AGENDA 2030: SURVEY ON SUSTAINABLE DEVELOPMENT GOALS THROUGH INVALSI DATA

Edited by Patrizia Falzetti

FrancoAngeli
OPEN ACCESS

INVALSI PER LA RICERCA
STUDI E RICERCHE

INVALSI

INVALSI PER LA RICERCA

La collana Open Access INVALSI PER LA RICERCA si pone come obiettivo la diffusione degli esiti delle attività di ricerca promosse dall'Istituto, favorendo lo scambio di esperienze e conoscenze con il mondo accademico e scolastico.
La collana è articolata in tre sezioni: "Studi e ricerche", i cui contributi sono sottoposti a revisione in doppio cieco, "Percorsi e strumenti", di taglio più divulgativo o di approfondimento, sottoposta a singolo referaggio, e "Rapporti di ricerca e sperimentazioni", le cui pubblicazioni riguardano le attività di ricerca e sperimentazione dell'Istituto e non sono sottoposte a revisione.

ISBN 9788835123330

ISBN 9788835123330

AGENDA 2030: SURVEY ON SUSTAINABLE DEVELOPMENT GOALS THROUGH INVALSI DATA

Edited by Patrizia Falzetti

INVALSI

FrancoAngeli

OPEN ACCESS

ISBN 9788835123330

Le opinioni espresse nei lavori sono riconducibili esclusivamente agli autori e non impegnano in alcun modo l'Istituto. Nel citare i contributi contenuti nel volume non è, pertanto, corretto attribuirne le argomentazioni all'INVALSI o ai suoi vertici.

Grafica di copertina: Alessandro Petrini

ISBN 9788835123330

Index

ISBN 9788835123330

ISBN 9788835123330

Introduction

by Patrizia Falzetti

Agenda 2030 for Sustainable Development is a shared action plan for people, for the planet and for prosperity, adopted in September 2015 by the 193 United Nations (UN) Member States. The guidelines of this journey are summarized by 17 Sustainable Development Goals (SDGs) and the associated 169 Targets approved by the UN, with the shared aim to reach them by 2030.

The SDGs take into account, in a well-balanced way, the three fundamental dimensions of sustainable development: economic growth, social inclusion and environmental protection. The wish is that the goals setting and achievement can promote interventions in these areas and ensure a better present and future for our planet and for all the people that live on it.

The SDGs are deeply interconnected: ensure inclusive and equitable quality education for all (Goal 4) means to offer equal opportunities to women and men (Goal 5); to ensure healthy lives and promote well-being (Goal 3) requires a healthy planet (Goal 6, 13, 14 and 15); decent work for all (Goal 8) needs to reduce the inequality (Goal 10). This volume, composed of 7 chapters, discusses 3 of the SDGs: to ensure inclusive and equitable quality education, to achieve gender equality and to reduce economic inequality within and across national borders.

To study these phenomena, the authors of the volume, that work all for the INVALSI Statistical Office, have exploited the databases of the Institute. The INVALSI national and international surveys provide a valuable resource to investigate the characteristics of the Italian school system, to define the potential support and development interventions and they allow comparing the students' achievement and the learning context of several Countries.

The first two chapters are tied to Goal 10. In the first chapter, the inequalities in the Italian educational field are described; these inequalities result

ISBN 9788835123330

in an unequal distribution in the Italian territory of the possibility to reach an adequate learning level. In the second chapter, the social and economic differences are analysed in relation to a topic as current as ever: the Covid-19 pandemic. The emergency situation has highlighted how the above-mentioned differences are not restricted to the healthcare system but they affect many other aspects of human life, including education.

The third and fourth chapter focuses on Goal 5 (achieve gender equality and empower all women and girls). In the third chapter, the gender gap is studied analysing the trend of student's academic performance from the first to the second cycle of education; in the fourth chapter, the same theme is investigated analysing the results in Maths with respect to the socio-economic background in order to verify if better conditions lead to a wider gap.

The fifth and the sixth chapter (Goal 4) analyses the academic achievement of the foreign students. The extended literature on this topic shows that performances are poorer for immigrant students than their native peers. In chapter five, it is exploited the data from the international survey, OECD PISA 2018, whereas, in the following chapter the learning differences between the two groups are studied using the data from the national survey of the scholastic year 2018/2019. The seventh and last chapter (Goal 4) investigates a problem that requires an appropriate and effective political response: the school dropout. The phenomenon is itself complex and it is characterized by the combination of personal, social, economic, and educational factors, usually associated with a condition of socio-economic disadvantage. The chapter allows monitoring the transition of the students between the lower and the upper secondary school, a delicate and fundamental step of the student's life in which dropout prevention programs are essential.

Aware that ensuring inclusive and equitable quality education for all is fundamental to improve the lives of people, we wish the volume reading encourages the discussion about possible ameliorative interventions and it is a starting point to measure potential progress.

ISBN 9788835123330

1. Territorial inequalities in education: some aspects related to the territory

by Emiliano Campodifiori, Patrizia Falzetti, Michele Marsili

This work aims to describe the existence of inequalities in Italy's educational field, which translate, in practice, into an unequally distributed possibility of achieving an adequate level of learning. Alongside the description of the phenomenon by means of cartograms that illustrate the regional percentage distribution of students in difficulty or in implicit drop-out[1] in the last year of upper secondary school, a research is conducted on the potential determinants of these territorial inequalities; for this purpose, multiple and logistic regression models are used to evaluate the impact on the proficiency level of a series of variables related to students (socio-economic cultural background, origin of the student etc.) and to the municipal area of the high school attended (unemployment rate, size of the Municipality, etc.). The proficiency level is measured by standardized test carried out by INVALSI (The Italian National Institute for the Evaluation of the Education and Training). The study ends with a cluster analysis in order to group the Italian municipalities by similar territorial characteristics and to highlight any associations with the percentage of students in difficulty. In the analyses presented, we went beyond the regional average, using variables at the municipal level in order to intercept determinants related to the sub-regional territory that have proved to be significant for the students' achievement; regression models suggest that municipal territorial characteristics such as the youth unemployment rate and the demographic size of the municipality have a significant impact on the level of learning achieved and/or on the probability of reaching the end of the upper secondary school without repetition; the cluster analysis

[1] By "students in difficulty" we mean those who do not reach the expected proficiency level in Italian or Mathematics at the end of the upper secondary school; with "students in implicit drop-out" we intend here those who do not reach the expected proficiency level at the end of the upper secondary school in any of the subjects covered by the INVALSI assessment.

ISBN 9788835123330

finally made it possible to identify territorial realities in contrast with the regional averages, “excellent” contexts within regions or macro-areas characterized by a low level of learning.

1. Literature

On September 25, 2015, the United Nations approved the 2030 Agenda for sustainable development and the 17 Sustainable Development Goals (SDGs in the English acronym), divided into 169 targets to be reached by 2030.

The document expresses a clear opinion on the unsustainability of the current development model on different levels, environmental, economic and social. All countries are called to participate, each according to the level of development achieved. In essence, each country is asked to define a sustainable development strategy that allows reaching the SDGs, accounting for the results obtained through a process coordinated by the UN. In order to achieve the objectives, it is necessary to involve all components of society from businesses to the public sector, from civil society to philanthropic institutions, from universities and research centers to information and culture operators.

In the list of 17 Sustainable Development Goals (SDGs), objective 10, Reducing inequality within and among nations is the one taken into consideration in this work.

2. Objective

The object of this work is to identify the existence of educational inequalities within Italy. A targeted study will show the situation throughout the national territory, demonstrating that access to education is not equally distributed throughout the national territory as well as the achievement of adequate levels of functional learning to the pursuit of active citizenship.

After the description of the aforementioned phenomena, the work continues by analyzing some of the potential determinants of this inequality in order to be able to get ideas for intervention policies, also going to “isolate” or circumscribe geographic environments characterized by common aspects that lead to educational failure. The whole study focuses on the identification of areas of the country characterized by the phenomenon of inequality in order to verify in the following years its reduction as foreseen in the objectives of Agenda 2030.

ISBN 9788835123330

3. Research hypotheses

The regional differences in Italy have been studied and are evident (INVALSI, 2019) as shown in Figures 1, 2 and 3. The first two figures shows, differentiating by school subject, the percentage of students in difficulty in the last year of upper secondary school, i.e. students who do not reach expected proficiency level[2] in Italian literacy or Mathematics at the end of the upper secondary school (Loth, 2019).

Figure 3 shows an even more worrying overview as it represents the distribution of the so-called "implicit drop-out" (Ricci, 2019b) phenomenon, that is when a student do not reach the expected proficiency level at the end of the upper secondary school in Italian literacy, Mathematic and English Language (Reading and Listening).

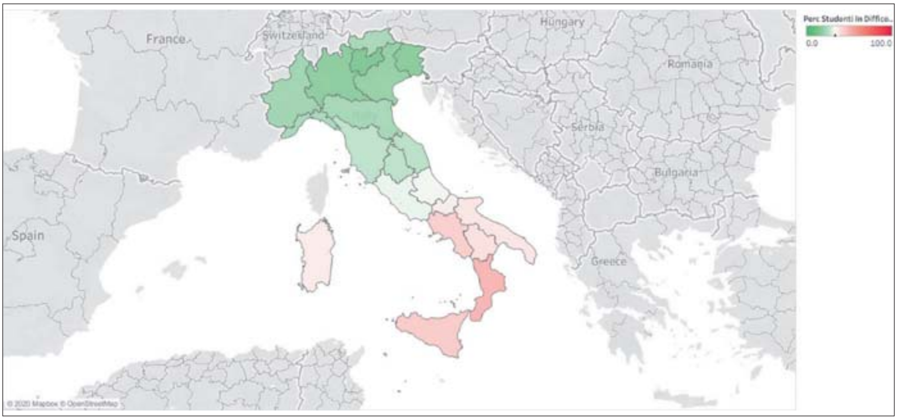


Fig. 1 – Percentage of students in difficulty in Italian literacy in the last year of the upper secondary school by region. The center value of the green (less critical) -red (more critical) diverging colour palette is 33%

[2] The proficiency level in a given subject is the level assigned to the student based on the score obtained in the standardized test administered by INVALSI, the National Institution that carries out the learnings assessment in Italy. Each proficiency level is associated with a description of the specific skills student has acquired in that subject; this classification includes 5 levels in Italian literacy and Mathematics (1 is the lowest proficiency level, 5 the highest) and 3 in English (Pre-B1, B1 and B2 with B2 the highest proficiency level); the expected level of competence is the level corresponding to the set of skills considered minimum or sufficient: level 3 in Italian literacy and Mathematics, level B1 in English.

ISBN 9788835123330

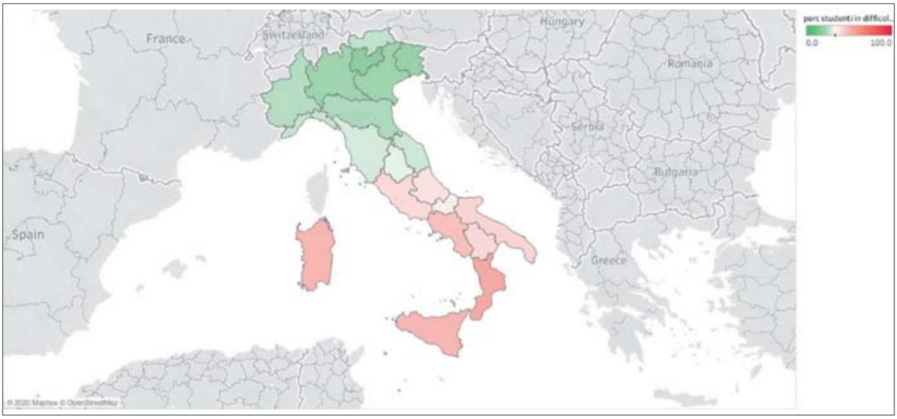


Fig. 2 – Percentage of students in difficulty in Mathematics in the last year of the upper secondary school by region. The center value of the green (less critical) -red (more critical) diverging colour palette is 33%

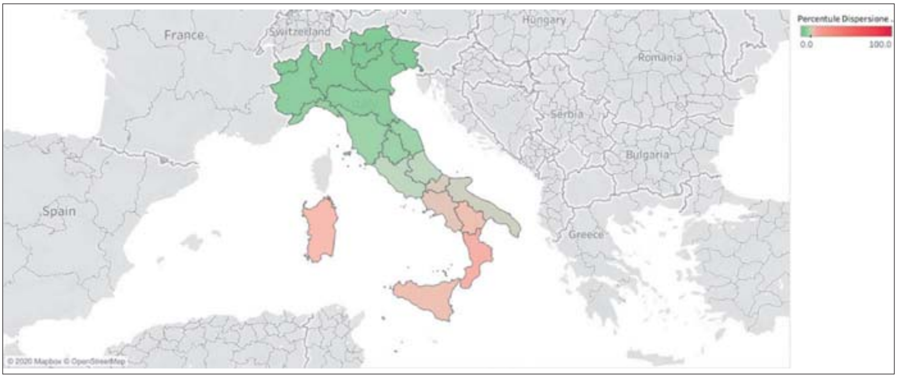


Fig. 3 – Percentage of students in implicit drop-out in the last year of the upper secondary school by region. The center value of the green (less critical) -red (more critical) diverging colour palette is 33%

The geographical differences are very clear if we consider the regional territories. The present work tries to find which characteristics can have more or less impact on the inequalities highlighted starting from a geographical breakdown at the municipal level, finer than regional data.

ISBN 9788835123330

4. Data

INVALSI periodically detects students' learning at the end of certain specific school grades.

In particular, in primary school, the survey, in paper format, takes place in grades[3] 2 and 5; in lower secondary school, it takes place in grade 8, up to the school year 2016/17 in paper form and by the SY 2017/18 in Computer Based Test (CBT) mode; in upper secondary school from the SY 2017/18 in CBT mode (previously on paper) involves grade 10 and, from the SY 2018/19 in CBT mode, also grade 13 (INVALSI, 2019).

The data used take into account all the cohort that in the SY 2013/14 completes grade 9 and which should therefore have reached the end of the SY 2018/19 in the fifth year of the upper secondary school.

The 2013/14 school cohort at 8th grade (III class of lower secondary school) is equal to 515,543 students, of whom about 67% take the grade 13 (last year of the upper secondary school) test in 2019, i.e. without having repeated a single year throughout the secondary school cycle. In 2016, 11% of students repeat grade 9, a slice is the one that goes to vocational training, the rest either repeats the years during the upper secondary school or leaves school[4].

The test taken by students in 8th grade in 2014 is paper-based, the one taken by students in grade 13 in 2019 is in CBT format. A clarification on the two types of models is due: the first in paper format must take into account the so-called *cheating* effect, the second instead is not affected by any type of external disturbance (Ricci, 2019a).

Cheating is a phenomenon that spread like wildfire for the first years of administration of the INVALSI tests and in some regions in particular; it is an opportunistic behaviour held in class by students (*student cheating*) or by the teacher (*teacher cheating*) that affects the result of some schools that find themselves having overestimated results compared to the real situation. *Student cheating* is the action of students to copy from other students or from books or other sources; *teacher cheating* is an action conducted by teachers, intended to provide answers to students, or to let students do *student cheating* (Agasisti, Falzetti and Freddano, 2015).

[3] By school "grades" we mean the year of schooling attended by the student: in primary school we have the first 5 grades, in lower secondary school we have grade 6 7 and 8 and finally in upper secondary school we have grades from 9 to 13.

[4] Obviously there is a small missing portion due to mismatch of SIDI codes (serial number attributed to the student upon entering the school system) of the student; private students and those who have not voluntarily taken the test must also be considered missing.

ISBN 9788835123330

To deal with this problem, INVALSI uses a procedure for identifying the so-called "propensity" to *cheating* index which is used to correct the observed score obtained by students (Longobardi, Falzetti and Pagliuca, 2018). Therefore, the student scores obtained in 2014 are re-modulated according to the procedure described above.

5. Methods

We used for this analisys two kind of models: a regession analysis (linear and logistic) and a cluster analisys (Anderson, 2003).

The first aspect analyzed was the effect on the dependet variable (the score in Italian literacy and in Mathematics at the end of the upper secondary school) of some individual and geographical variables.

This effect has been studied through a multivariate linear regression model, such as the one indicated below:

$$Y_{\mathrm{i}} = \alpha + \beta_1 X_{i1} + \beta_2 X_{i2} + \ldots + \beta_n X_{in} + \varepsilon_i \qquad [1]^5$$

The second aspect addressed is the analysis of the student's probability to reach the last year of the upper secondary school without repetition.

Logistic regression[6] (Mccullagh and Nelder, 1989) is used when the dependent variable Y is categorical. In the specific case of this work, the dependent variable is dichotomous, that is, it assumes only two values: it will

[5] Where:
Y_i = performance of the "i" student for i=1,..,m;
α = the intercept, i.e. the average value assumed by Y when $\{X_1, \ldots, X_n\} = 0$;
β_1 = the first *linear regression coefficient* that expresses the average variation of the variable Y for a unit variation of the variable X_1;
β_n = n-th *linear regression coefficient* that expresses the average variation of the variable *Y* for a unit variation of the variable X_n;
X_{i1} = first explanatory variable, for example the "i" student's entry score;
X_{in} = n-th explanatory variable, for example the "i" student's socio-economic cultural background;
ε_i = is the error term, i.e. the difference between the Y value actually observed and the theoretical value predicted by (1). This difference represents the part of data variability that the set of explanatory variables cannot explain.

[6] Logit models take a general form of:

$$P(Y = 1|X_1, \ldots, X_n) = \frac{e^u}{1 + e^u}$$

where is the classic linear regression
$u = A + B_1X_1 + B_2X_2 + \ldots + B_kX_k$

ISBN 9788835123330

be “1” when the student has reached grade 13 without losing any year during the 5 years of upper secondary school, “0” in the remaining cases.

With this method it is possible to calculate the probability that the dependent variable is equal to 1 conditional on different explanatory variables inserted in the models.

The third and final aspect dealt with the identification of groups of similar schools by profile through cluster analysis.

Cluster analysis is an unsupervised data mining technique that “learns” from the available data without receiving examples from already labelled data.

A *clustering* algorithm[7] is able to group together objects having similar characteristics, therefore clustering means the segmentation of a heterogeneous group into homogeneous subgroups (clusters).

6. Results and discussion

The first approach presented in this paper is the one that takes into account the Rasch scale score (Bond and Fox, 2007) obtained at the end of upper secondary school as a dependent variable and a set of explanatory variables expression of student characteristics (entry score, origin of the student, etc.) or characteristics of the high school attended (type of high school, private or public school) or, finally, of the municipality of the high school (unemployment rate, municipality size, etc.) (ISTAT, 2011).

It should be noted that in these models only students who arrive in 13th grade and without any missing values on all variables are taken into consideration, in particular, the score at the 8th grade INVALSI tests in 2014. This means that the target population is the students reaching 13th grade without repeating.

