	66,805
	Std. Deviation	90.467	.871
Total	Mean	476.940	-.218
	N	510,149	508,767
	Std. Deviation	94.074	.917

TABLE 5. Comparison of the means between the possibility of repetition (Grade_repetition) and the results in the tests in Student Reading (Reading) and the index of socio-economic and cultural status (ESCS). Reading * Grade Repetition: Eta = .314; Eta square = .099; number of weighted valid cases = 510,149; ANOVA between variance = 446,428,764.800; within variance = 4,068,364,778.000; F value = 55,979.280; sign. = .000 – Index of economic, social and cultural status * Grade Repetition: Eta = .186; Eta square = .035; number of weighted valid cases = 508,767; ANOVA between variance = 412,846.901; within variance = 427,611.020; F value = 18,194.308; sign. = .000

Source: Author's elaboration from PISA 2018 dataset

Looking at reading scores, students who have experienced one or more repetitions in their school career get a score 75 points lower than the average score. A first explanation could be related to a lower determination to succeed (Jackson, 2013), but could also be related to other factors not linked to the motivational aspects of the students. The situation is different for those who did not experience repetition, scoring 12 points above average. The first consideration of the results concerns the presence of a clear association between those who experience repeats in their school career and modest performance in Reading. Looking at the results for social background in terms of socio-economic and cultural status, is evident that the value of the social origin level index is lower in students who have experienced repeats. The distance is significant compared to the sample average and to the value recorded by students who have not experienced any repetition. Seems evident the importance of family socio-economic and cultural level in children's educational achievements as theoretical analyses shown (Bourdieu, et al., 1971). No conclusive results, of course, also because at the age of fifteen students have already spent many years at school, so it will be interesting to see what differences will be shown according to their school choices. The next analysis shows how these differences are distributed in the different geographical areas of our country. Performed using the status (ESCS) and reading performance (Reading) variables, comparing the averages with the geographical areas (Area_Geo).

Area_Geo		Index of economic, social and cultural status	Reading
North west	Mean	-0.149	498.244
	N	119,256	123,102
	Std. Deviation	0.886	90.051
North east	Mean	-0.160	501.467
	N	95,894	97,511
	Std. Deviation	0.872	87.169
Center	Mean	-0.095	484.105
	N	95,151	96,605
	Std. Deviation	0.887	89.335
South	Mean	-0.342	452.790
	N	118,831	121,130

	Std. Deviation	0.951	89.511
South and Islands	Mean	-0.358	439.257
	N	80,483	82,876
	Std. Deviation	0.954	98.427
Total	Mean	-0.219	476.285
	N	509,614	521,223
	Std. Deviation	0.916	93.853

TABLE 6. Comparison of the averages between the geographical area (*Area_geo*) and the socio-economic and cultural status index (*ESCS*) and the results in the student reading tests (*Reading*). Index of economic, social and cultural status * *Area_geo*: *Eta* = .116; *Eta square* = .013; number of weighted valid cases = 509,614; ANOVA between variance = 5,722.461; within variance = 422,280.268; *F value* = 1,726.470; *sign.* = .000 – *Reading* * *Area_geo*: *Eta* = .259; *Eta square* = .067; number of weighted valid cases = 521,223; ANOVA between variance = 307,611,073; within variance = 4,283,535,310; *F value* = 9,357.475; *sign.* = .000 Source: Author’s elaboration from PISA 2018 dataset

Table 6 shows northern Italy student' reading score is higher than in the rest of the peninsula. The North-East students experiences a better average score than in the North-West (3 points difference) and a total of 25 points above the national average. The gap between the performance results of students in the North and those in the South of Italy is truly remarkable, ranging from 49 points in the South to 62 in the areas including the Islands. Considering that the school curriculum indication is decided at a national level, such a marked difference is not attributable to the teaching factor alone. Students in Central Italy are in an intermediate position between the areas just described. Factors like the socio-economic and cultural conditions of the different areas are also analysed. The difference in the *ESCS* score level of students belonging to Northern Italy compared to the South is immediately apparent, although in this case, the highest average value is for those belonging to Central Italy. Describes very well the complex socio-economic and cultural situation experienced by those from different Italian regions. Shows how the family's socio-economic and cultural factors still play a crucial role in their children's educational pathways. This study wants to bring to light the persistence of differences that cannot be attributed to this factor alone and that relate to contextual element. For example, are analysed school choices with the geographical areas where the students live.