In particular, the model 1 involves only two explanatory variables at the student level: the score at the grade INVALSI tests and the socio-economic

with the constant *A*, coefficients B_j and predictors X_j. The coefficients can be estimated via maximum likelihood estimation; the aim is to find the best linear combination of predictors that maximizes the likelihood of obtaining the frequencies of the observed results.

[7] The clustering algorithms can be divided into two group according on the technique the clusters are generated: agglomerative clustering algorithms that begin by placing each object in its own cluster and then group iteratively until a specific condition is reached and divisive clustering algorithms that begin by inserting all the objects in the collection into a single cluster and then iteratively separate it into smaller clusters until a specific condition is reached. In both cases, the result is a set of clusters containing one or more objects. Regardless of the approach used, the clustering algorithms are all based on a purely geometric metric, which allows you to identify how similar two objects are to each other.

ISBN 9788835123330

background, ESCS index (Campodifiori *et al.*, 2010). In the following model other variables are included, some of them related to students and others to the school type; in the third model, the variable entered relates only to the municipality size to which the school attended belongs; finally, in the fourth model, the variables connected to the socio-economic aspects of the municipalities themselves are included.

We conducted two different analyses, one for each INVALSI test subject (Italian literacy and Mathematics) and results are shown below.

In Table 1 (Italian literacy test) the R^2 adjusted increases from the Model 1 to the model 4. If we consider the first two explanatory variables, the 8th grade score and the family background we have R^2 value equal to 0,34, it means that the model explain the 34% of the variability of the dependent variable.

Tab. 1 – Results of the OLS regression model applied to Italian Literacy test data for students in grade 13

Variables	*Model 1*		*Model 2*		*Model 3*		*Model 4*	
	Beta	*Sig.*	*Beta*	*Sig.*	*Beta*	*Sig.*	*Beta*	*Sig.*
Grade 8 Mathematics entry score	0.53	0.00	0.45	0.00	0.45	0.00	0.41	0.00
Grade 13 student ESCS index	0.16	0.00	0.09	0.00	0.09	0.00	0.07	0.00
Foreign student			0.01	0.00	0.01	0.00	-0.02	0.00
Repeating student at lower secondary school			-0.04	0.00	-0.04	0.00	-0.04	0.00
High school type: Lyceum			0.18	0.00	0.18	0.00	0.20	0.00
High school type: Vocational Institute			-0.13	0.00	-0.13	0.00	-0.13	0.00
Attending a private school at grade 8			0.05	0.00	0.05	0.00	0.03	0.00
Attending a private school at grade 13			-0.06	0.00	-0.06	0.00	-0.06	0.00
Change of the municipality from grade 8 school to grade 13 school			0.02	0.00	0.03	0.00	-0.01	0.00
Municipality with more than 100,000 inhabitants*					0.04	0.00	0.03	0.00
Youth unemployment rate*							-0.14	0.00
Number of museums or similar institutions per square km of the municipality*							0.04	0.00
Number of active No-Profit local units per 1,000 inhabitants of the municipality*							0.06	0.00
Percentage of single-parent families*							0.01	0.00
Percentage of employees in the industry*							0.05	0.00
Adjusted R-squared	0.34		0.40		0.40		0.43	

Note: * 2011 Population census data (ISTAT).

ISBN 9788835123330

In the next two models the variability explained by the model increases to 40%. We observed a small impact of the variable added from the second to the third model; in the fourth model, the addition of a set of variables related to the social and economic aspects of the municipality increase the explained variability to 43%.

Regarding the Mathematics' model (Table 2), we did not find a large difference. The most important differences are in of the first model, which is higher for Mathematics than Italian literacy and in the coefficients of the first two explanatory variables: the 8th grade score than the family background.

Tab. 2 – Results of the OLS regression model applied to Mathematics test data for students in grade 13

Variables	*Model 1*		*Model 2*		*Model 3*		*Model 4*	
	Beta	*Sig.*	*Beta*	*Sig.*	*Beta*	*Sig.*	*Beta*	*Sig.*
Grade 8 Mathematics entry score	0.57	0.00	0.53	0.00	0.53	0.00	0.49	0.00
Grade 13 student ESCS index	0.12	0.00	0.08	0.00	0.08	0.00	0.06	0.00
Foreign student			0.02	0.00	0.02	0.00	0.00	0.03
Repeating student at lower secondary school			-0.03	0.00	-0.03	0.00	-0.03	0.00
High school type: Lyceum			0.03	0.00	0.03	0.00	0.04	0.00
High school type: Vocational Institute			-0.17	0.00	-0.17	0.00	-0.17	0.00
Attending a private school at grade 8			0.04	0.00	0.04	0.00	0.02	0.00
Attending a private school at grade 13			-0.04	0.00	-0.05	0.00	-0.05	0.00
Change of the municipality from grade 8 school to grade 13 school			0.02	0.00	0.03	0.00	-0.01	0.00
Municipality with more than 100,000 inhabitants*					0.02	0.00	0.03	0.00
Youth unemployment rate*							-0.12	0.00
Number of museums or similar institutions per square km of the municipality*							0.04	0.00
Number of active No-Profit local units per 1,000 inhabitants of the municipality*							0.05	0.00
Percentage of single-parent families*							0.00	0.13
Percentage of employees in the industry*							0.06	0.00
Adjusted R-squared	0.36		0.40		0.40		0.43	

Note: * 2011 Population census data (ISTAT).

ISBN 9788835123330

Tab. 3 – Results of the logistic regression model

Variables	*Model 1*			*Model 2*			*Model 3*			*Model 4*		
	B	*Sign.*	*Exp(B)*	*B*	*Sign.*	*Exp(B)*	*B*	*Sign.*	*Exp(B)*	*B*	*Sign.*	*Exp(B)*
Qualification of the student's mother: middle school	-0.48	0.00	0.62	-0.44	0.00	0.65	-0.44	0.00	0.64	-0.47	0.00	0.62
Qualification of the student's mother: university degree	0.27	0.00	1.31	0.29	0.00	1.33	0.30	0.00	1.35	0.28	0.00	1.33
Employment of the student's mother: household. unemployed	-0.08	0.00	0.93	-0.02	0.02	0.98	-0.02	0.02	0.98	-0.15	0.00	0.86
Employment of the student's mother: professional. manager	-0.06	0.00	0.94	-0.09	0.00	0.92	-0.08	0.00	0.92	-0.09	0.00	0.91
Employment of the student's mother: entrepreneur. self-employed	-0.05	0.00	0.95	-0.04	0.01	0.96	-0.05	0.00	0.95	-0.10	0.00	0.91
Qualification of the student's father: middle school	-0.39	0.00	0.68	-0.42	0.00	0.65	-0.43	0.00	0.65	-0.44	0.00	0.64
Qualification of the student's father: university degree	0.16	0.00	1.17	0.19	0.00	1.21	0.21	0.00	1.23	0.20	0.00	1.23
Employment of the student's father: household. unemployed	-0.46	0.00	0.63	-0.36	0.00	0.70	-0.35	0.00	0.70	-0.44	0.00	0.65
Employment of the student's father: professional. manager	0.17	0.00	1.19	0.05	0.00	1.06	0.06	0.00	1.06	0.07	0.00	1.08
Employment of the student's father: entrepreneur. self-employed	0.03	0.00	1.03	-0.02	0.06	0.98	-0.02	0.02	0.98	-0.03	0.01	0.97
Grade 8th low perforiming student in Italian and Mathematics	-1.17	0.00	0.31	-1.03	0.00	0.36	-1.03	0.00	0.36	-1.09	0.00	0.34
Repeating student at lower secondary school				-1.63	0.00	0.19	-1.64	0.00	0.19	-1.62	0.00	0.20
Foreign student				-0.86	0.00	0.42	-0.86	0.00	0.42	-0.73	0.00	0.48

(to be continued)

ISBN 9788835123330

Tab. 3 – Results of the logistic regression model

Variables	*Model 1*			*Model 2*			*Model 3*			*Model 4*		
	B	*Sign.*	*Exp(B)*	*B*	*Sign.*	*Exp(B)*	*B*	*Sign.*	*Exp(B)*	*B*	*Sign.*	*Exp(B)*
Attending a private school at grade 8^{th}				-0.16	0.00	0.85	-0.10	0.00	0.91	0.00	0.98	1.00
Medium sized municipality							-0.03	0.01	0.97	-0.09	0.00	0.91
Medium-big sized municipality							-0.14	0.00	0.87	-0.19	0.00	0.82
Big sized municipality							-0.34	0.00	0.71	-0.38	0.00	0.68
Metropolitan municipality							0.06	0.00	1.06	0.09	0.00	1.10
Percentage of commuters for work reasons										-0.03	0.00	0.97
Percentage of couples with children										0.00	0.00	1.00
Number of active no-profit local units per 1.000 inhabitants of the municipality										0.01	0.01	1.01
Number of museums or similar institutions per 1.000 inhabitants of the municipality										0.26	0.00	1.30
Number of hotels and accommodation facilities per 1.000 inhabitants of the municipality										0.00	0.82	1.00
Percentage of employees in the industry										0.18	0.00	1.19
Constant	1.37	0.00	3.92	1.55	0.00	4.70	1.59	0.00	4.90	2.36	0.00	10.57

ISBN 9788835123330

Generally, the Model 4 suggests that the most important positive factors for both the subjects in grade 13 students' score are their social, economic and cultural background, the 8th grade score, the high school type (Lyceum vs Vocational Institute), while the youth unemployment rate of the municipality has a negative impact; also other territorial characteristics such as the municipality size, the number of the museums or no-profit associations, the percentage of the industry sector employees are significant but have a smaller positive impact.

The next logistic regression (Agresti, 2018) shows the factors related to the territorial level on learning in order to evaluate the impact on the achievements of the 13th grade not repeating students. The dependent variable used is being or not in 13th grade in 2019 and the explanatory variables concerns the individual characteristics of the students, such as elements relating to the student's family and others relating to the municipal area of belonging.

The strong negative impact of being a repeating students at lower secondary school is very clear: the estimated odds is 0.2 times that of students not repeating. Being a foreigner also has a negative impact: for every native student who reaches the last year of high school, only 0.48 foreign students are able to do that; the other fundamental variable is having had low scores in Italian literacy and Mathematics at INVALSI test at the end of the low secondary school, these scores prove to be particularly predictive of the student's subsequent academic success. Students with a top quartile score in both subjects in 8th grade tests in 2014 are less likely to reach grade 13 on schedule.

The variables at the municipal level which appear most relevant in a positive way are the size of the municipality, the percentage of employees in the industry sector and the greater presence of museums.

All the other variables have a lower incidence, but are statistically significant, with the exception of the greater presence of hotels.

Before proceeding to the cluster analysis we observe on maps the distribution of the above mentioned phenomena, considering the municipality classified according to the number of their students. The images below are to be compared to the above mentioned regional maps (Figure 1, 2 and 3). Below there are the distribution of the low-performing students on the whole territory. This distribution considers only the student's cohort that has performed the 8th grade test in 2014 and the 13th grade test in 2019. For this reason, Figures 4 to 11 show only the municipalities with students belonging to the cohort defined before (1,060 municipalities). The first graphs of Italian (Figure 4) and Mathematics (Figure 8), shows the distribution of all the municipalities. The next three graphs (Figures 5, 6 e 7 for Italian, Figures 9, 10 e

ISBN 9788835123330

11 for Mathematics) divide the municipalities into three categories according to the amount of the residing students. The three categories are:
1) up to 100 students;
2) between 101 and 250 students;
3) more then 250 students.

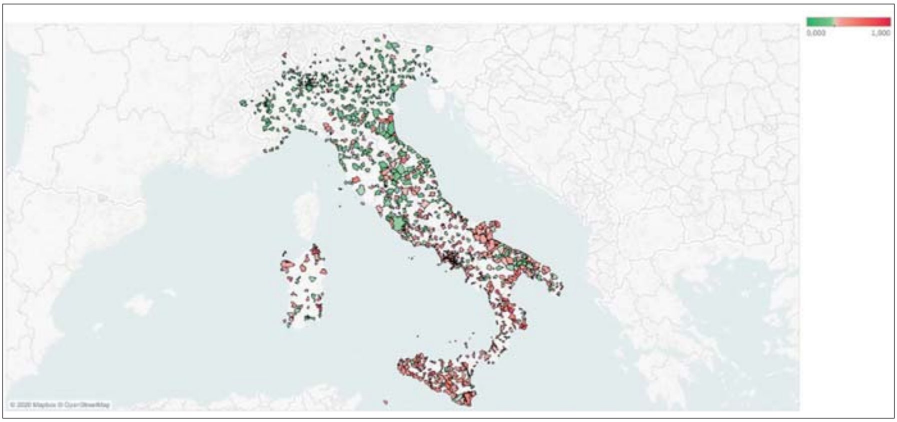


Fig. 4 – Distribution of students in difficulty in Italian literacy the last year of upper secondary school for all municipalities. The center value of the green (less critical) -red (more critical) diverging colour palette is 33%

As we can see from the graph above, the distribution is not homogeneous across the regions if we consider categories of students.

This means that even within areas with a low "level" of inequality, there are some municipalities that instead show strong differences compared to their surrounding context.

Both in Italian and Mathematics, for example, there are groups of neighbouring municipalities between Puglia and Basilicata which have relatively small percentages of low-performing students. In Puglia they are mainly concentrated in the provinces of Bari, Taranto, Barletta-Andria-Trani and Lecce, in Basilicata they are concentrated in the province of Matera. This group mainly concerns municipalities with a student population which do not exceed 250 students (Figures 5-6, 9-10), including some provincial capitals. In particular, in Italian literacy the municipalities of Matera, Trani and Barletta show low percentages of low-performing students (Figure 7). Similar groups are not observed in other southern regions. All provincial capitals have more than a third of students which are low performing in both subjects. The exceptions are Cagliari, Nuoro and Ragusa in Italian literacy and only Ragusa in Mathematics.

ISBN 9788835123330

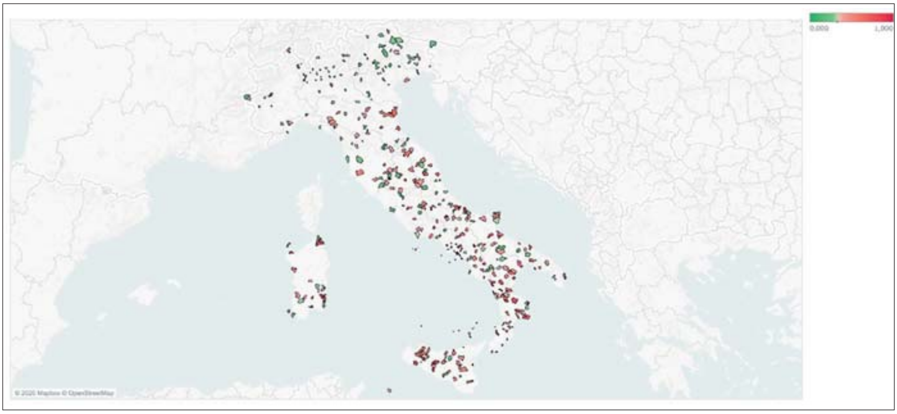


Fig. 5 – Distribution of students in difficulty in Italian literacy in the last year of upper secondary school for all municipalities with 1 to 100 students. The center value of the green (less critical) -red (more critical) diverging colour palette is 33%

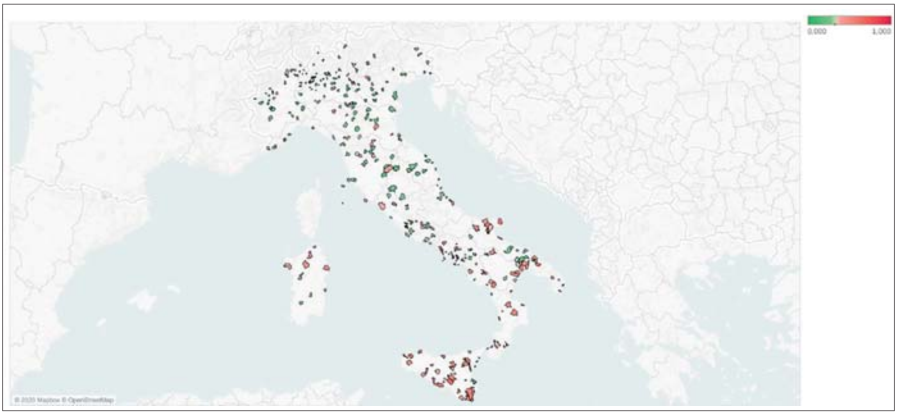


Fig. 6 – Distribution of students in difficulty in Italian literacy in the last year of upper secondary school for all municipalities with 101 to 250 student. The center value of the green (less critical) -red (more critical) diverging colour palette is 33%

ISBN 9788835123330

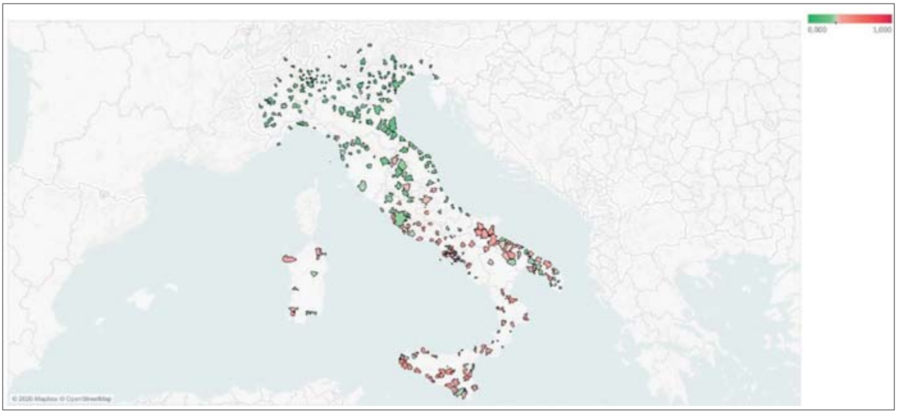


Fig. 7 – Distribution of students in difficulty in Italian literacy in the last year of upper secondary school for all municipalities with more than 250 students. The center value of the green (less critical) -red (more critical) diverging colour palette is 33%

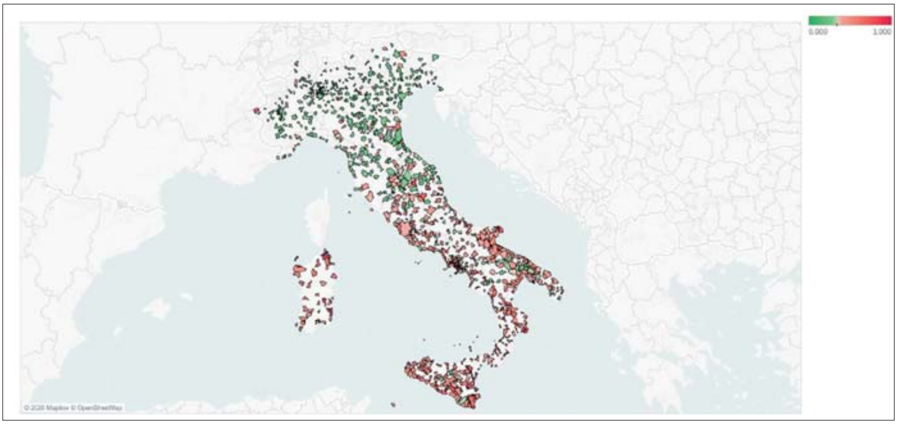


Fig. 8 – Distribution of students in difficulty in Mathematics in the last year of upper secondary school for all municipalities. The center value of the green (less critical) -red (more critical) diverging colour palette is 33%

ISBN 9788835123330

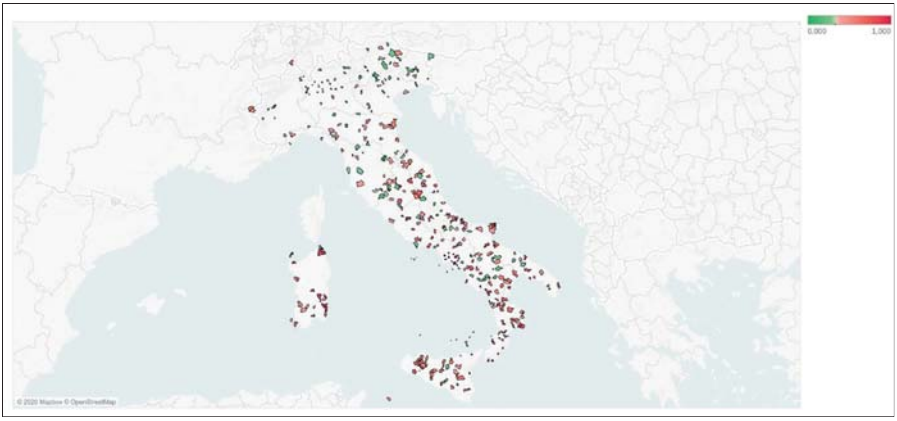


Fig. 9 – Distribution of students in difficulty in Mathematics in the last year of upper secondary school for all municipalities with 1 to 100 students. The center value of the green (less critical) -red (more critical) diverging colour palette is 33%

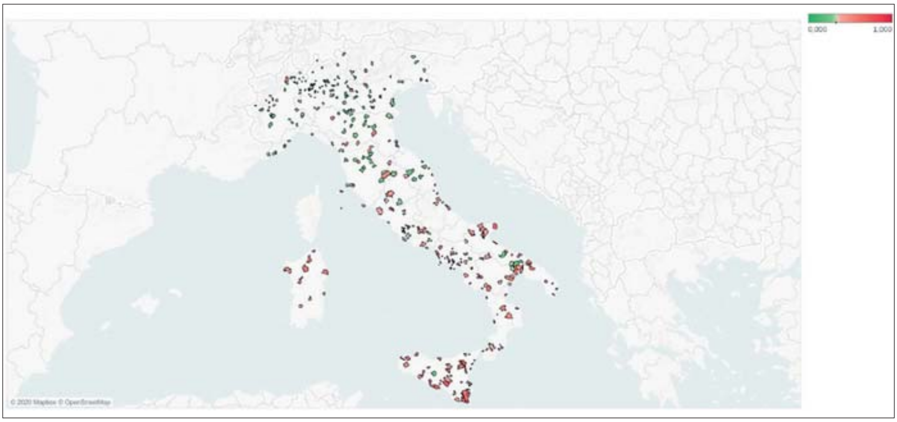


Fig. 10 – Distribution of students in difficulty in Mathematics in the last year of upper secondary school for all municipalities with 101 to 250 students. The center value of the green (less critical) -red (more critical) diverging colour palette is 33%

ISBN 9788835123330

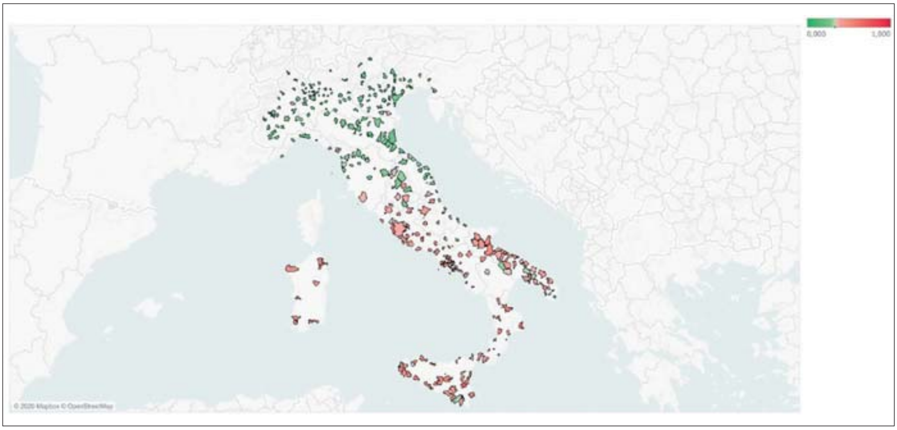


Fig. 11 – Distribution of students in difficulty in Mathematics in the last year of upper secondary school for all municipalities with more than 250 students. The center value of the green (less critical) -red (more critical) diverging colour palette is 33%

Another aspect observed at the regional level (Figure 3) is the so-called implicit drop-out of students in the last year of secondary school (SY. 2018-19). In this case, all outgoing students are considered, including those who have repeated several years in secondary school. For this reason figures 12 to 15 shows only the municipalities with at least one upper secondary school (1,437 municipalities).

In the first graph (Figure 12) all the municipalities are mapped, in the others (Figures 13, 14 and 15) the classification of the municipalities indicated above is used.

As can be seen from the figures below (Figures 12, 13, 14 and 15) the differences between macro-areas are quite clear, with the south having the highest percentages of implicit drop-out regardless of the municipal size taken into consideration.

In particular, in Puglia in the provinces of Taranto and Foggia, in Calabria in the provinces of Cosenza and Crotone and in Sicily in the province of Agrigento. Even in municipalities with a larger student population, the share of students who do not reach the minimum proficiency levels is greater than a third.

Most of the municipalities in central and northern Italy have percentages of less than 10%. Exceptions are some groups of municipalities between Tuscany and Liguria, some of the largest municipalities in the province of Rome (Figure 15), groups of municipalities between the province of Frosinone and that of Latina and some small or medium-sized municipalities in Umbria, mainly concentrated in the province of Perugia (Figures 13 and 14).

ISBN 9788835123330

As observed in the maps of the distribution of students in difficulty, even in the case of implicit drop-out it is possible to identify groups of municipalities in the south with values in contrast with the regional data. This is the case of the province of Lecce, Bari and Barletta-Andria-Trani in Puglia, the province of Palermo and Ragusa in Sicily and some small neighbouring municipalities in the provinces of Salerno and Avellino (Figures 13 and 14).

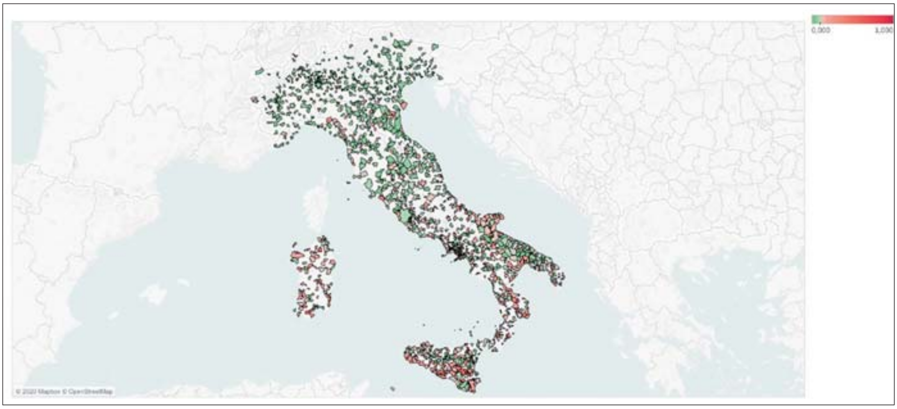


Fig. 12 – Implicit drop-out distribution in the last year of upper secondary school for all municipalities. The center value of the green (less critical) -red (more critical) diverging colour palette is 10%

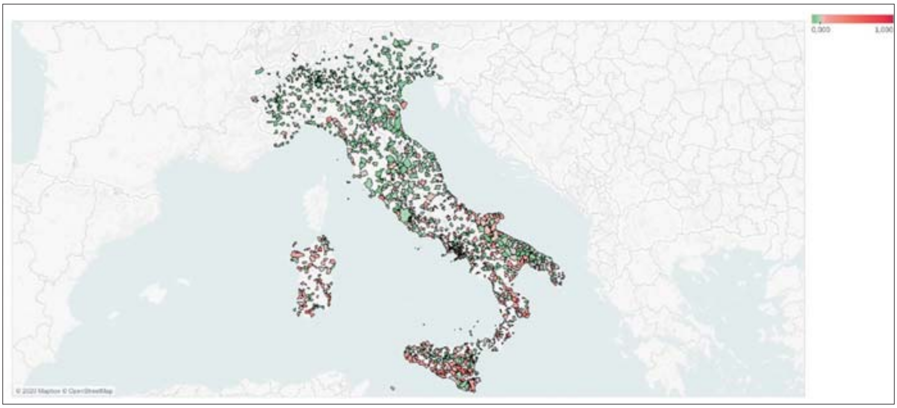


Fig. 13 – Implicit drop-out distribution in the last year of upper secondary school for all municipalities with 1 to 100 students. The center value of the green (less critical) -red (more critical) diverging colour palette is 10%

ISBN 9788835123330

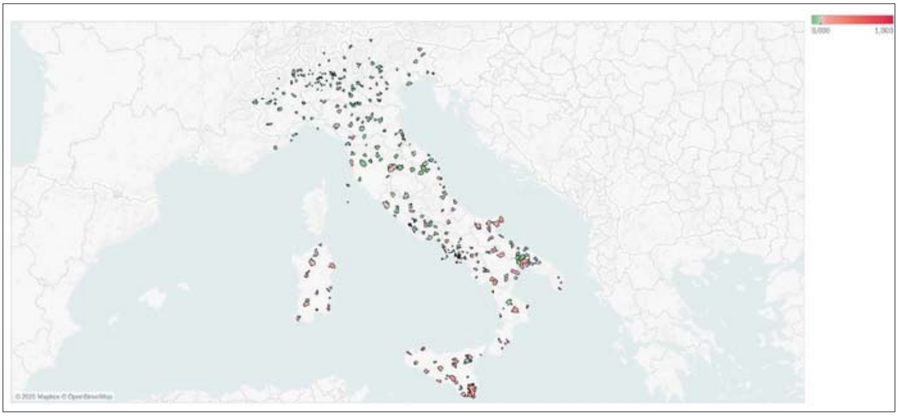


Fig. 14 – Implicit drop-out distribution in the last year of upper secondary school for all municipalities with 101 to 250 students. The center value of the green (less critical) -red (more critical) diverging colour palette is 10%

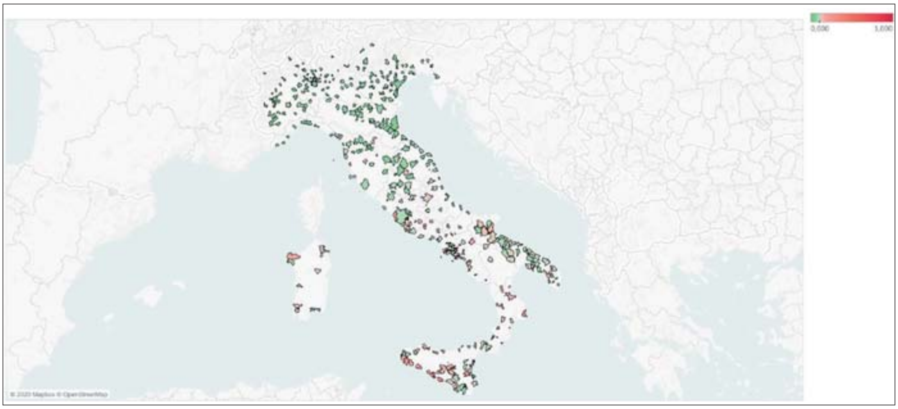


Fig. 15 – Implicit drop-out distribution in the last year of upper secondary school for all municipalities with more than 250 students. The center value of the green (less critical) -red (more critical) diverging colour palette is 10%

The last analysis performed, the cluster analysis, attempted to group the municipalities by similar characteristics and highlight any associations with the phenomena of interest studied so far, that is, percentage of students who do not reach the last year of upper secondary school without repeating a school year, percentage of students in difficulty in Italian literacy or in Mathematics.

In particular, two distinct cluster analysis were carried out using the municipalities as units; for the first clustering the following variables were used: number of hotels, museums per inhabitant, employees in the industrial sector,

ISBN 9788835123330

percentage of commuters, youth unemployment rate and percentage of students in difficulty in Italian (Figures 16-19); for the second clustering were used: number of hotels, museums per inhabitant, employees in the industrial sector, percentage of commuters, youth unemployment rate and percentage of students in difficulty in Mathematics (Figures 20-23).

Below we show some exemplificative images of how the regional realities in some cases are "bypassed" and mixed environments are created.

As highlighted by the cartograms below (Figures 16-17), the municipalities of Cluster 1 are mainly located in the south (in more than 90% of cases), the municipalities of Cluster 2 and Cluster 3, on the other hand, are mainly located in the Center-North.

The largest municipalities belong to Cluster 3, including all the metropolitan cities of the Center-North.

The municipalities of Cluster 2 are on average smaller in size, with an average population per municipality of about 20 thousand inhabitants.

The municipalities of Lazio mainly belong to Cluster 1 and Cluster 3 (the municipalities of the province of Latina and Frosinone). Some municipalities such as Cisterna di Latina and Aprilia are exceptions, characterized by many residents employed in the industry sector, which belong to Cluster 2.

In the south, the only provincial capital that does not belong to Cluster 1 is Ragusa. As previously mentioned, membership in Cluster 3 in this case is partially justified by relatively low percentages of students in difficulty in the two subjects.

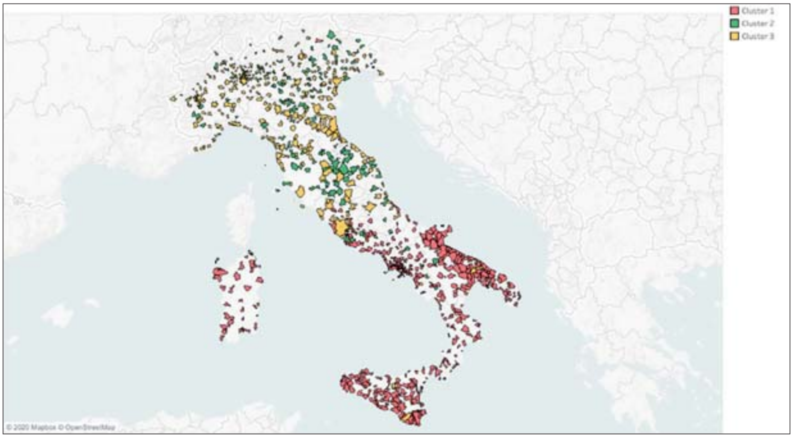


Fig. 16 – Percentage of students in difficulty in Italian literacy in the last year of upper secondary school divided into clusters represented throughout the national territory

ISBN 9788835123330

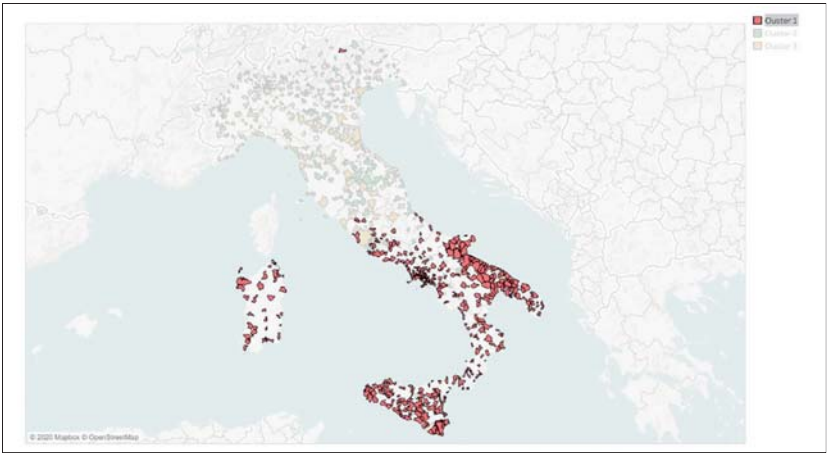


Fig. 17 – Percentage of students in difficulty in Italian literacy in the last year of upper secondary school belonging to cluster 1

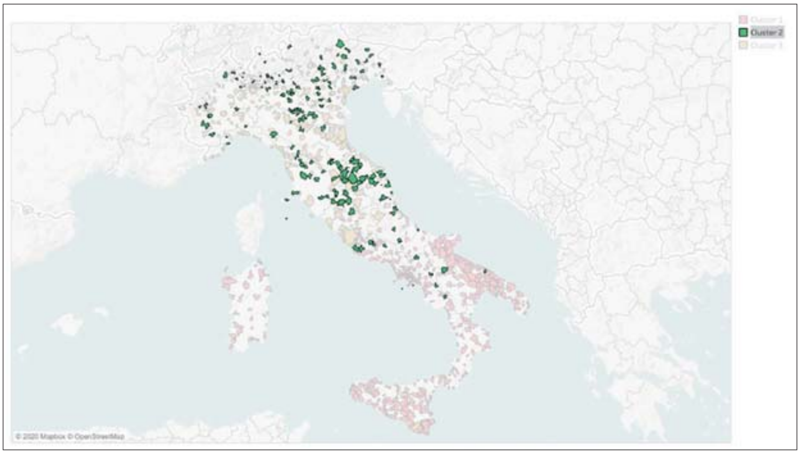


Fig. 18 – Percentage of students in difficulty in Italian literacy in the last year of upper secondary school belonging to cluster 2

ISBN 9788835123330

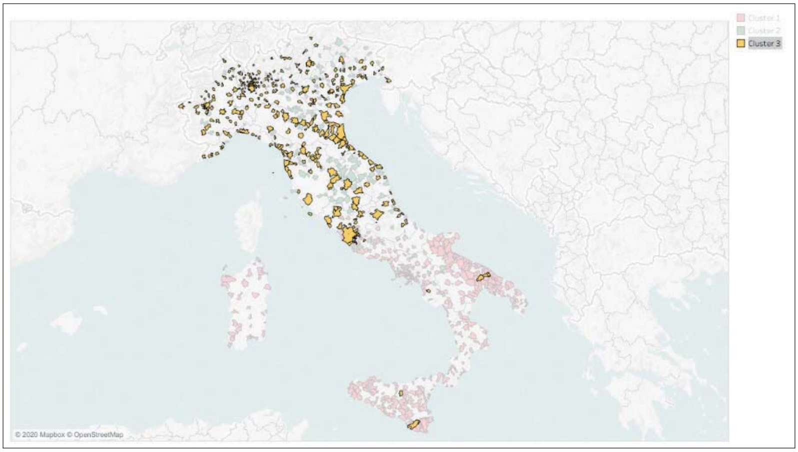


Fig. 19 – Percentage of students in difficulty in Italian literacy in the last year of upper secondary school belonging to cluster 3