		Type of school rec			Total
		High School	Technical	Professional	
Area_Geo	North west	22.3%	22.8%	21.1%	22.3%
	North east	15.8%	21.9%	19.9%	18.4%
	Center	21.9%	16.7%	12.3%	18.8%
	South	22.9%	24.4%	25.8%	23.8%

	South and Islands	17.1%	14.1%	20.9%	16.8%
Total		100.0% (259,086)	100.0% (153,818)	100.0% (75,250)	100.0%

TABLE 7. Comparison Table of contingency between the choices in terms of secondary school (*Type_of_school_rec*) and the geographical area of origin (*Area_geo*). Values in percent of column. Number of weighted valid cases = 488,154; chi-square of Pearson = 7,176.973; contingency coefficient = .120; sign. = .000

Source: Author's elaboration from PISA 2018 dataset

Table 7 shows high school tracks are predominantly chosen by students in the South of Italy and the North-West, followed by the Centre, the Islands and, lastly, the North-East. Technical courses are chosen mainly by the South and the North-West, but in this case, the percentage in the North-East grows significantly and decreases in the Centre and the Islands. Enrolments in Vocational Institutes are very frequent in the South and Islands as well as in the North-West, and, if we consider what emerged in the analyses on the territorial distribution of socio-economic and cultural status, shows how school choices are more differentiated in these regions than in the Centre and North-East. Although the percentages of choice of Technical and Vocational schools are very high in the South, in purely quantitative terms High school enrolments are twice as high as in the Technical schools and about three times as high as in the Vocational ones. This choice is prevalent and in line with the phenomenon of 'liceizzazione' highlighted in many studies (Giancola & Salmieri, 2020). To study the contribution to the Reading performance made by each of the variables previously used a linear regression model is constructed. Reading' scores is the dependent variable. The analysis of this model will help to show the direction of the contribution of these factors in terms of an increase or decrease in the likelihood of students achieving superior performance on the Reading test. Will also be a valuable reference in comparison with multilevel analyses that follow. Table 8 shows the reference categories used for the independent variables entered.

Variables	Reference categories
ESCS	Cardinal variable
Area_geo	Center
Gender (of students)	Male
Immigration background	Native
Repeats	None
Type of school	Professional

TABLE 8. Independent variables included in Model 1 of linear multiple regression

Next table shows the estimated determinants of performance in Reading.

Model 1	Coefficients not standardised		Standardised coefficients	t	Sign.	95% Confidence interval for B	
	B	Standard error				Exp(B)	Lower bound
(Costant)	409.810	0.385		1,063.491	0.000	409.055	410.565
TFGender=F	7.830	0.215	0.042	36.474	0.000	7.409	8.251
IMMIG=First-Generation	-25.920	0.505	-0.058	-51.292	0.000	-26.910	-24.929
IMMIG=Second-Generation	-22.592	0.461	-0.055	-49.012	0.000	-23.495	-21.688
TFStudyProg=High school	108.524	0.297	0.581	364.929	0.000	107.942	109.107
TFStudyProg=Technical school	52.753	0.304	0.257	173.493	0.000	52.157	53.349
repeat_ric=Repeats	-50.073	0.320	-0.181	-156.455	0.000	-50.700	-49.445
ESCS low	-2.336	0.259	-0.012	-9.027	0.000	-2.843	-1.829
ESCS high	6.263	0.256	0.032	24.481	0.000	5.761	6.764
Area-geo_5=North west	30.489	0.318	0.138	95.840	0.000	29.866	31.113
Area-geo_5=North east	34.693	0.335	0.146	103.592	0.000	34.037	35.350
Area-geo_5=South	-23.315	0.321	-0.105	-72.723	0.000	-23.944	-22.687
Area-geo_5=South and Islands	-38.431	0.353	-0.149	-108.919	0.000	-39.123	-37.740