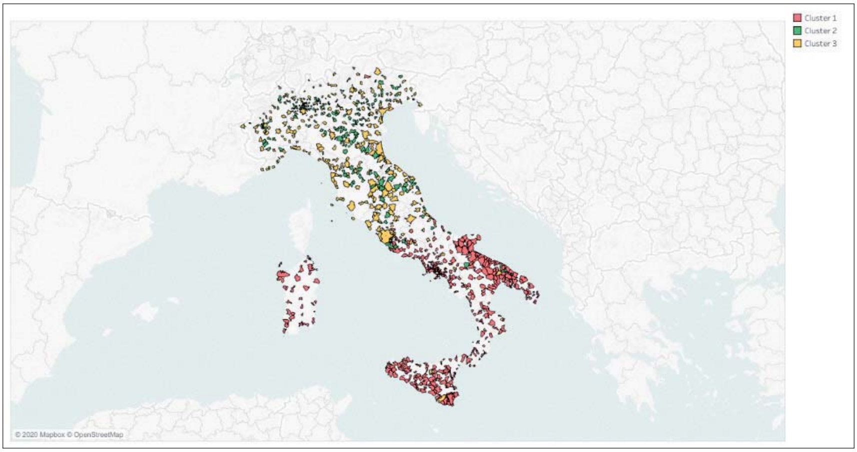


Fig. 20 – Percentage of students in difficulty in Mathematics in the last year of upper secondary school divided into clusters represented throughout the national territory

ISBN 9788835123330

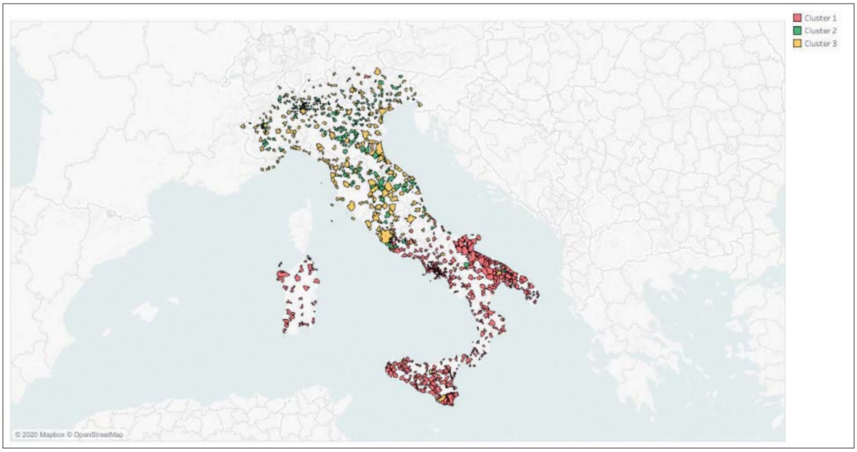


Fig. 21 – Percentage of students in difficulty in Mathematics in the last year of upper secondary school belonging to cluster 1

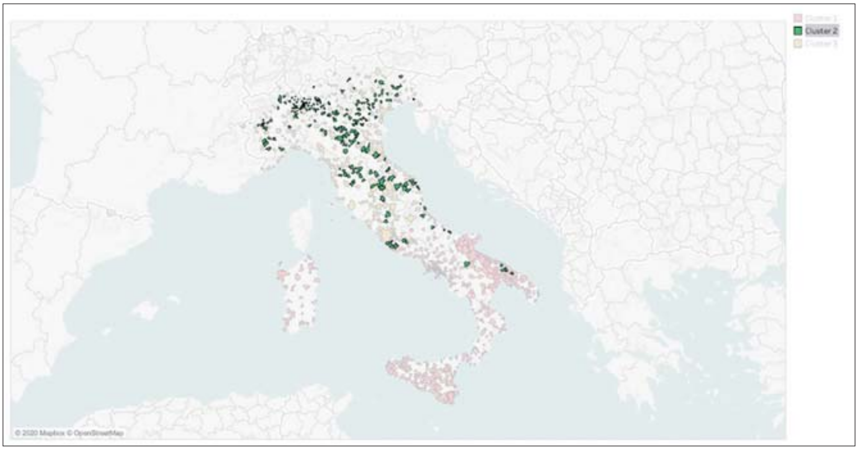


Fig. 22 – Percentage of students in difficulty in Mathematics in the last year of upper secondary school belonging to cluster 2

ISBN 9788835123330

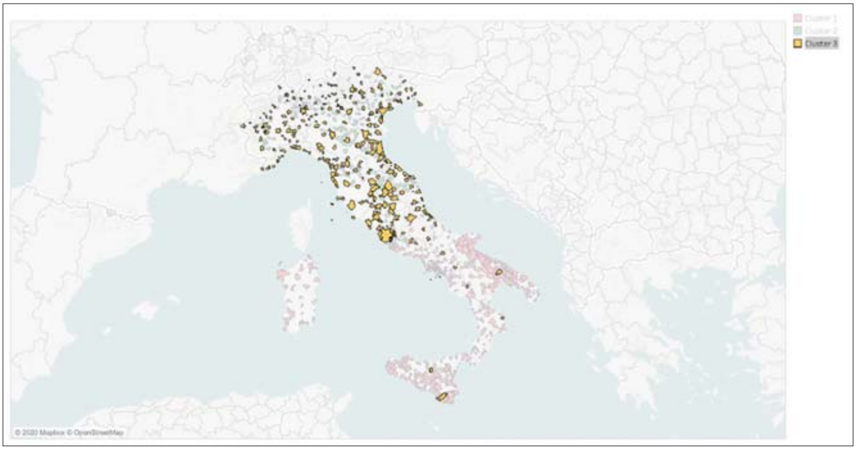


Fig. 23 – Percentage of students in difficulty in Mathematics in the last year of upper secondary school belonging to cluster 3

Table 4 shows the averages of the variables within each group. In this way it is possible to understand the characteristics of the units within each group.

Tab. 4 – Averages of the variables within each cluster

Variables	*Cluster*		
	1	*2*	*3*
Number of hotels and accommodation facilities per 1,000 inhabitants of the municipality	1.12	0.77	6.04
Number of museums or similar institutions per 1,000 inhabitants of the municipality	0.06	0.06	0.16
Percentage of employees in the industry	11.61	31.07	12.63
Percentage of commuters for work reasons	25.23	36.62	34.43
Percentage of low-performing students (Italian)	42.36	19.33	21.11
Youth unemployment rate	47.40	25.01	26.94

Cluster 1 includes municipalities with a low number of hotels, a low number of museums per inhabitant, a low percentage of employees in the industrial sector, a low percentage of commuters, high rates of youth unemployment and a high percentage of students in difficulty in both subjects.

Cluster 2 includes municipalities with a low number of hotels, a low number of museums per inhabitant, a high percentage of employees in the industrial sector, a high percentage of commuters, low youth unemployment rates and a low percentage of students in difficulty in both subjects.

ISBN 9788835123330

Finally, Cluster 3 includes municipalities with a high number of hotels, a high number of museums per inhabitant, a low percentage of employees in the industrial sector, a high percentage of commuters, low rates of youth unemployment and a low percentage of students in difficulty in both subjects.

Cluster 1 is further away than Cluster 2 and Cluster 3 (Table 4). The greatest distance is observed between Cluster 1 and Cluster 2 (Euclidean distance 39.24), while Cluster 2 and Cluster 3 are much closer (Euclidean distance 19.48).

7. Conclusions

The complete regression model (Model 4) achieves a good level of variability explained and highlights a significant impact of variables related to the municipal territorial reality on the students' final score achieved: the youth unemployment rate first, but also other municipal characteristics, although with a weaker impact such as the number of inhabitants of the municipality, the more or less accentuated presence of museum institutes and/or local units of no-profit organizations and the percentage of industrial workers. In summary, the regression confirms the presence of a significant link between the student's performance in Italian and Mathematics and the socio-economic characteristics of the municipality of residence of the high school attended.

The logistic model has highlighted that, in addition to individual factors, also the variables at the territorial context have a significant impact on the probability of reaching the last year of upper secondary school within the foreseen times. In particular, the variables on the employment context and the demographic ones seem to be the most relevant.

The analyses conducted show profound inequalities at territorial level. The following cluster analysis allow us to better identify the differences between the municipalities.

In particular, the Cluster Analysis made it possible to "isolate" three geographic environments characterized by common aspects. Although the municipalities with the most difficult schools are almost exclusively concentrated in the south, some territorial realities emerge that seem to go against the trend, mainly in Puglia.

In central Italy, on the other hand, the municipalities of 'lower' Lazio are in contrast with the rest of the municipalities in the macro-area. More than half of these municipalities, in fact, belong to the Southern School Cluster.

In Italy, in order to satisfy the Goal 10 of the Agenda 2030, with the aim of monitoring the inequalities, it is necessary to reduce to a minimum territo-

ISBN 9788835123330

rial breakdown. It is also one of the principles of Agenda 2030: "no leaving one behind". A municipal-level analysis in the educational outcome is capable of facing this challenge.

References

Anderson T.W. (2003), *An Introduction to Multivariate Statistical Analysis*, John Wiley, New York.

Agasisti T., Falzetti P., Freddano M. (2015), "L'uso dei risultati delle Rilevazioni Nazionali per l'autovalutazione delle scuole", *Rassegna Italiana di Valutazione*, XIX, 61, pp. 28-48.

Agresti A. (2018), *Statistical Methods for the Social Sciences*, Pearson, London.

Bond Trevor G., Fox Christine M. (2007), *Applying the Rasch Model: Fundamental Measurement in the Human Sciences*, Lawrence Erlbaum Associates, Mahwah.

Campodifiori E., Figura E., Papini M., Ricci R. (2010), "Un indicatore di status socio-economico-culturale degli allievi della quinta primaria in Italia", *INVALSI Working Paper*, 2.

INVALSI (2019), *Rapporto prove INVALSI 2019*, retrieved on Dicember, 30, 2020, from https://invalsi-areaprove.cineca.it/docs/2019/Rapporto_prove_INVALSI_2019.pdf.

ISTAT (2011), *15° Censimento generale della popolazione e delle abitazioni*, retrieved on Dicember, 30, 2020, from https://www.istat.it/it/files/2012/12/volume_popolazione-legale_XV_censimento_popolazione.pdf.

Longobardi S., Falzetti P., Pagliuca M.M. (2018), "Quis custodiet ipsos custodes? How to detect and correct teacher cheating in Italian student data", *Statistical Methods and Applications*, 27 (3), pp. 515-543.

Loth A. (2019), *Visual Analytics with Tableau*, John Wiley, New York.

Mccullagh P., Nelder A.J. (1989), *Generalized Linear Models*, Chapman & Hall, London.

Ricci R. (2019a), *Dalla licenza media alla maturità. Il percorso visto attraverso i dati INVALSI*, retrieved on Dicember, 30, 2020, from https://www.invalsiopen.it/.

Ricci R. (2019b), *La dispersione scolastica implicita*, retrieved on Dicember, 30, 2020, from https://www.invalsiopen.it/.

ISBN 9788835123330

2. Doing school but not at school: territorial peculiarities and socio-economic inequalities in access to distance learning

by Cecilia Bagnarol, Silvia Donno, Veronica Riccardi

Covid-19 pandemic has strongly highlighted how social and economic differences among and within countries can contribute to increase many inequalities in health care and also in many other aspects of people's lives, including education.

This study aims to investigate, within the Italian national territory, how the transfer of all educational activities in the digital classroom affected home as the only physical learning environment. In fact, domestic environment, characterized by possession of adequate technological spaces and tools, can be discriminatory for certain sections of the population.

Census data from the INVALSI 5th grade Student Questionnaire for the academic year 2018/2019 are used. Through georeferencing of schools, we brought to light some territorial peculiarities, not always predictable, based on socio-economic and cultural background of students. To deepen the size and dimension of these differences can be useful for setting the prospective strategies of action.

1. Home as a learning environment

Attention to the learning space has prompted reflections, pedagogical and beyond, for some decades already. If we consider that the founding fathers of pedagogical activism, putting the student in the foreground, clearly identified a key role of the context in the teaching and learning processes, we can say that school environment is one of the undisputed protagonists of research, experimentation and clashes of ideas by philosophers, psychologists, pedagogists and architects for more than a century.

ISBN 9788835123330

To define the learning environment, it is useful to resort to the constructivist paradigm that defines it as the place “where students can work together and help each other to learn how to use a multiplicity of tools and information resources in the common pursuit of learning objectives and problem-solving activities” (Wilson, 1996, p. 5). The space is therefore considered as a “third educator” (Malaguzzi, 2010) and plays a decisive role in determining the quality of learning: classrooms, laboratories, corridors, architectural form of the building, colors of the walls, furnishings, teaching materials. All this, and much more, creates the context in which students live, learn, experience, and enter into relationship with others (Močinić and Moscarda, 2016).

Thinking about the Italian context, to which this study is dedicated, it is important to note that in the National Guidelines for the curriculum of the pre-primary school and the first cycle of school education (MIUR, 2012), it is argued that “a good primary and secondary school is a suitable context to promote meaningful learning and to guarantee the educational success of all students”. In the paragraph concerning the learning environment, a specific space was dedicated to the implementation of adequate interventions regarding diversity (in particular teaching of Italian Language to students with non-Italian citizenship and integration of students with disabilities), to ensure that diversity does not become inequality. We will avoid entering in specific features that characterize the National Guidelines on this specific topic, for each segment of education, but it is worth underlining and highlight how the learning space, if properly thought out and structured, can be a tool for inclusion and educational success for all students, without exception. Class and school, therefore, are perceived as a privileged space (to think, plan, structure) where specific responses to the educational needs of the students are found. So, it is important searching strategies to hinder inequalities – whether they are related to the social and family context where the students come from or to specific situations of disability – and to value differences. Teachers and school managers paid great attention to these aspects and strive daily to improve learning spaces even more, also through specific projects, studies, didactic experiments and the use of new technologies.

March 5, 2020: in Italy, the school (intended as face-to-face classroom teaching) shut. Covid-19 has forced political decision-makers to interrupt all activities in the country, including schools, to face the spread of the virus. For a few weeks, homes became the only living environment and, at school level, the only way to “go to school”, but “not at school” (note MIUR n. 388, of 17 March 2020), keeping the school community alive by totally shifting the learning environments from the physical class to the virtual class. The rules of the game have totally changed: for this period, teachers could not

ISBN 9788835123330

benefit from the teaching resources present at school, but they had to rely, almost totally, on the computer as a mediator of their work and student activities. In this context, the danger was that the differences, especially those related to the social and family environment of the students, could become more and more discriminating and that the poorest, in terms of the material and cultural resources of the family, could be left behind.

The study of the influence of socio-cultural factors on school learning is a rather classic topic in pedagogical, sociological and economic literature. There is a substantial agreement that the family environment is one of the main factors involved in determining interpersonal differences in learning achievement: in fact, more educated parents can more effectively follow their children's school activities, also having available more economic resources to be allocated to them too (Cappellari, 2006; Checchi, 2010; Checchi, Fiorio and Leonardi, 2006; Parziale, 2016). However, academic degree of parents and their economic possibilities are not the only matter, but we consider how these factors translate into the living conditions that the student experiences in their lives. Holding a quiet place to study, a desk, a computer, and an Internet connection can be granted for some students and much less for others, especially in a pandemic period. This is precisely the object of this study: through territorial comparisons, it intends to examine the domestic environment as a "new learning environment", investigating the nature and extent of the distance existing between those who have, in their own home, all the resources suitable for promoting distance learning and those who, on the other hand, are excluded, partially or totally. We will move in two directions: on one hand we investigate domestic environment as a place where learning takes place (therefore the presence or absence of appropriate spaces) and, on the other, the possession of some digital devices (mainly the computer and the Internet connection) as essential tools for the distance learning. We decided to carry out territorial analyses to verify whether the territorial dualism (Centre-North vs South) characterizing Italy on many aspects related to the education system (student learning levels, school choices, etc.) is also detectable on our investigation topic.

2. Data and Methods

2.1. Data

The dataset realized for the analysis contains the answers provided by all the national student population of 5th grade of primary school to the Student

ISBN 9788835123330

Questionnaire arranged by INVALSI for the academic year 2018/2019. Due to the health emergency caused by Covid-19 pandemic, in fact, the INVALSI National Survey for the academic year 2019/2020 was suspended, so the data from the previous school year are used as a proxy for the data of non-recognition.

In particular, the dataset refers to students who have carried out at least one of the INVALSI tests of Italian Language and Mathematics and who have provided a valid answer to the Q4 question battery of Student Questionnaire, which investigates the possession of some tools at home: a total of 480,474 students were examined. This dataset has been linked to additional information: some of a socio-economic and cultural nature, related to individual students such as the profession and the qualifications of both parents; others of a territorial type such as the address, the municipality and the province of the schools that students have attended, necessary for the geo-referencing procedure of schools.

In order to study the material and technological equipment of the students, two synthetic indicators have been calculated: the indicator relating to the learning environment (hereinafter referred to as AA) and the indicator relating to the possession of digital devices (hereinafter referred to as DD).

To calculate the AA indicator, Q4A and Q4C questions were taken into consideration, asking the student respectively "At home you have: a quiet place to study" and "At home you have: a desk to do homework". The indicator takes value 1 when the student has at least one of the options, a quiet place to study or a desk to do homework, vice versa it takes value 0.

To calculate the DD indicator, Q4B and Q4E questions were taken into consideration, asking the student respectively "At home you have: a computer that you can use to study" and "At home you have: an Internet connection". The indicator takes value 1 when the student has at least one of the options, a computer for study and an Internet connection, vice versa it takes value 0.

In addition, for the purposes of the survey, some socio-economic and cultural variables have been recoded and included in the analyses. The first calculated variable, *StudioGenitori*, is an indicator that recodes the variable INVALSI "academic degree" of father and mother, and it is worth 1 when at least one parent of the student has a higher academic degree than the diploma, while it is worth 0 in other cases. The academic degree reached by parents is, in fact, the most used indicator to capture the cultural background of students (Triventi, 2014).

To calculate the second variable, *DigitalGenitori*, the variables INVALSI "father and mother profession" were considered, taking the categories "Manager, university professor, civil servant or military officer", "Entrepreneur/

ISBN 9788835123330

agricultural owner", "Professional employee or freelancer" and "Teacher, military employee etc" as professions potentially related to the use of digital devices, while the categories "Unemployed", "Homeowner", "Self-employed worker (trader, direct farmer, artisan, mechanic, etc.)", "Worker, services/cooperative member" and "Pensioner" as professions not bound by the use of digital devices.

The new indicator takes value 1 when at least one parent of the student has a profession that involves the use of digital devices while assuming value 0 in other cases. Although research shows that over the years, especially regarding the use of digital devices and the Internet, the differences between managers, entrepreneurs freelance professionals and workers are gradually becoming blurred (ISTAT, 2019), we decided to include an indicator of this type because it can be crucial to understand whether parents are familiar with technologies, not just for recreational or social use.

All the selected information were aggregated at school level, then it was applied a data-cleaning procedure for the removal of non-representative schools (schools with less than 20 students) and schools located on small islands. These schools, due to their distance from the mainland, could have a distortive effect in the calculation of the distance measure. Altogether the dataset has 6,026 schools distributed throughout the national territory.

The variables used in the study are:

– the percentage of students with a value of 1 for the AA indicator;
– the percentage of students with a value of 1 for the DD indicator;
– the percentage of students with a value of 1 for the StudioGenitori indicator;
– the percentage of students with a value of 1 for the DigitalGenitori indicator.

2.2. Notes on the Univariate and Bivariate Global and Local Spatial Autocorrelation Indices

Spatial autocorrelation (or spatial association) is a concept that derives from the consideration that the values assumed by a variable under study are not distributed independently over the territory but, on the contrary, tend to concentrate in certain areas (Demarinis *et al.*, 2011). In particular, we talk about:

– *positive spatial autocorrelation* when similar values of a variable tend to group close to each other;
– *negative spatial autocorrelation* when dissimilar values of a variable tend to group close to each other;

ISBN 9788835123330

– *absence of spatial autocorrelation* (or spatial independence) when the values of a variable are distributed randomly over the territory.

Among spatial autocorrelation techniques, we differentiate the methods for measuring on the entire set of localities (*global measures*) and the methods for measuring on a spatially delimited subset of localities (*local measures*).

The most frequently used measure to test the degree of *global spatial autocorrelation* for a dataset is the *Moran's I* statistic (Moran, 1948), expressed by the formula:

$$\mathrm{I} = \frac{N}{\sum_i \sum_j w_{ij}} \frac{\sum_i \sum_j w_{ij} \, (x_i - \bar{x})(x_j - \bar{x})}{\sum_i (x_i - \bar{x})^2}$$

where:

– N is the number of observations (i.e. localities or geographical units);
– x_i is the value of the observed variable in locality i;
– x_j is the value of the observed variable in the locality j;
– $\overline{x}$ is the sample mean;
– $(x_i - \overline{x})$ represents the deviation from the average of the observed variable;
– w_{ij} is a weight assigned to the relationship between locality i and locality j.

Moran's I statistic is structurally similar to the correlation coefficient and assumes values between -1 and +1.

The Moran's I can be extended to the multivariate case: in this way the index represents the systematic association between the values of a variable observed x in a given area of interest and the values of another variable y observed in neighbouring areas.

The *bivariate Moran index* of *x* with respect to y is thus obtained:

$$I_b = \frac{\sum_{i=1}^{N} \sum_{j=1}^{N} w_{ij} \, z_i v_j}{\left(\sum_{i=1}^{N} \sum_{j=1}^{N} w_{ij}\right)\left(\sum_{i=1}^{N} z_i^2\right) / N}$$

Local measures, statistics for measuring *spatial autocorrelation at the local level*, allow us to identify the contribution of each location on global pattern and therefore allow to study the variations of spatial autocorrelation within the territory. By focusing on each location, these techniques can be used to identify the presence of spatial clusters.

The most frequently used *Local Indicator of Spatial Association (LISA)* is represented by the local version of Moran's I index and it is expressed by the following formula:

ISBN 9788835123330

$$I_i = N \frac{x_i - \bar{x}}{\sum_i (x_i - \bar{x})^2} \sum_{i=1, i \neq j} w_{ij} (x_j - \bar{x})$$

Positive values of local index indicate the presence of clusters in which the observations have similar intensities (i.e. they all have high values or low values). On the contrary, negative values indicate the presence of clusters in which the observations have different intensities (i.e. the observations with high values are located close to observations with low intensity or vice versa). Finally, when combining the significance information provided by the LISA in a map, we obtain the *Moran Significance Map*, that reports areas with significant LISAs associated with the relative positive or negative local spatial autocorrelation value and the *Moran Cluster Map*, that show areas with significant LISAs and relative indication of spatial association type. Areas are distributed in four different group of observations: a statistically significant cluster of high- values (High-High), a cluster of low values (Low-Low), an outlier in which a high value is surrounded primarily by low values (High-Low), and an outlier in which a low value is surrounded primarily by high values (Low-High) (Anselin, 1995).

3. Results

3.1. Descriptive Statistics

About learning environment and digital devices, PISA data (collected before the health crisis) shown that, in Italy, 90% of 15-year-old students have a computer to do homework and 91% of them have a quiet place to study at home. These values, which are similar to the OECD media, have strong variations based on the social and economic background of families (OECD, 2020; Palmerio and Caponera, 2020).

A general overview of our data, involving students of 5th grade of primary school, shows that 8% of students (about 38,300 students) have neither a computer to study or an Internet connection, while 5.2% (about 25,097 students) don't have a quiet place to study or a desk to do homework at their homes. In addition, only 25.2% of students have at least one parent with a higher academic degree, less than 10% have both parents graduates while 53.5% don't have neither parent who work with technological devices.

The distribution of the variables at regional level offers interesting points of reflection: the highest percentages of the AA index equal to 0 (percentage greater or equal to 6%) are found in the regions of Northern Italy (Valle

ISBN 9788835123330

D'Aosta, Veneto and Friuli-Venezia Giulia), while the highest percentages of the DD index equal to 0 (higher or equal to 9%) are found in Friuli-Venezia Giulia, Puglia, Calabria, Sicily and Sardinia.

Looking at the distribution of the *StudioGenitori* indicator it is possible to notice that less than 25% of students in Campania, Puglia, Calabria, Basilicata, Sicily and Sardinia have at least one parent with higher education than the diploma. Moreover, in Campania and Sicily, less than 40% of scholars have at least one parent working with technological devices. The regional details of the two indicators are given in Table 1.

Tab. 1 – Percentage distribution of AA, DD, StudioGenitori e DigitalGenitori by region

	AA		*DD*		*StudioGenitori*		*DigitalGenitori*	
	0	*1*	*0*	*1*	*0*	*1*	*0*	*1*
Valle D'Aosta	6.0	94.0	8.8	91.2	71.1	28.9	44.5	55.5
Piedmont	4.7	95.3	8.0	92.0	74.0	26.0	50.4	49.6
Liguria	5.1	94.9	8.6	91.4	71.9	28.1	50.0	50.0
Lombardy	5.0	95.0	6.8	93.2	71.3	28.7	47.0	53.0
Veneto	6.0	94.0	8.9	91.1	73.4	26.6	48.2	51.8
Friuli-Venezia Giulia	6.4	93.6	9.2	90.8	70.0	30.0	46.1	53.9
Emilia-Romagna	5.1	94.9	7.9	92.1	69.0	31.0	48.4	51.6
Tuscany	5.4	94.6	8.0	92.0	74.9	25.1	56.6	43.4
Umbria	4.2	95.8	7.3	92.7	67.6	32.4	48.1	51.9
Marche	5.3	94.7	7.7	92.3	71.3	28.7	51.8	48.2
Lazio	5.1	94.9	7.0	93.0	73.4	26.6	53.1	46.9
Abruzzo	4.8	95.2	7.7	92.3	67.7	32.3	49.5	50.5
Molise	4.0	96.0	7.3	92.7	72.2	27.8	54.7	45.3
Campania	5.5	94.5	7.3	92.7	80.8	19.2	62.7	37.3
Puglia	5.6	94.4	9.3	90.7	78.6	21.4	57.8	42.2
Basilicata	4.6	95.4	8.8	91.2	77.0	23.0	59.1	40.9
Calabria	4.0	96.0	9.0	91.0	77.4	22.6	59.3	40.7
Sicily	5.4	94.6	9.5	90.5	82.9	17.1	61.7	38.3
Sardinia	5.0	95.0	9.2	90.8	79.0	21.0	56.5	43.5
Aut. Prov. Bolzano (l. it.)	4.5	95.5	8.2	91.8	67.3	32.7	43.2	56.8
Aut. Prov. Trento	6.0	94.0	8.9	91.1	73.8	26.2	49.9	50.1
Italy	5.2	94.8	8.0	92.0	74.8	25.2	53.5	46.5

ISBN 9788835123330

3.2. Spatial autocorrelation analysis: Moran's I and LISA Cluster Map

The Italian country, because of its breadth, complexity and heterogeneity, may not allow descriptive analysis to bring out different territorial realities, catchment areas of equally different educational institutions. For this reason, we considered appropriate to deepen the investigation through the spatial autocorrelation method. We applied Moran's I index, Moran's Scatterplot (Anselin, 1996) and Local Indicator of Spatial Association (LISA) (Anselin, 1995), using GeoDa, a software for analysis of spatial data (Anselin, 2003) and Tableau software (v. 2020.1) for graphical representation on map.

Moran's I provides an indication of the degree of linear association between the observed values of the study variable and the spatially delayed values. This type of spatial autocorrelation measures requires the construction of the weight matrix that defines a neighbourhood for each geographical unit. The value of the study variable for each unit is compared with the average weight of the values of the neighbouring units.

To construct the matrix of weights, based on the spatial distances between two points expressed in terms of latitude and longitude, the *great circle distance* or *arc-distance* were used, taking into account the terrain curvature. The average number of analysed "neighbours" was 165, the median 93.

The univariate Global Moran's I revealed the presence of spatial autocorrelation in the distribution of AA scores ($I \approx 0.035$), DD scores ($I \approx 0.058$), *StudioGenitori* ($I \approx 0.114$) and *DigitalGenitori* ($I \approx 0.125$), in all of cases the index was significant (the pseudo p-value is < 0.001).

As previously mentioned, based on the assumption that in a given geographical unit the values of the observed variable show a systematic association with another observed variable in adjacent geographical units, the bivariate Moran's I was used to explore and analyze the spatial dependence between the AA (and DD) score and the *StudioGenitori* and *DigitalGenitori* indicators.

The bivariate Global Moran's I was significant (the pseudo p-value is $<$ 0.001) for each couple of indicators and revealed the presence of spatial dependence of negative sign ($I \approx -0.010$ for AA and SG; $I \approx 0.000$ for AA and DG; $I \approx -0.041$ for DD and SG; $I \approx -0.035$ for DD and DG).

However, the Moran's I index does not allow to verify whether spatial dependence generates clusters of schools by level of educational performance, nor to identify geographical boundaries of these clusters. In order to overcome these limits, the Moran's Scatterplot and local autocorrelation measures were applied.

ISBN 9788835123330

To measure the interdependence with other schools and to indicate the type of the observed spatial dependence and its significance the bivariate Local Indicator of Spatial Association (LISA) was calculated.

This type of analysis makes it possible to provide research with a deeper level of geographical detail and to make both the administrative geographical boundaries and the morphological characteristics of a territory more flexible.

The results of the bivariate LISA indicator were displayed in the map named *cluster map* (Fig. 1A-4A), in which 4 different clusters of educational institutions were identified:

– the first cluster (High-High, it is red in the maps) is composed by schools that had AA (or DD) scores above the national average and are surrounded by institutions whose *StudioGenitori* (or *DigitalGenitori*) is above the average. This type of cluster is present in the Centre-North of Italy;
– the second cluster (Low-Low, blue in the maps) is made up by schools that register AA (or DD) scores lower than the national average and are adjacent to schools with the *StudioGenitori* (or *DigitalGenitori*) lower than the national average; primarily it includes the regions of Southern Italy and the Islands;
– the outlier group in purple in the map (Low-High) is composed by schools with AA (or DD) scores lower than the national average that seem to not benefit from their geographical location. In fact, they are surrounded by schools with average higher values of the *StudioGenitori* (or *DigitalGenitori*);
– the last High-Low outliers group (pink in the map), is formed by schools with AA (or DD) scores above the national average that are surrounded by schools with *StudioGenitori* (or *DigitalGenitori*) values below the national average;
– schools with not significant LISA are showed in grey in the map.

Cartograms shown in Figures 1B-4B, called *significance map* represent localization of institutions with significant LISA in different shades of green, depending on the level of significance.

In Figure 1, it is immediately evident that, by relating the Digital Device (DD) and *StudioGenitori* (SG) indicators, there is a strong concentration of significant LISAs in correspondence with large cities and regional and provincial capitals (e.g. Milan, Rome, Naples).

The cluster map shows a clear division of Italy into two sections: in the Centre-North regions the High-High (H-H) cluster and the Low-High (L-H) outlier are more present, while in the South, the Low-Low (L-L) cluster and High-Low (H-L) outliers.

ISBN 9788835123330

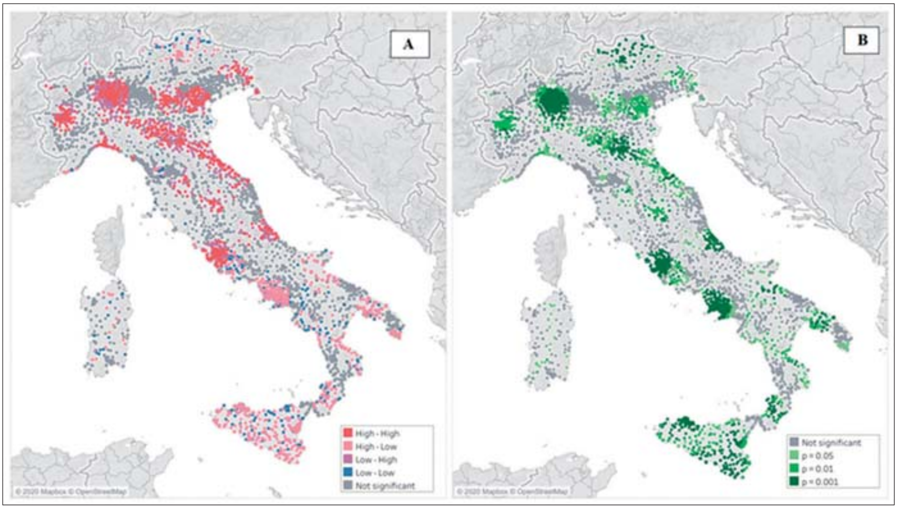


Fig. 1 – Bivariate LISA cluster map and significance map: Digital Device (DD) and StudioGenitori (SG)

In particular, the H-H and L-H clusters are present in Valle d'Aosta, in north-eastern Piedmont (especially in the province of Turin), in some municipalities in the provinces of Biella and Novara on the boundary with western Lombardy and the provinces of Milan and Pavia, then continue in the provinces of Bergamo, Monza-Brianza, Como, Lecco, Lodi. The same group of schools also includes some institutes in the province of Brescia and Mantua on the border with Veneto, mainly Verona. The provinces of Padua, Venice and Vicenza are affected in Veneto. H-H and L-H clusters also include a group of schools along the Emilian Apennines (provinces of Modena, Reggio-Emilia and Bologna), along the Adriatic coast of Emilia, some institutes in the provinces of Ravenna, Forlì-Cesena and Rimini, in the northern Marche in the province of Ancona and Pesaro-Urbino and finally in the Abruzzo provinces of Pescara, Chieti and Teramo.

Along the western axis, the mentioned clusters are present in the province of Genoa, in the Tuscan provinces of Florence, Arezzo and Siena, and in the province of Perugia. Lazio region is clearly divided into two: the H-H and L-H clusters characterize the province of Rome while the L-L and H-L clusters characterize the provinces of Frosinone and Latina. In some region of Northern Italy, however, there are clusters of L-L and H-L schools that are typical of the South: some schools in the province of Bolzano, others in the province of Trento and Belluno on the boundary with Bolzano; and in Tuscany, some institutes in Lucca-Pisa-Pistoia triangle.

ISBN 9788835123330

In the South, the L-L and H-L clusters are present in: Campania, mostly on the Tyrrhenian coast (Caserta, Naples and Salerno); Puglia, in Brindisi, Taranto and Barletta-Andria-Trani provinces, some schools in the province of Foggia and a group of schools in southern Salento; Calabria, the province of Reggio Calabria, schools on the Tyrrhenian coast in Vibo Valentia and on the Ionian coast as Cosenza and Crotone; some Basilicata schools on the boundary with Campania, Calabria and Puglia are also affected too. In Sicily, the most important concentrations are present in the provinces of Palermo and Catania, but the two clusters are widespread throughout the region, as well as in Sardinia.

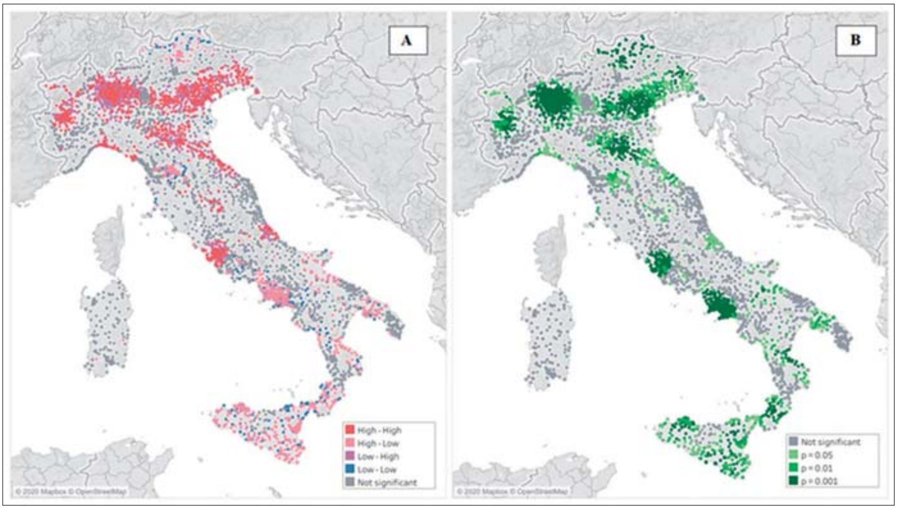


Fig. 2 – Bivariate LISA cluster map and significance map: Digital Device (DD) and DigitalGenitori (DG)

In Figure 2, which represents the relationship between the Digital Device (DD) and DigitalGenitori (DG) indicators, there is also a strong concentration of significant LISAs in correspondence with large cities and regional and provincial capitals, as well as a clear division of the Centre-North and South of Italy. In this case, the H-H and L-H clusters in the North are characterized by a certain continuity, creating an almost unicum between Valle d'Aosta, north-eastern Piedmont (especially the province of Turin), some municipalities in the provinces of Biella and Novara on the boundary with western Lombardy and the provinces of Milan and Pavia, then continuing in the provinces of Bergamo, Monza-Brianza, Como, Cremona, Lecco, Lodi. Some schools in the province of Brescia and Mantua on the boundary with

ISBN 9788835123330

Verona act as a bridge between Lombardy and Veneto. In Veneto the mentioned clusters are highly dense, affecting the provinces of Padua, Venice, Vicenza, Treviso and Belluno. It also includes a group of schools along the Emilia Apennines (provinces of Modena, Reggio-Emilia, Bologna), the Adriatic coast of Emilia in the provinces of Ravenna, Forlì-Cesena and Rimini and the northern Marche in the province of Pesaro-Urbino. In Abruzzo the two clusters are represented by some schools in the coasting areas of the provinces of Pescara, Chieti and Teramo.