a. Dependent variable: Reading

TABLE 9. Determinants of Reading performances of Italian students (Reading) - Linear regression model. Model 1. Number of weighted valid cases = 500,470; R = .631; R² = .398; R² adapted = .398; standard error = 72.413; degrees of freedom of the model = 12; F value = 27,572.439 sign. = .000

Source: Author's elaboration from PISA 2018 dataset

The estimates shown in Table 9 tell us that: among the school paths chosen, having attended a High school has advantages in terms of being more likely to achieve high results in Reading performance than the other paths (Technical and Vocational). These results echo the evidence shown in recent years in the literature reporting how school pathways influence performance (Giancola & Salmieri, 2022). The literature also shows how having experienced less linear paths (failures/repeats) influences performance and how this is linked to the social background of the students. Also visible into the linear regression model. The results of the model reveal a low impact of social origin on performance for both categories (Low and High). The influence of socio-economic background releases its effects along the educational years, playing a role in the choice of secondary education, producing significant impact on future performance and choices. Contextual variables, such for example geographical areas of origin and having a migration background, also contribute to impacting performance. Research on the topic shows that being a first- or second-generation immigrant entails a performance gap concerning achievement. The model's results shows that the disadvantage of second-generation students decreases in the time. The massive growth of participation in school by immigrant students in recent years has fostered integration. The hypothesis is that first-generation parents have encouraged their children to acquire an elaborate linguistic code close to that of the dominant culture (Bernstein, 1971). As far as geographic areas are concerned, being part of a Northern area has more significant advantages in achieving good performance than students in the South. While there are no distinct differences between areas in the North, in the South, these differences seem to widen, especially when put in relation to the islands; live in these areas would entail lower Reading results than their peers of other areas of Southern Italy. It is assumed that these low values are due to the presence of the regions Calabria, Sicily, and Sardinia, in the South and Islands category, which are the lowest performing regions (INVALSI 2022). Finally, the model confirms gender difference to the advantage of women in Reading results, in line with literature.

5. Multilevel Analysis

5.1. Variance Component Model (null model)

Starting from equation [2.2] has been constructed the first multilevel model (or mixed model). It will be used as a reference in comparison with the second model that has explanatory variables (predictors) included. The model is applied to the reference database.

Criteria	Value
-2 Restricted Log Likelihood [Note 1]	5,860,751.157
Akaike's Information Criterion (AIC)	5,860,757.157
Hurvich and Tsai's Criterion (AICC)	5,860,757.157
Bozdogan's criterion (CAIC)	5,860,793.648

Schwarz's Bayesian Criterion (BIC)	5,860,790.648
---	---------------

TABLE 10. Information criteria (null model)

Origin	Gl numerator	Gl denominator	F	Sign.
Intercept	1	507.657	22,571.228	.000

a. Dependent variable: Reading.

TABLE 11. Type III test of fixed effects (null model)

Parameter	Estimate	Standard error	gl	T	Sign.	Confidence interval 95%	
						Lower bound	Upper bound
Intercept	464.209	3.090	507.657	150.237	.000	458.139	470.279

a. Dependent variable: Reading.

TABLE 12. Estimates of fixed effects (null model)

The table shows the value of the average of the performance averages of all the intercepts/schools $\gamma_{00} = 464.209$. This is the predicted school average for performance in Reading and, by extension, the predicted score for performance in Reading for any random student sampled at Level 1.

Parameter	Estimate	Standard error	Z of Wald	Sign.	Confidence interval 95%		
					Lower bound	Upper bound	
Residual	4,463.314	8.748	510.178	.000	4,446.200	4,480.494	
Intercept [subject = School_ID]	Variance	4,821.519	305.447	15.785	.000	4,258.528	5,458.939

a. Dependent variable: Reading.