In the western regions, the clusters are present in the province of Genoa and La Spezia, in the Tuscan provinces of Arezzo and Siena, on the boundary with the province of Perugia in Umbria. Lazio is again clearly divided into two: the H-H and L-H clusters characterize the province of Rome, the L-L and H-L clusters are present in the provinces of Frosinone and Latina.

The same clusters, typical of the southern regions, are present once again in the northern regions in some schools in the province of Bolzano and Trento on the boundary with Bolzano; in Tuscany, in the provinces of Florence, Lucca, Prato and Pistoia. In the South, the L-L and H-L clusters are present in Campania, mostly on the central and northern Tyrrhenian coast: Caserta, Naples and Salerno. In Puglia the clusters are present in the provinces of Brindisi and Taranto, Barletta-Andria-Trani and in some schools in the province of Foggia; in Calabria, in the province of Reggio Calabria, in some schools on the Tyrrhenian coast of Vibo Valentia and on the Ionian coast of Cosenza and Crotone. Basilicata is present only for some schools on the boundary with Campania, Calabria and Puglia. In Sicily, there are concentrations mainly in the provinces of Palermo and Catania, but clusters are widespread throughout the region. Sardinia does not seem to be affected by any significant cluster.

In Figure 3, linking the indicators AA and *StudioGenitori* (SG), is clearly visible a pattern similar to that already observed in Figure 1, as well as in Figure 4, which shows the relationship between the indicator AA and *DigitalGenitori* (DG) similar to Figure 2.

ISBN 9788835123330

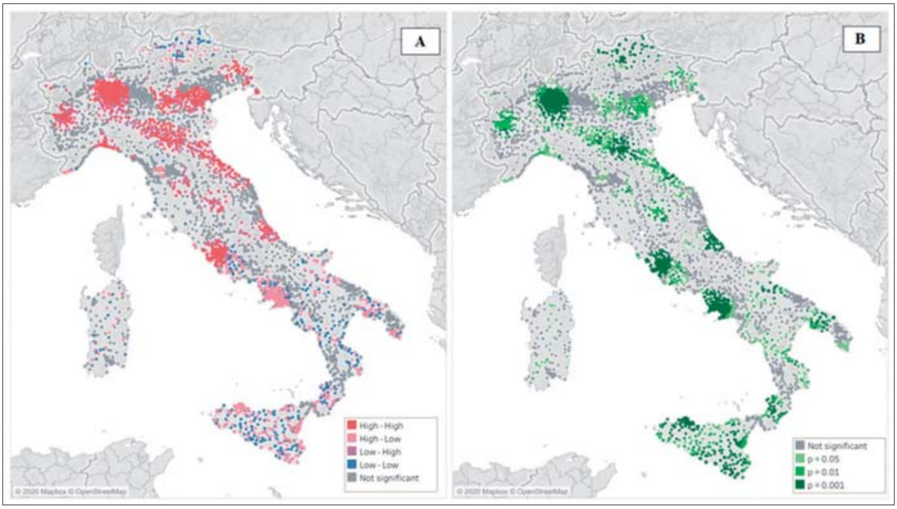


Fig. 3 – Bivariate LISA cluster map and significance map: Ambiente di Apprendimento (AA) e StudioGenitori (SG)

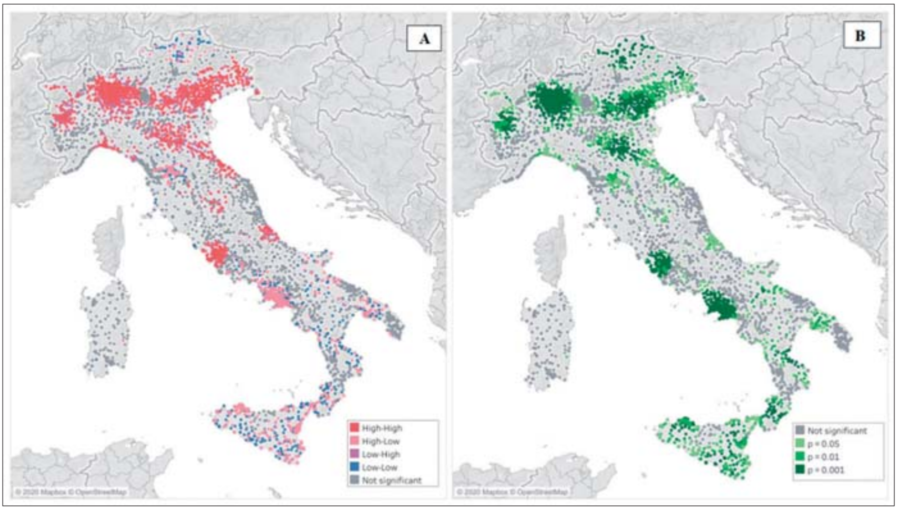


Fig. 4 – Bivariate LISA cluster map and significance map: Ambiente di Apprendimento (AA) and DigitalGenitori (DG)

ISBN 9788835123330

4. Discussion

Taken together, our data show that, considering the four indicators, Italy appears divided into two, but not in a clear way as could be expected at a theoretical level. On the one hand, in fact, research confirms that there is a certain territorial dualism between Centre-North (characterized by cluster H-H) and South (characterized by cluster L-L), on the other hand, instead, the boundaries between these two zones are not always so clear and there are territories of one zone that show typical behaviors of the other.

For example, the provinces of Trento and Bolzano which, although in the North, show clusters similar to the schools of the South for all the indicators considered. Within these two macro-zones (Centre-North and South), moreover, it is evident the presence of outliers (cluster L-H for the Centre-North and H-L for the South) and this shows that Italy has four different speeds: in the Centre-North there are clusters of schools in which students do not have an adequate learning environment and a good supply of digital devices (below average), surrounded by schools attended by students with highly educated parents with satisfying digital skills (above average). Vice versa in the South there are clusters of schools with students who have an adequate learning environment, equipped with digital devices (above average), surrounded by schools attended by students whose parents have a level of education and digital skills lower than average.

It is important to note that, in addition to the North-South contrast, there are strong differences between cities of different demographic size. This can be easily explained, especially as regards access to the network, through some infrastructure differences that characterize the country: in metropolitan areas broadband access rates reach 78.1% while in municipalities up to 2,000 inhabitants this share drops to 68.0% (ISTAT, 2019). Moreover, since the cost of living in metropolitan areas is higher than in small countries (ISTAT, 2017), it is also likely that more wealthy families, formed by well-educated people with a good job, will be concentrated there.

Parent's profession, in any case, confirms to be a fundamental indicator on the possession of material and technological goods. Indeed, the territorial distribution of the relationship between AA and *DigitalGenitori* is very similar to the one between DD and *DigitalGenitori*. This evidence can be attributed to the different distribution of productive activities/services on the national territory and to the different employment situation. Overall, we can say that in the Centre-North of Italy the weight of highly skilled occupations is growing, part-time can often represent a free choice of the worker and there are increasing economic well-being and employment rate, while the South

ISBN 9788835123330

is characterized by a significant process of general job de-qualification, difficulty in finding full-time jobs and a decrease in economic well-being and employment rate (CNEL, 2019).

The educational level and the profession are closely linked. Participation in the labour market is increasingly reserved for those who have a high level of education, which in recent years has been a guarantee against the crisis that has hit the markets (Loriga *et al.*, 2017).

However, the fact that in our study the indicators relating to the qualification and profession of parents, related to AA and DD, provide very similar results confirms that they are both very influential indicators in defining the social and economic background of the families in which Italian students grow up (Campodifiori *et al.*, 2010; Ballarino and Schadee, 2006). Parents with a high level of education and good employment can provide their children with a comfortable home environment, equipped with digital devices, and help them interact with these tools.

Children and adults are in fact accustomed to a radically opposite use of the computer: fun/expressive/recreational aid normally reserved for leisure time for the first one (video games, download and listening to music, download and video viewing), the computer is instead linked to the work environment for parents, but not for all parents in the same way and in equal measure, just because they have different professions (Feola, 2015) and qualifications.

Parents with a qualified profession and a high academic degree, who are familiar with the network and good computer skills, can more easily help their children in the digital activities (Murru, 2012; Livingstone and Helsper, 2008). On the other hand, in the most disadvantaged families, parents tend to use the Internet sporadically or to use it only for entertainment, they have fewer skills and, consequently, are less able to help and guide their children (De Almeida, Nuno De Almeida and Carvalho, 2011; Aroldi, 2012). These data confirm what emerged in the ISTAT 2020 Report, according to which the digital gap between households is due to social and territorial factors: the probability that there is at least one member with high digital skills in a family is eight times higher if there is a graduate member in the family than those in which the highest qualification is at most the lower secondary school education license. The probability is double in families where the head of the household is an entrepreneur manager or freelancer compared to those where the head of household is a worker (ISTAT, 2020).

Nevertheless, recalling the four different speeds mentioned at the beginning of this paragraph, we can say that, in the two areas of the country, the cultural and economic status of families is not always an infallible predictor of a certain level of possession of goods. In particular, in the South, the

ISBN 9788835123330

widespread presence of H-L outliers indicates that, although the social and geographical context in which these schools are located, families formed by people with low educational qualifications and low-skilled occupations are able to guarantee their children a serene domestic environment and the possession of some material goods, including technological ones.

5. Conclusions

Home and technology were certainly the two words that characterized the quarantine period due to COVID-19 medical emergency for all Italian families. House is often considered as a place to rest, to refresh yourself but for a few months it became the only living space for the whole family. Furthermore, technologies have been literally "invaded by reality" playing a substantial role in maintaining the essential functions of society and sociality: home working, online teaching, all kinds of care services carried out through telematic means, communication through social channels, online payments and fundraising campaigns through digital platforms, etc. In essence, the essentials of the economic, working, relational and educational activities of our societies have been transferred to the network (Benanti, Darnis and Sciarrone Alibrandi, 2020).

Changes that have taken place (and still existing) should risk to amplify the already existing inequalities. School, which clearly has a primary role in tackling any kind of inequalities, and in guaranteeing the educational success of all students has demonstrated the full height of a completely new and unpredictable challenge in this occasion as well. Moving into the houses, however, school had to deal more than ever with the living conditions of students and with their material difficulties.

This study has shown, for example, that almost 40 thousand students in the fifth grade of primary school, with significant territorial differences, may have encountered difficulties with distance learning due to the lack of indispensable tools (DD = 0). We use a hypothetical sentence because dataset of this study was collected at the end of 2018/19 school year, therefore several months before the start of the health emergency, and this is undoubtedly a limit. Another limitation is linked to the fact that, with the data at our disposal, we have been able to investigate only the possession of a PC and the use of the latter for the study, neglecting all the other technological devices that may be available to students (in first of all cell phones and tablets).

One of the peculiarities of this study was the investigated school grade (5^{th} grade), since research of this type has often focused on older students (Micheli, 2015; Gui and Argentin, 2011).

ISBN 9788835123330

Possible developments could therefore be linked to similar analyses but carried out on students attending higher school grades. It should also be noted that the Italian government, in this period of health emergency, has allocated several million euros to make available to the poorest students, on loan free of charge, individual digital devices and even households have clearly activated in this sense: an additional research trail could be obtained by repeating the analyses one or two years later to study the outcome of these processes.

At a time like the present, when economic and cultural differences could threat the effective right to education (compulsory and free of charge) enshrined in the Constitution, it is crucial and undoubtedly urgent to study the impact of the policies implemented, also in relation to the transnational objectives that Italy shares, such as the reduction of disparities related to income, sex, age, disability, race, religion, economic status included as goal 10 within the 2030 Agenda.

References

Anselin L. (1995), "Local indicators of spatial association – LISA", *Geographical Analysis*, 27, 2, pp. 93-115.

Anselin L. (1996), "The Moran scatterplot as an ESDA tool to assess local instability in spatial association", in M.M. Fischer, H. Scholten, D. Unwin (*eds.*), *Spatial analytical perspectives on GIS in environmental and socio-economoc sciences*, Taylor and Francis, London, pp. 111-125.

Anselin L. (2003), *GeoDa 0.9 User's Guide. Spatial Analysis Laboratory (SAL)*, Department of Agricultural and Consumer Economics, University of Illinois, Urbana-Champaign (IL).

Aroldi P. (2012), "Ripensare il rapporto tra media e generazioni: concetti, indicatori, modelli", in F. Colombo, G. Boccia Artieri, L. Del Grosso Destrieri, F. Pasquali, M. Sorice (a cura di), *Media e generazioni nella società italiana*, FrancoAngeli, Milano, pp. 33-64.

Ballarino G., Schadee H.M.A. (2006), "Espansione dell'istruzione e diseguaglianza delle opportunità educative nell'Italia contemporanea", *Polis*, 2, pp. 207-232.

Benanti P., Darnis J.P., Sciarrone Alibrandi A. (2020), "Per una resilienza con la tecnologia. Appunti per il post Covid-19", in C. Caporale, A. Pirni (a cura di), *Pandemia e resilienza. Persona, comunità e modelli di sviluppo dopo la Covid-19*, CNR, Roma, pp. 113-121.

CNEL (2019), *Rapporto mercato del lavoro e contrattazione collettiva 2019*, Roma.

Campodifiori E., Figura E., Papini M., Ricci R. (2010), "Un indicatore di status socio-economico-culturale degli allievi della quinta primaria in Italia", *INVALSI Working paper*, 2.

ISBN 9788835123330

Cappellari L. (2006), “Background familiare, scelte formative e transizione scuola-università”, in G. Ballarino, D. Checchi (a cura di), *Sistema scolastico e disuguaglianza sociale. Scelte individuali e vincoli strutturali*, il Mulino, Bologna, pp. 57-90.

Checchi D. (2010), “Percorsi scolastici e origini sociali nella scuola italiana”, *Politica economica*, 26, 3, pp. 359-388.

Checchi D., Fiorio C.V., Leonardi M. (2006), “Sessanta anni di istruzione in Italia”, *Rivista di Politica Economica*, 7, pp. 285-318.

De Almeida N. D., Nuno De Almeida A., Carvalho T. (2011) “Children and Digital Diversity: From ’Unguided Rookies’ to ’Self-reliant Cybernauts’”, *Childhood*, 19, 2, pp. 219-234.

Demarinis G., Iaquinta M., Leogrande D., Viola D. (2011), “Analisi quantitativa della mobilità studentesca negli atenei italiani. Confronto territoriale fra domanda e offerta di formazione universitaria”, in D. Viola (a cura di), *Valutazione e qualità degli atenei: Modelli, metodi e indicatori statistici*, Università degli studi di Bari, Bari, pp. 273-303.

Feola E.L. (2015), “La famiglia come luogo di costruzione di modelli d’uso e di concezioni di consumo mediatico dei ragazzi”, *In-formazione*, X, 13, pp. 46-49.

Gui M., Argentin G. (2011), “Digital skills of internet natives: Different forms of digital literacy in a random sample of Northern Italian high school students”, *New Media & Society*, 13, pp. 963-980.

ISTAT (2017), *Annuario statistico italiano*, Roma.

ISTAT (2019), *Cittadini e ICT. Anno 2019*, 18 dicembre, Roma.

ISTAT (2020), *Rapporto annuale 2020. La situazione del Paese*, Roma.

Livingstone S., Helsper E.J. (2008), “Parental Mediation of Children’s Internet Use”, *Journal of Broadcasting & Electronic Media*, 52, 4, pp. 581-599.

Loriga S., Spizzichino A., Gisbert Martì O., Franco I. (2017), “Popolazione e titolo di studio: nuovi dati per analisi di lungo periodo”, *Rivista Italiana di Economia Demografia e Statistica*, LXXI, 4, October-December, pp. 21-30.

Malaguzzi L. (2010), *I cento linguaggi dei bambini. L’approccio di Reggio Emilia all’educazione dell’infanzia*, Junior, Bergamo.

Micheli M. (2015), “L’appropriazione di Internet da parte degli adolescenti: tra riproduzione sociale e mutamento culturale”, *Quaderni di Sociologia*, 69, pp. 7-32.

MIUR (2012), “Indicazioni nazionali per il curricolo della scuola dell’infanzia e del primo ciclo d’istruzione”, *Annali della Pubblica Istruzione*, LXXXVIII, retrieved on July 10, 2020, from http://www.indicazioninazionali.it/wp-content/uploads/2018/08/Indicazioni_Annali_Definitivo.pdf.

Močinić S., Moscarda C. (2010), “L’ambiente come fattore di apprendimento nella scuola dell’infanzia”, *Studia Polensia*, 5, 1, pp. 1-20.

Moran P. (1948), “The Interpretation of Statistical Maps”, *Journal of the Royal Statistical Society*, 10, pp. 243-251.

Murru M.F. (2012), “La mediazione sociale”, in G. Mascheroni (a cura di), *I ragazzi e la rete. La ricerca EU Kids Online e il caso Italia*, La Scuola, Brescia, pp. 237-260.

ISBN 9788835123330

OECD (2020), *L'istruzione scolastica al tempo del Covid-19: gli insegnanti e gli studenti erano preparati? Country note OCSE*, retrieved on July 10, 2020, from https://www.invalsi.it/invalsi/ri/pisa2018/Country_Note.pdf.
Palmerio L., Caponera E. (2020), *La situazione di studenti e insegnanti in relazione all'uso delle tecnologie dell'informazione e della comunicazione nel periodo precedente l'emergenza sanitaria da Covid-19*, retrieved on July 10, 2020, from https://www.invalsi.it/invalsi/ri/pisa2018/situazione_studenti_insegnanti.pdf.
Parziale F. (2016), *Eretici e respinti. Classi sociali e istruzione superiore in Italia*, FrancoAngeli, Milano.
Triventi M. (2014), "Le disuguaglianze di istruzione secondo l'origine sociale. Una rassegna della letteratura sul caso italiano", *Scuola democratica*, 2, May-August, pp. 321-341.
Wilson B.G. (1996), *Constructivist Learning Environments. Case Studies in Instructional Design*, Educational Technology Publications, Englewood Cliffs (NJ).

ISBN 9788835123330

3. Gender asymmetries: an analysis of the trend from I to II cycle of education

by Andrea Bendinelli, Michele Cardone, Patrizia Falzetti

Based on the 17 *Sustainable Development Goals* (SDGs), this work focuses on Goal 5: "Achieving gender equality, for the empowerment of all women and girls", deepening the gender differences in the high school segment. Using INVALSI data, the students in the 8th grade of 2014 were matched with those in the 13th grade of 2019, considering three dichotomous as output variables: the completion of the last school year without repeating, the achievement of a level of competence attributed by INVALSI above/below "3 out of 5" which represents the threshold of sufficiency, or above/below "4 out of 5" which represents the threshold of the top performers). Using these three dichotomous variables, three logistic regression models were constructed respectively, using as predictors some students' characteristics: socio-demographics (gender, origin, geographical area and family background), school performance (repetition at the low secondary school and score at the INVALSI 8th grade tests of Italian language and Mathematics) and the desired educational qualification.

The models built on the levels of competence show a better representativeness than the first one based on the achievement of grade 13, which we intend to re-evaluate in the future.

The results in part confirm what has already emerged in literature, in part they provide some new food for thought. From literature we know that females obtain better results in literary subjects and males in Mathematics and the gap in favour of females is reduced considering the top performers. We also know that the geographical area, the origin of the pupil and the family background are main factors on performance: all these assumptions are already confirmed by the descriptive statistics provided. The proposed models, that is when we check on the basis of the proposed explanatory variables, however, also suggest other reflections: the geographic area remains influen-

ISBN 9788835123330

tial but the north-south gap drops considering the top performers; the family background has a lower effect, the family origin the pupil becomes almost irrelevant when we exclude repeating and drop out students, while the desired qualification and 8th grade test score have the most influential odds ratios of all.

Regarding gender, an interesting result also emerges: if we check with respect to the other predictor variables, the effect in favour of females in Italian language almost disappears (odds ratio around 1), while that in favour of males in Mathematics remains strong. In other words, considering those students reaching the last high school year without repeating, with the same demographic and social conditions, same school results in grade 8th and same qualification desired, a male student is twice as likely to reach a level at least equal to "3" (sufficient) or at least equal to "4" (top performer) in Mathematics.

1. Theoretical framework and Literature

On September 25, 2015, the United Nations approved the 2030 Agenda[1] for Sustainable Development and the 17 *Sustainable Development Goals* (SDGs), divided into 169 Targets to be achieved by 2030. The document expresses a clear judgment on the unsustainability of the current development model on different levels, environmental, economic and social. All countries are called to participate, each according to the level of development achieved. In essence, each country is asked to define a sustainable development strategy that allows reaching the SDGs, accounting for the results obtained through a process coordinated by the UN. In order to achieve the objectives it is necessary to involve all components of society, from businesses to the public sector, from civil society to philanthropic institutions, from universities and research centers to information and culture operators.

In the list of 17 SDGs, the Goal 5: "Achieving gender equality, for the empowerment of all women and girls" is what this work focuses on. As well as access to work, we wonder what access to education is for girls, considering an entire segment that goes beyond compulsory school, the so called in Italy "second cycle" (high school).

This intent is inscribed in the vast literature on the subject focused on comparative international research promoted mainly by the OECD (Organization for Economic Cooperation and Development) and by the IEA (Inter-

[1] https://unric.org/it/wp-content/uploads/sites/3/2019/11/Agenda-2030-Onu-italia.pdf.

ISBN 9788835123330

national Association for the Evaluation of Educational Achievement), which show how the best results tend to be associated with males when the subject regards scientific-mathematical skills, with females when, instead, the subject regards literary skills.

In the field of educational research, the topic of gender differences in school learning has been widely studied in recent decades (among all Sammons, 1995; Fryer Jr. and Levitt, 2009; Stoet and Geary, 2013). The studies conducted so far agree that males achieve better school performance in the mathematical-scientific sphere, while females in the linguistic sphere. Gender differences in school results are the subject of different interpretations which on the one hand involve individual factors, on the other factors affecting the social and family context and the socialization processes through which they are conveyed. In the case of individual variables, in addition to the different intellectual maturation process that affects males and females (Cicero, 2004), the interest and the way of approaching the subjects of study are fundamental. Male pupils show a safer attitude and lower anxiety levels towards mathematical and spatial skills (Poliandri *et al.*, 2001); they also show greater self-control in dealing with particularly stressful and unexpected situations: an interpretation is that boys achieve higher scores in standardized tests because they have always been more used to competitions (Steele, 1997). Conversely, female pupils appear more predisposed for the sphere of language, facing problematic situations with greater stress and concern (Else-Quest, Hyde and Linn, 2010; Zaniello, 2012). These propensities can in turn be influenced by aspects that materialize in the educational models that structure gender identity and, not least, in the transmission of the relative stereotypes. For example, the attitude through which parents implicitly communicate their beliefs about their children's school skills affects not only their immediate performance, but also guides future ones, representing a predictor of subsequent study and career choices. In fact, the association, even implicit and symbolic, between gender and school subjects intervenes in the acquisition of skills and competences, and also structures the perception of their achievement (Bleeker and Jacobs, 2004; Tomasetto, Alparone and Cadinu, 2011; Tomasetto, 2013). Other interpretations f the same kind are a) standardized tools adapt with more difficulty to the "feminine" style which requires to argue one's own answers without the typical restrictions of closed-response modes (Sternberg and Williams, 1995); b) the stress generated by gender stereotypes in Mathematics is added to the "competition" stress (Davies and Spencer, 2005).

The empirical evidence about gender differences is also confirmed by international surveys on learning. The PISA (*Program for International Stu-*

ISBN 9788835123330

dent Assessment) survey promoted by OECD and conducted on samples of fifteen-years-old students from different countries, shows the specificity of the patterns deriving from the gender effect in Mathematics and Reading results[2]. In particular, in Italy these gaps between females and males were all statistically significant, both in Reading and Mathematics, in the last three PISA editions[3]. The situation is similar for Science, where about two out of three top performer students are boys (INVALSI, 2016c).

The IEA PIRLS (*Progress in International Reading Literacy Study*) and TIMSS (*Trends in International Mathematics and Science Study*) surveys also highlight specific dynamics affecting gender factors. The PIRLS 2016 results confirm those of other previous surveys on Reading: the average difference in scores on the overall scale is in favour of females, settling, for Italian students of the 4th grade of schooling, at 7 points compared to 19 points in the participating countries. This trend is confirmed by the scores on the partial scales that follow the international reference averages for Italy, but with slightly lower values (INVALSI, 2017). The data of the IEA-TIMSS survey over time do not show any particular changes for the results of Mathematics at the 4th grade, but the advantage of males over females in the last year of survey appears to be significant in Italy, the largest among all countries involved (INVALSI, 2016a). This trend is also found for the 8th grade, both as regards the mathematical and scientific fields (INVALSI, 2016b).

Interesting effects on the performance of boys and girls can also be observed by studying the relationship between social origin and academic performance. In addition to the main finding showing how students who come from families with a low level of socio-economic-cultural status (ESCS) achieve more modest academic results than students with a higher background (Sirin, 2005; Eurydice, 2010), the interaction between these variables and gender highlights specific behaviour patterns. In general, the influence of socio-economic-cultural status appears more marked in a positive sense for males than females because the former are more sensitive to resources present in the family and learning context. This gap tends to decrease for lower ESCS levels (Legewie and Di Prete, 2012).

[2] In particular, according to PISA 2018, in 36 over 36 OECD countries females scores better than males in Reading, with 34 differences statistically significant, while males scores better than females in 31 OECD countries over 37, but with only 8 of those differences are statistically significant (our elaboration of PISA data).

[3] The gap in favor of females in Reading was +39 (2012), +16 (2015), +25 (2018); the gap in favor of males in Mathematics was +18 (2012), +20 (2015), +15 (2018). PISA scores are on a scale of 500.

ISBN 9788835123330

2. Research hypotheses

A longitudinal analysis of the results obtained by the students of the secondary school – those who in 2014 took the state exam in grade 8 and in 2019 took the final exam – showed that 1 out of 5 pupils did not managed to complete his studies at the year expected, because he repeated one or more years, or even dropped out, abandoning the training experience ahead of time and never returning to it[4].

The hypothesis behind this research work is that gender differences in the segment related to the first cycle of education see girls ahead of men in the comprehension tests. This advantage tends to decrease with the progress of the school path up to almost zero at the end of the second cycle of education. The situation partly changes when considering Mathematics tests: differences are moderate in the first cycle of education, which sees males slightly ahead of females, then diverge in favour of males reaching important differences at the end of the school course.

The aim of this work is to verify gender differences in the second cycle of the school curriculum (high school), considering 8th grade male and female students reaching 13th grade without repeating, and comparing their level of competence achieved at the end of the cycle, taking into account different characteristics such as demographics (gender, italian origin or not, ESCS and geographic area), related to school performance (repetition at lower secondary school and 8th grade INVALSI test scores) and personal aspiration (desired educational qualification).

3. Data, methods and assumptions

The data used for this study are those of the national learning assessment administered by INVALSI to all the students of the grades concerned. The dataset is structured with a starting year coinciding with the tests administered at the end of the first cycle of education (8th grade) in 2014 and ends with the tests taken at the end of the second cycle in 2019 (13th grade). The connection of single students in the cohorts was made possible thanks to the unique SIDI code provided by the Ministry of Education, in particular this code allows each student to be traced anonymously longitudinally during the various surveys. In the case of this work, the pupils participating in the 2014

[4] https://www.invalsiopen.it/wp-content/uploads/2019/11/Editoriale2_mediamaturita%CC%80.pdf.

ISBN 9788835123330

8th grade test (all of them) were linked via SIDI to the pupils who took the 10th grade test in 2016 and the 13th grade test in 2019 (therefore in order with the course of studies).

Of the initial 8th grade cohort only a part performed the 13th grade test; the absence of the data may have occurred for one or more of the following reasons:

- repeating pupils between grade 9 and grade 12 (repeating students take the13th grade test at least in the 2019/20 school year);
- pupils who have not taken the INVALSI test at the end of grade 13 in 2019 by choice or for personal reasons;
- pupils enrolled in vocational training, who took a different test;
- private students, not included in the data used;
- dropout pupils and pupils whose SIDI for some reason does not coincide between the two surveys.

The outcome variables considered are:

- **Y1**: dependent variable built on the entire population of 8th grade 2014 pupils, which assumes a value of 1 if the student has reached the 13th grade final exam in 2019, "0" if not;
- **Y2**: dependent variable constructed on the population of 13th grade pupils with an assigned level of competence at the 13th grade test in 2019, which assumes value "1" if the student has obtained a competence level equal or higher than level 3 (level 3 is considered the level of "sufficient"), "0" if lower than 3;
- **Y3**: dependent variable constructed on the population of 13th grade pupils with an assigned level of competence at the 13th grade test in 2019, which assumes value "1" if the student has obtained a competence level equal or higher than level 4 (level 4 is considered the threshold of high performers), "0" if lower than 4.

These dependent variables, considered as "response" in the models proposed further on, are built by connecting pupils between 8th grade 2014 and 13th grade of 2019, and are affected by the physiological discrepancy (for the reasons mentioned before) between the population of 8th grade pupils in 2014 and the part of them actually matched through SIDI code to the 13th grade 5 years after. This discrepancy was quantified in Tab. 1, which shows also the amount of students taken into account in each context. Here it's useful to remember that a) the discrepancy mainly affects the response variable Y1, for which with "0" (no grade 13 test taken) we encode several other realities, previously described; and b) the data used for model 2 and model 3 decreases because we include only those students with no missing in any variable.

ISBN 9788835123330

Tab. 1 – Data and dependent variables used

Context	*Students*	*%*	*DV*
Total pupils 8th grade 2014:	515,543	100.0	–
of which matched with 13th grade 2019*	344,459	66.8	–
of which used for model 1	283,890	55.1	Y1
of which used for model 2	242,042	46.9	Y2
of which used for model 3	242,042	46.9	Y3

* This percentage refers to a data match which can fail for different reason, so it's not an indication of the effective school dropout.

The independent variables considered (i.e. predictors in the models), displayed in Tab. 2 can be grouped in two aspects.

Socio-demographics:

– "family origin" indicator (based on the place of birth of the pupil and parents): native, first generation foreigner, second generation foreigner;
– ESCS socio-economic cultural background index detected at 10th grade: categorized in quartiles, from the lowest to the highest;
– geographical area: North-West, North-East, Centre, South, South-Islands[5];
– desired qualification (from the student questionnaire at 10th grade 2016)[6]: categorized in "no university degree", "university degree or similar".

Related to school performance:

– "career regularity" indicator at the 1st school cycle (low secondary school): non repeating, repeating;
– 8th grade 2014 test score in Italian language and Mathematics: categorized in quartiles from lowest to highest: they represent the level of entry competence.

Two important considerations arise from the descriptive statistics of Tab. 2. The first is that the variables "desired qualification" and "ESCS", since they are detected during the grade 10 test, which in 2016 was less "participated" than in the last few years, reduce the study population by 344,459 pupils (i.e. successful match based on the SIDI code) to less than 290,000 used in the proposed models, but it was still chosen to include them as deemed important as predictors.

[5] This classification guarantees a balanced distribution of the pupils in 5 almost equal parts (around 20%).

[6] The question administered was: "Q12 – What is the qualification you intend to achieve?". A dichotomous recoding of the various options was carried out.

ISBN 9788835123330

Tab. 2 – Descriptive statistics of the predictors with respect to gender (8th grade students 2014) – In bold the predictors more characterized by gender

Predictors		*Gender*			
		Male	*Female*	*Total*	
			Row %		
Family origin indicator	Italian native	50.2	49.8	100.0	461,695
	Foreigner I g.	50.4	49.6	100.0	26,647
	Foreigner II g.	50.2	49.8	100.0	24,108
	Total	50.2	49.8	100.0	512,450
ESCS	1° quartile	47.6	52.4	100.0	72,480
	2° quartile	47.7	52.3	100.0	72,218
	3° quartile	48.7	51.3	100.0	69,021
	4° quartile	49.9	50.1	100.0	75,673
	Total	48.5	51.5	100.0	289,392
Geographical area	North-West	50.0	50.0	100.0	124,014
	North-East	50.0	50.0	100.0	91,967
	Centre	50.4	49.6	100.0	94,467
	South	50.4	49.6	100.0	117,081
	South-Island	50.3	49.7	100.0	85,655
	Total	50.2	49.8	100.0	513,184
Desired qualification (10th Grade Qst)	No Univ. degree	57.8	42.2	100.0	116,808
	Univ. degree/ more	42.0	58.0	100.0	168,654
	Total	48.4	51.6	100.0	285,462
Repeating student at lower secondary school	Not repeating	49.1	50.9	100.0	466,413
	Repeating	60.9	39.1	100.0	46,651
	Total	50.2	49.8	100.0	513,064
2014 8th grade INVALSI Italian language test score	1° quartile	56.2	43.8	100.0	128,246
	2° quartile	52.1	47.9	100.0	128,233
	3° quartile	48.8	51.2	100.0	129,193
	4° quartile	43.7	56.3	100.0	127,512
	Total	50.2	49.8	100.0	513,184
2014 8th grade INVALSI Mathematics test score	1° quartile	47.4	52.6	100.0	128,261
	2° quartile	48.3	51.7	100.0	128,236
	3° quartile	49.7	50.3	100.0	133,962
	4° quartile	55.6	44.4	100.0	122,725
	Total	50.2	49.8	100.0	513,184

The second consideration is that among the six predictors proposed are three (values highlighted in bold) those most characterized by gender: one is always the *desired qualification*, with females more eager of a "Degree or

ISBN 9788835123330

similar" than males (16 percentage points above); another is the *score in the 2014 8th grade test*, which is very characterized both by gender and by discipline: in particular, observing the two extreme bands (1st and 4th quartiles), the percentages are diametrically opposite between the subjects: as regards Italian language test, in the best quartile females are 56.3% (against 43.7% of males) and, on the contrary, in the worst quartile males are 56.2% (compared to 43.8% of females); the opposite occurs for the Math test. The last predictor characterized by gender is *Repeating student at lower secondary school*, with 61% males repeating against 31% females. It is clear that when commenting the models' coefficients, this relationship must be taken into account.

Regarding the analysis, as a first step of results we produce descriptive statistics of all the (independent) variables crossed with the three outcome variables (Y1, Y2 and Y3) displayed from Tab. 4 to Tab. 8: a specific focus is dedicated to gender, computing its odds ratio for all the crossings and comparing them. As a second step we use regression models with Y1, Y2 and Y3 as DV and the other as predictors, in order to check if the gender odds ratio change net for the other IV considered (Tab. 9 to Tab. 13).

The model considered is the logistic regression models: this model is used in phenomena in which the response variable is a dichotomous variable. In the present study, the response variable is a binary variable, that is, it assumes two values: 0 or 1 in terms of absence/presence of the characteristic in question. As highlighted before, a first model was constructed considering whether or not a student reaches the final exam of the 13th grade (dependent variable Y1); in the second instance, a second formulation of the same model was built considering only the students who obtained or not at least a level of competence of "3" assigned by INVALSI according to national indications (dependent variable Y2). A final formulation of the model instead focused on excellence, skill levels "4" and "5" according to national indications (dependent variable Y3).

The data used for the models proposed are therefore a part of the 2014 students' cohort, and meet the requirements for the use of logistic regression:

– *Ratio of cases to variables and missing data*. The data does not have any of the critical issues regarding the ratio of cases to variables. As regards the missing data in one or more predictors (missing), the default option has been left (*Exclude*), therefore the cases that have not enhanced all the variables of the model are excluded. So, the different models work on a slightly different number of pupils as explained in Tab. 1;
– *Adequacy of expected frequencies*. Goodness of fit statistics may have little power if expected frequencies are too small. When using tests based on the chi-square, it is best if all expected frequencies are greater than

ISBN 9788835123330

one, and that no more than 20% are less than five. The inverse risk which will be taken into account in the rest of this work is that each model is always significant by adding any variable;
- *Absence of outliers in the solution* (*if fit inadequate*): outlier values, which can affect the model fit, are not present in the database used since the predictors are all discrete or categorized (ESCS and test score);
- *Multicollinearity*: logistic regression, like all varieties of multiple regression, is sensitive to extremely high correlations among predictor variables, signalled by exceedingly large standard errors for parameter estimates. Among the predictors used in the models, there are non-continuous variables; moreover, given that we find very small standard errors (< 0,041), we can assume absence of multicollinearity;
- *Linearity in the logit*: since continuous predictors are not taken into account, the assumption is not violated;
- *Independence of errors*: it is assumed that the answers of the different cases (pupils) are independent of each other, i.e. each answer comes from an independent case. The present data can be considered independent because they're scores coming from independent (but equivalent) standardized tests administered individually to each pupil.

Not having specific hypotheses on the order of importance of predictors in the model, the "direct" or standard insertion method is used, i.e. each predictor is evaluated as if it entered the equation of the model last.

The *Omnibus* test compares the complete model compared to the one with the constant only: in all models the test is positive, that is, the models proposed are significantly better than those with the constant only. However, it should be considered that the fact of having such large "n" cases tends to make any modification to the model significant in terms of added predictors[7]. Therefore, for the verification of the goodness of the model (effect size) other measures will also be observed:
- the pseudo R-squares: Cox-Snell, and Nagelkerke (an adjusted Cox-Snell that can reach 1);
- classification tables will be used but only in terms of comparison between the models, since the approach of this work is causal and not predictive[8];

[7] «Sample size also is relevant because if sample size is very large, almost any difference between models is likely to be statistically significant even if the difference has no practical importance and classification is wonderful with either model» (Tabacknick-Fidell, 2007, p. 456).