TABLE 13. Estimates of the covariance parameters (null model)

Table 13 contains the estimated within-group and between-group variances (level 1 and level 2). The within-group variance in the test scores is σ_W^2 4,463.314 while, the between-group variance (reflecting the variation in the intercepts, which are the group mean on the dependent variable) is σ_B^2 4,821.519. The values predict a higher variability of results between schools than within the same school and can be seen as evidence of the existence of a clustering of level 1 units within level 2 clusters. This concludes the first step of the multilevel analysis.

5.2. Intraclass Coefficient (ICC) Calculation

Before proceeding with the second step and thus with the construction of the saturated model, the Interclass Correlation Coefficient (ρ), or ICC was calculated (see Formula 2.3).

ICC value is $\rho = 0.52$ indicating the possibility of continuing with the analysis at the next levels.

5.3. Model with predictor insertion (saturated model)

As a second step of multilevel analysis, final multilevel model (see Formula 2.4) is used for estimation of the impact on student performance of individual, social, geographical, and contextual factors. The model was constructed using the following parameters: **Fixed Effects**: all independent variables except ESCS; **Random Effects**: Covariance Type = Unstructured [Note 2], Include intercept, Factors = ESCS; 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 [Note 1]	5,592,721.392
Akaike's Information Criterion (AIC)	5,592,751.392
Hurvich and Tsai's Criterion (AICC)	5,592,751.393
Bozdogan's criterion (CAIC)	5,592,933.261
Schwarz's Bayesian Criterion (BIC)	5,592,918.261

TABLE 14. Information criteria (null model)

Origin	Gl numerator	Gl denominator	F	Sign.
Intercept	1	686.421	8,197.192	.000
High_school	1	14,706.853	2,224.947	.000
Technical_school	1	15,991.896	1,184.366	.000
Gender_F	1	501,065.802	5,285.161	.000
bk_migr 1st_gen.	1	500,891.579	3,920.683	.000
bk_migr 2nd_gen.	1	499,974.347	4,122.695	.000

area_geo North_west	1	461.839	24.832	.000
area_geo North_east	1	483.546	34.161	.000
area_geo South	1	465.767	17.966	.000
area_geo South & Islands	1	468.312	26.277	.000
ESCS_mean	1	715.273	27.967	.000

a. Dependent variable: Reading.

TABLE 15. Type III test of fixed effects (saturated model)

The first table to be analysed is the estimated fixed effects on reading performance due to the variables considered.

Parameter	Estimate	Standard error	gl	t	Sign.	Confidence interval 95%	
						Lower bound	Upper bound
Intercept	395.147	4.364	686.421	90.538	.000	386.577	403.716
High school	118.413	2.510	14,706.853	47.169	.000	113.493	123.334
Technical school	69.889	2.031	15,991.896	34.415	.000	65.908	73.869
Gender_F	15.085	.207	501,065.802	72.699	.000	14.678	15.491
bk_migr 1st_gen.	-27.560	.440	500,891.579	-62.615	.000	-28.423	-26.697
bk_migr 2nd_gen.	-31.739	.494	499,974.347	-64.208	.000	-32.708	-30.770
area_geo North_west	29.556	5.931	461.839	4.983	.000	17.901	41.211
area_geo North_east	28.436	4.865	483.546	5.845	.000	18.876	37.995
area_geo South	-25.717	6.067	465.767	-4.239	.000	-37.639	-13.794
area_geo South & Islands	-27.179	5.302	468.312	-5.126	.000	-37.597	-16.760
ESCS_mean	20.505	3.877	715.273	5.288	.000	12.893	28.118

a. Dependent variable: Reading.

TABLE 16. Estimates of fixed effects (saturated model)

Compared to the results shown above in the linear regression model, the impact of the variables as a fixed effect shows differences between geographical areas become smaller and, to a greater extent, between the southern areas. This is because of school composition, which tends to level out the differences by making them homogeneous within the same institutes and keeps the distances unchanged between institutes. Weighing further in the model is the average school ESCS, which shows how the influence of the socio-economic status of students belonging to the same school, and thus the degree of homogeneity within the school, has advantages for them in terms of results. Migration background estimation results show a disadvantage in terms of the possibility of good performance for second-generation students. Being female acquires significance in determining better results than in the model presented above. Finally, the importance of the chosen secondary school in determining a higher probability of success in test results is confirmed.

Parameter		Estimate	Standard error	Z of Wald	Sign.	Confidence interval 95%	
						Lower bound	Upper bound
Residual		4,081.921	8.163	500.046	.000	4,065.953	4,097.952
Intercept + ESCS [subject = School_ID]	UN (1,1)	1,580.773	108.571	14.560	.000	1,381.679	1,808.555
	UN (2,1)	295.636	47.299	6.250	.000	202.932	388.340
	UN (2,2)	487.419	35.466	13.743	.000	422.636	562.132

a. Dependent variable: Reading.

TABLE 17. *Estimates of the covariance parameters (saturated model)*

In the table, the value UN (1,1) corresponds to the variance of the intercept, UN (2,2) corresponds to the variance of the slope of the ESCS regression coefficient line, and UN (2,1) corresponds to the covariance between the intercepts and slopes. In the model, the value 0 (zero) is not included in the confidence interval for UN (2,1); thus, we can reject the null hypothesis. The table shows that the value of the between-school variance is higher than the within-school variance. The value of the composite variance with the effects of ESCS variable is also reported. The school selected is characterised by an internal homogeneity that indirectly influences all students, showing a school composition effect. This is not happened between schools, as heterogeneity does not allow such a direct impact. These data are essential to understand why, given student performance, some schools seem to perform well and others less so. In addition of the estimation of the intercept of the average socio-economic and cultural status of the student on a school basis, the model also includes the estimated slope variability on the student' ESCS to investigate the composition and individual effects simultaneously. The covariance value (295.636) high-

lights how both school mean and individual ESCS effects impact student performance through what might be called a chaining effect between the dimensions examined. Concerning the variance values, results have confirmed the importance of the intra-school homogeneity component in favouring the indirect transfer of the school's socio-economic status to students. The model's confidence interval on estimating the variability of the predictors is entirely significant for both fixed effects and covariances. Below is the graphical representation (Figure 4) of the fixed values of the estimates of the variables considered.

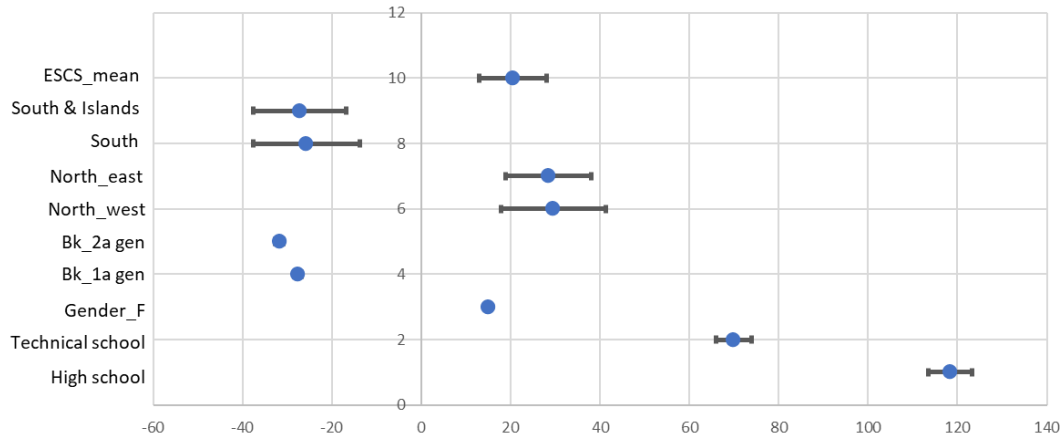


FIGURE 4. *Multilevel model saturated - Fixed effects and confidence intervals*

Conclusion

Analyses confirmed the co-presence of mechanisms interacting at different levels showing how they produce both primary and secondary effects on the outcomes in Reading and maintenance of educational inequalities. Starting with the primary effects, at the individual level, analyses show that students with high socio-economic and cultural status are oriented towards high school rather than technical or vocational paths. Students enrolled in these paths experience lower chances of repetition and achieve significantly higher performance on Reading tests. These results tend to confirm Bourdiesian theoretical positions (1972) and show how, in Italy, ascribed characteristics and family background still have a substantial weight in influencing the choices of secondary school. This entails direct effects at the level of school composition, feeding the chaining effects due to the indirect impacts produced by the variability between schools. The multilevel analysis allowed to estimate how much of the total variability can be attributed to factors for each level. It also unveils the existence of different relationships between individual and contextual characteristics (i.e., attributable to belonging to a given group) and their reciprocal relationships (interaction between levels). The application of the model shows that performance in Reading is influenced by a set of mechanisms acting in conjunction

with individual student characteristics. School composition in terms of average socio-economic and cultural status ties in with individual factors, indirectly impacting the performance of individual students. This influence is linked both to the choice of pathway, which we know to derive from family background and the choices made by the student in terms of costs and benefits (Boudon, 1974). The introduction of the second level of analysis (school level) shows how the direct impact of individuals' family origins reduces its effects once to school because contextual variables absorb it. These concatenated effects between the factors are the product of the social capital that students possess (Coleman, 1988). The complexity of the mechanisms impacting school performance described in this study does not provide easy policy directions to implement. The policy instruments adopted in recent years in Italy aimed at rebalancing the secondary school chains through the adoption of curricular reforms and managerial tools for the Technical and Professional tracks, but they have not had the desired effect. Students enrolled in High school maintain a substantial advantage in test results even when the variability between different schools and regions is considered. There is a sort of segregation effect where the aggregate utility through the choices of individuals concatenates with the tracking effect, determined by the structure of the Italian educational system, producing cumulative effects that are strongly determinant for the educational performance of Italian students and the persistence of social inequalities in education. The present study is to be considered a prototype analysis and will be supplemented in the future with further comparisons using the new data collected by OECD-PISA in the school year 2021/22. The intention will be to study how inequalities in education will have also changed because of the possible impacts of the pandemic and the resulting social-economic crisis.

Notes

[1] We can compare different models by looking at the difference in the value of $-2LL$ (Log Likelihood).

[2] SPSS Statistics v25 has 17 covariance structures that you can use. In the model is used “unstructured” covariance structure selection.

References

- Aitkin, M., & Longford, N. (1986). Statistical Modelling Issues in School Effectiveness Studies. *Journal of the Royal Statistical Society*, 149, 1-43.
- Ballarino, G., & Checchi, D. (2006). *Sistema scolastico e disuguaglianza sociale. Scelte individuali e vincoli strutturali*. Bologna: Il Mulino.
- Ballarino, G., & Panichella, N. (2016). Social stratification, secondary school tracking and university enrolment in Italy. *Contemporary Social Science*, 1(14).
- Barone, C., Barg, K., & Ichou, M. (2021). Relative risk aversion models: How plausible are their assumptions? *Rationality and Society*, 33(2), 143–175.

- Benadusi, L., Fornari, R., & Giancola, O. (2010). La questione dell'equità scolastica in Italia. *Fondazione Giovanni Agnelli, Working Paper* (26).
- Bernstein, B. B. (1971). *Class, Codes and Control - Theoretical Studies towards a Sociology of Language (Volume I)*. London and New York: Routledge & Kegan Paul Ltd.
- Boudon, R. (1974). *Education, opportunity and social inequality*. New York: Wiley.
- Bourdieu, P. (1983). Vous avez dit "populaire"? *Actes de la recherche en sciences sociales*, 46(pp. 98-105).
- Bourdieu, P., & Passeron, J.-C. (1972). *La riproduzione. Per una teoria dei sistemi di insegnamento*. Rimini: Guaraldi.
- Brint, S. (1999). *Scuola e società*. Bologna: Il Mulino.
- Cavalli, A., & Facchini, C. (2001). *Scelte cruciali - Indagine LARD su giovani e famiglie di fronte alle scelte alla fine della scuola secondaria*. Bologna: Il Mulino.
- Checchi, D. (2006). *The economics of education: Human capital, Family Background and Inequality*. Cambridge MA: Cambridge University Press.
- 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. (2011, June). Credential Inflation and the Future of Universities. *Italian Journal of Sociology of Education*, 3(2), 228-251.
- Giancola, O., & Salmieri, L. (2020). *Family Background, School-Track and Macro-Area: the Complex Chains of Education Inequalities in Italy*.
- Giancola, O., & Salmieri, L. (2022, maggio-agosto). Chain Effects in Diachronic Perspective. Social Inequalities and School-Tracks-Choices Affecting Educational Outcomes in Italy. *Scuola democratica*(2), pp. 385-409.
- Goldstein, H. (1987). *Multilevel Models in Educational and Social Research*. London: Griffin.
- Hedstrom, P. (2005). *Dissecting the Social - On the Principles of Analytical Sociology*. New York: Cambridge University Press.
- INVALSI. (2019). *Rapporto Nazionale e Sintesi dei risultati PISA 2018*. Retrieved from https://www.invalsi.it/invalsi/ri/pisa2018/docris/2019/Rapporto_Nazionale.pdf
- INVALSI. (2021). Retrieved from <https://www.invalsi.it/invalsi/index.php>
- Jackson, M. (2013). *Determined to succeed? - Performance versus Choice in Educational Attainment*. Stanford, California: Stanford University Press.
- Lin, N. (1999). *Building a Network Theory of Social Capital*. Retrieved from [insna.org: http://www.insna.org/PDF/Connections/v22/1999_I-1-4.pdf](http://www.insna.org/PDF/Connections/v22/1999_I-1-4.pdf)
- Lucas, S. R., & Beresford, L. (2010). Naming and Classifying: Theory, Evidence, and Equity in Education. *Review of Research in Education*, 34(25).

- Mare, R. D. (1980). Social background and school continuation decisions. *Journal of the American statistical association*, 75 (370), 295-305.
- Mare, R. D. (1981). Change and stability in educational stratification. *American sociological review*, 72-87.
- OECD. (2012). *Grade Expectations: How Marks and Education Policies Shape Students' Ambitions, PISA*. OECD Publishing.
- OECD. (2019). PISA 2018 Reading Framework. In OECD, *PISA 2018 Assessment and Analytical Framework*. OECD.
- OECD. (2019). *PISA 2018 Results (Volume I)*. OECD.
- Panichella, N., & Triventi, M. (2014). Social Inequalities in the Choice of Secondary School. *European Societies*, 16(5 (pp. 666-693)).
- 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.
- Perfetti, C., Landi, N., & Oakhill, J. (2005). The Acquisition of Reading Comprehension Skill. In M. Snowling, & C. Hulme, *The Science of Reading: A Handbook*. Oxford: Blackwell Publishing Ltd.
- PISA 2018 Technical Report. (2018). *PISA 2018 Technical Report*. (OECD) Retrieved from <https://www.oecd.org/pisa/data/pisa2018technicalreport/>
- Schizzerotto, A., & Barone, C. (2006). *Sociologia dell'istruzione*. Bologna: Il Mulino.
- Snijders, T. A., & Bosker, R. J. (2012). *Multilevel analysis – An introduction to basic and advanced multilevel modelling - 2nd Edition*. London: Sage.
- Tilly, C. (1977). *From Mobilization to Revolution*. Michigan: University of Michigan.
- Vergati, S. (2005). *La costruzione intergenerazionale delle aspettative*. Roma: Bonanno.