[8] That is, the aim is not to create a predictive model to anticipate where to which group the students will take part, but to analyse the influence of certain variables on the response variable.

ISBN 9788835123330

– finally, the models will be evaluated based on the “strength” of the individual coefficients, based on how much the respective odds-ratios differ from 1.

The predictors included in the model reported in Tab. 3 with their respective “contrast” variable used for the analysis.

Tab. 3 – Categorical variables coding in the models (the “contrast” categories in bold)

Variable	*Categories*	*Parameter codes*			
Geographical area	North-West	1	0	0	0
	North-East	0	1	0	0
	Centre	0	0	0	0
	South	0	0	1	0
	South-Islands	0	0	0	1
2014 8th grade INVALSI test score	1° quartile	0	0	0	
	2° quartile	1	0	0	
	3° quartile	0	1	0	
	4° quartile	0	0	1	
ESCS	1° quartile	0	0	0	
	2° quartile	1	0	0	
	3° quartile	0	1	0	
	4° quartile	0	0	1	
Family origin indicator	Italian native	0	0		
	Foreigner I g.	1	0		
	Foreigner II g.	0	1		
Desired qualification (10th Grade Qst)	No Univ. degree	0			
	Univ. degree/more	1			
Repeating student at lower secondary school	Not repeating	0			
	Repeating	1			
Gender	Male	0			
	Female	1			

4. Results

The following tables provide the crosstabs of the independent variables (predictors in the logistic regressions) crossed with the three dependent variables Y1, Y2 and Y3.

ISBN 9788835123330

Tab. 4 – Descriptive statistics of the predictors with respect to the dependent variable Y1

Predictors *No*		*Y1: 13^{th} grade 2019 exam reached after 5 years (without repeating)*		
		Yes	*Total*	
Gender	Male	38.4%	61.6%	100.0%
	Female	27.9%	72.1%	100.0%
	Female odds ratio		1.61	
Family origin indicator	Italian native	30.2%	69.8%	100.0%
	Foreigner I g.	66.3%	33.7%	100.0%
	Foreigner II g.	52.5%	47.5%	100.0%
Repeating student at lower secondary school	Not repeating	28.8%	71.2%	100.0%
	Repeating	78.3%	21.7%	100.0%
Desired qualification (10^{th} Grade Qst)	No Univ. degree	22.9%	77.1%	100.0%
	Univ.degree/more	8.9%	91.1%	100.0%
Geographical area	North-West	37.6%	62.4%	100.0%
	North-East	34.3%	65.7%	100.0%
	Centre	30.6%	69.4%	100.0%
	South	29.1%	70.9%	100.0%
	South-Island	33.9%	66.1%	100.0%
ESCS	1° quartile	53.0%	47.0%	100.0%
	2° quartile	39.4%	60.6%	100.0%
	3° quartile	26.5%	73.5%	100.0%
	4° quartile	13.6%	86.4%	100.0%
2014 8^{th} grade INVALSI Italian language test score	1° quartile	54.2%	45.8%	100.0%
	2° quartile	40.2%	59.8%	100.0%
	3° quartile	25.8%	74.2%	100.0%
	4° quartile	12.3%	87.7%	100.0%
2014 8^{th} grade INVALSI Mathematics test score	1° quartile	53.0%	47.0%	100.0%
	2° quartile	39.4%	60.6%	100.0%
	3° quartile	26.5%	73.5%	100.0%
	4° quartile	13.6%	86.4%	100.0%

Tab. 4, which considers cases of reaching the finale high school exam without having repeated a school year, shows that females are 10 percentage points above males or, considering the odds ratio, they're 61% more likely than males to reach the final exam after 5 years (from grade 9 to 13 without repeating). We remind, again, that the code "No" of Y1, meaning "Not reached the 13^{th} grade 2019 final exam" include all SIDI mismatch, all drop-out or rejected students from grade 9 2014/15 onwards.

ISBN 9788835123330

Tab. 5 – Descriptive statistics of the predictors with respect to the dependent variable Y2 (not sufficient/sufficient competence level) – Italian language

Predictors *< “3”*		*Assigned competence level in 13^{th} grade 2018-19 Italian language test*		
		>= “3”	*Total*	
Gender	Male	33.2%	66.8%	100.0%
	Female	26.6%	73.4%	100.0%
	Female odds ratio		1.37	
Family origin indicator	Italian native	29.1%	70.9%	100.0%
	Foreigner I g.	43.7%	56.3%	100.0%
	Foreigner II g.	32.6%	67.4%	100.0%
Repeating student at lower secondary school	Not repeating	28.7%	71.3%	100.0%
	Repeating	61.0%	39.0%	100.0%
Desired qualification (10^{th} Grade Qst)	No Univ. degree	43.0%	57.0%	100.0%
	Univ. degree/more	15.4%	84.6%	100.0%
Geographical area	North-West	17.4%	82.6%	100.0%
	North-East	18.2%	81.8%	100.0%
	Centre	28.4%	71.6%	100.0%
	South	40.0%	60.0%	100.0%
	South-Island	44.7%	55.3%	100.0%
2014 8^{th} grade INVALSI Italian language test score	1° quartile	63.6%	36.4%	100.0%
	2° quartile	45.5%	54.5%	100.0%
	3° quartile	22.3%	77.7%	100.0%
	4° quartile	7.2%	92.8%	100.0%
ESCS	1° quartile	67.6%	32.4%	100.0%
	2° quartile	50.2%	49.8%	100.0%
	3° quartile	28.1%	71.9%	100.0%
	4° quartile	9.0%	91.0%	100.0%

If we consider the subset of those students (still without ever repeating a school years in the period under investigation) obtaining a competence level at least equal to “3” in the 13^{th} grade Italian language test (Tab. 5), there is a reduction of the gap from 10 (in the previous table) to 7 percentage points still in favour of females. In terms of odds ratio, the decrease is from 1.61 to 1.37, i.e. females are 37% more likely than males to reach a competence level at least of “3”. As underlined by the vast literature on the subject, females achieve sufficient marks to a greater extent than their male colleagues.

ISBN 9788835123330

Tab. 6 – Descriptive statistics of the predictors with respect to the dependent variable Y2 (not sufficient/sufficient competence level) – Mathematics

Predictors		*< “3”*	*Assigned competence level in 13th grade 2018-19 Math test*	
			>= “3”	*Total*
Gender	Male	28.6%	71.4%	100.0%
	Female	38.9%	61.1%	100.0%
	Female odds ratio		0.63	
Family origin indicator	Italian native	33.9%	66.1%	100.0%
	Foreigner I g.	42.3%	57.7%	100.0%
	Foreigner II g.	34.1%	65.9%	100.0%
Repeating student at lower secondary school	Not repeating	33.3%	66.7%	100.0%
	Repeating	60.2%	39.8%	100.0%
Desired qualification (10^{th} Grade Qst)	No Univ. degree	42.7%	57.3%	100.0%
	Univ. degree/more	21.3%	78.7%	100.0%
Geographical area	North-West	20.5%	79.5%	100.0%
	North-East	19.7%	80.3%	100.0%
	Centre	34.5%	65.5%	100.0%
	South	45.0%	55.0%	100.0%
	South-Island	51.9%	48.1%	100.0%
2014 8^{th} grade INVALSI Math test score	1° quartile	67.6%	32.4%	100.0%
	2° quartile	50.2%	49.8%	100.0%
	3° quartile	28.1%	71.9%	100.0%
	4° quartile	9.0%	91.0%	100.0%
ESCS	1° quartile	41.6%	58.4%	100.0%
	2° quartile	30.5%	69.5%	100.0%
	3° quartile	25.7%	74.3%	100.0%
	4° quartile	20.7%	79.3%	100.0%

The situation varies depending on the subject. Tab. 6 repeats the descriptive statistics commented on for the previous table but considering in this case the Math test. This subject sees a gap in favour of males: 10 percentage points more than females obtaining a result at least equal to sufficiency (competence level “3”). The result is in fact completely reversed, looking at the odds ratio: females are 37% less likely than males to reach a competence level at least of “3”.

ISBN 9788835123330

Tab. 7 – Descriptive statistics of the predictors with respect to the dependent variable Y3 (low performers/high performers) – Italian language

Predictors		*Assigned competence level in 13th grade 2018-19 Italian language test*		
< "4"		*>= "4"*	*Total*	
Gender	Male	61.6%	38.4%	100.0%
	Female	57.0%	43.0%	100.0%
	Female odds ratio		1.21	
Family origin indicator	Italian native	58.5%	41.5%	100.0%
	Foreigner I g.	74.7%	25.3%	100.0%
	Foreigner II g.	64.1%	35.9%	100.0%
Repeating student at lower secondary school	Not repeating	58.3%	41.7%	100.0%
	Repeating	85.5%	14.5%	100.0%
Desired qualification (10th Grade Qst)	No Univ. degree	75.1%	24.9%	100.0%
	Univ. degree/more	43.3%	56.7%	100.0%
Geographical area	North-West	46.0%	54.0%	100.0%
	North-East	47.0%	53.0%	100.0%
	Centre	59.1%	40.9%	100.0%
	South	70.2%	29.8%	100.0%
	South-Island	73.7%	26.3%	100.0%
2014 8th grade INVALSI Italian language test score	1° quartile	86.6%	13.4%	100.0%
	2° quartile	82.5%	17.5%	100.0%
	3° quartile	60.9%	39.1%	100.0%
	4° quartile	27.2%	72.8%	100.0%
ESCS	1° quartile	85.5%	14.5%	100.0%
	2° quartile	77.2%	22.8%	100.0%
	3° quartile	55.6%	44.4%	100.0%
	4° quartile	22.9%	77.1%	100.0%

If we analyse the distribution of students who reach a high level of competence (at least level "4"), for Italian language (Tab. 7) it can be seen that the positive gap in favour of females, which was 10 percentage points for Y1 and 7 for Y2, it decreases again to 5.4 percentage points for Y3. In other words, females are 21% more likely than males to reach a competence level at least of "4". As literature says about it, the positive gap in favour of females in linguistic subjects decrease when considering higher results.

ISBN 9788835123330

Tab. 8 – Descriptive statistics of the predictors with respect to the dependent variable Y3 (low performers/high performers) – Mathematics

Predictors		*Assigned competence level in 13th grade 2018-19 Math test*		
< "4"		*>= "4"*	*Total*	
Gender	Male	48.1%	51.9%	100.0%
	Female	62.1%	37.9%	100.0%
	Female odds ratio		0.57	
Family origin indicator	Italian native	55.3%	44.7%	100.0%
	Foreigner I g.	65.9%	34.1%	100.0%
	Foreigner II g.	56.7%	43.3%	100.0%
Repeating student at lower secondary school	Not repeating	54.9%	45.1%	100.0%
	Repeating	79.3%	20.7%	100.0%
Desired qualification (10th Grade Qst)	No Univ. degree	65.9%	34.1%	100.0%
	Univ. degree/more	41.5%	58.5%	100.0%
Geographical area	North-West	41.7%	58.3%	100.0%
	North-East	40.7%	59.3%	100.0%
	Centre	56.4%	43.6%	100.0%
	South	67.0%	33.0%	100.0%
	South-Island	73.1%	26.9%	100.0%
2014 8th grade INVALSI Math test score	1° quartile	85.5%	14.5%	100.0%
	2° quartile	77.2%	22.8%	100.0%
	3° quartile	55.6%	44.4%	100.0%
	4° quartile	22.9%	77.1%	100.0%
ESCS	1° quartile	64.0%	36.0%	100.0%
	2° quartile	52.8%	47.2%	100.0%
	3° quartile	47.0%	53.0%	100.0%
	4° quartile	40.3%	59.7%	100.0%

The situation again is reversed when we analyse the distribution of students who reach a high level of competence in the Math test (Tab. 8). While for the Italian language test the gap in favour of females tends to decrease considering the higher levels of competence, for Math the gap in favour of males increases to 14 percentage points (from 10 detected in model 2). In other words, females are 43% less likely than males to reach a competence level at least of "4".

In summary, it can be stated that, considering grade 8 as the starting point, females reach the 13th grade test after 5 years (thus without dropping out or repeating) to a greater extent than males (10 percentage points more, as showed in Tab. 4), but when we go to detail the phenomenon according to the

ISBN 9788835123330

levels of competence the results change. In fact, the gap in favour of females remains only for Italian, actually reducing if we consider the achievement of a competence level at least equal to “3” (‘sufficient’) or at least equal to “4” (respectively +7 and +5 percentage points, displayed in Tab. 5 and Tab. 7). On the other hand, males are less likely to reach grade 13 without dropping out or repeating, but then, contrary to what happens for Italian language with females, for Maths the gap in favour of males tends to increase as the level of competence obtained increases (+10 and +14 percentage points more respectively showed in Tab. 6 and Tab. 8).

These results obtained by descriptive statistics, will be now checked by controlling for the predictors included in the logistic regressions.

On the basis of the tables further on above, we proceed to an examination of the three models proposed by considering first of all the pseudo-R^2: the first model reports unsatisfactory values of the pseudo-R^2, lower than 0.3 (Tab. 9), considered a threshold of acceptability of these measures in logistic regression. The fact that all the coefficients are significant is not necessarily an indication of a good model since, as already mentioned, for very large “n” statistics based on the chi-square tend to always be significant. The little strength (or “effect size”) of this model is also represented by the odds-ratio values: many of them deviate little from “1” (no influence) and, in general, are of lesser extent than those found in the following models with other response variables.

The two subsequent models, in fact, have satisfactory pseudo-R^2 values, as they are around or higher than 0.3. Moreover, in these models the coefficients are all significant, with the exception of the pupil’s origin indicator, in fact this predictor is not significant in both models of Mathematics, and also for Italian it does not appear to have much effect, with odds ratio close to 1: it could therefore be eliminated from the model without major losses in representativeness. In these 2 models, several coefficients present odds-ratios very different from “1”, a signal of influence on the response variable.

Although the approach of this work is not of a predictive type, the classification tables of the three models (Tab. 10) were considered only to add some reflection on the difference between the models: the first model, if used with a predictive approach, would classify only about 5% of pupils who fail to complete the 13th grade test correctly, raising doubts about the reliability of the response variable thus calculated. The other two models would work much better in a predictive approach, in particular model 3, which correctly classifies 3 out of 4 pupils in general, well distributed among top-performers and lower levels.

ISBN 9788835123330

Tab. 9 – Model summary of the three logistic regression models

	Model summary		
Model 1	-2 Log likelihood	213,868.4	
	Cox & Snell	0.08	
	R-square	0.14	
		Italian language	*Mathematics*
Model 2	-2 Log likelihood	203,256.5	222,533.1
	Cox & Snell	0.26	0.26
	R-square	0.38	0.36
Model 3	-2 Log likelihood	249,878.1	253,052.2
	Cox & Snell	0.29	0.29
	R-square	0.39	0.39

Note: ** p-value < 0.01.

Tab. 10 – Classification tables of the three logistic regression models

Model proposed		*% correct*	
Italian language		*Math*	
Model 1 (Y1)	Reaches 13th grade exam without repeating	99.3%	
	Reaches 13th grade exam repeating or drop out	4.7%	
	Average	85.4%	
Model 2 (Y2)	Competence level 1, 2	46.4%	48.8%
	Competence level 3, 4, 5	91.9%	89.2%
	Average	80.3%	77.4%
Model 3 (Y3)	Competence level 1, 2, 3	79.4%	67.8%
	Competence level 4, 5	69.5%	84.5%
	Average	75.0%	77.8%

Division value: 0.5.

At this point, we can move on to the interpretation of the coefficients obtained by considering, in addition to the odds ratios. As regards model 1 (Tab. 11), considering the odds ratio column (exp (B)) for the gender, which is 1.63 (underlined in the table), we can say that the possibility that a female, compared to a male, took the 13th grade test, is 63% more likely than a male student. Another relevant issue that emerges from the analyses is the geographical distribution of students. In particular, it is known that a student from the South or South and Islands, compared to a student from the Centre, has respectively 17 and 12 percent more opportunities to participate in the 2019 grade 13 test, thus not repeating any year in the period considered. On the other hand, a North-East or North-West student, compared to a student

ISBN 9788835123330

from the Centre, has respectively 9 and 18 percent less chance to participate in the Grade 13 test. The major hypothesis around the latter evidences is that southern regions could be less strict with their students, so that they're more likely to reach grade 13 without being rejected.

Tab. 11 – Model 1. Logistic regression coefficients with Y1 outcome variable: 13th grade final exam reached without repeating

Variable	*Categories*	*B*	*s.e. (B)*	*Exp(B)= odds ratio*	*π*
Constant		0.66**	0.02	1.94	0.66
Geographical area	North-West	-0.20**	0.02	0.82	0.61
	North-East	-0.09**	0.02	0.92	0.64
	Centre			1.00	
	South	0.16**	0.02	1.17	0.69
	South-Islands	0.11**	0.02	1.12	0.68
2014 8th grade INVALSI test score	1° quartile			1.00	
	2° quartile	0.22**	0.02	1.25	0.71
	3° quartile	0.59**	0.02	1.80	0.78
	4° quartile	1.14**	0.02	3.12	0.86
ESCS	1° quartile			1.00	
	2° quartile	0.19**	0.02	1.21	0.70
	3° quartile	0.28**	0.02	1.32	0.72
	4° quartile	0.29**	0.02	1.33	0.72
Family origin indicator	Italian native			1.00	
	Foreigner I g.	-0.42**	0.03	0.66	0.56
	Foreigner II g.	-0.57**	0.02	0.57	0.52
Desired qualification (10th Grade Qst)	No Univ. degree			1.00	
	Univ. degree/more	0.69**	0.01	1.98	0.79
Repeating student at lower secondary school	Not repeating			1.00	
	Repeating	-0.85**	0.02	0.43	0.45
Gender	Male			1.00	
	Female	0.49**	0.01	1.63	0.76

Turning to Tab. 12, as a "success" event it is considered the achievement of a competence level at least equal to "3" (out of 5), considered as "sufficient". In this case, net of the other variables of the model, the possibility that a sufficient competence level was obtained from a female, compared to a male is only 10% higher at the Italian language test, while 46% less in Mathematics test (odds ratio 1.11 and 0.54 underlined in the table). The gender gap still changes in favour of males in the third model (Tab. 13)

ISBN 9788835123330

in which the dependent variable discriminates between top performers and others, i.e. takes on a value of 1 if the student reaches the test in grade 13 obtaining a level of competence at least equal to “4” (out of 5). The positive gap in favour of females in Italian language is disappeared, while the one in favour of males in Mathematics increase (female odds ratio is 0.97 and 0.44 respectively).

There are some other results involving the other predictors of the models. For the geographical area, southern pupils seem to have more probability than northern to reach grade 13 without being rejected, probably an evidence of a lesser tendency in southern schools to reject pupils (model 1 in Tab. 11). Then, looking at models 2 and 3, even net of the other predictors considered, the well-known highest relationship of the regions of Northern Italy with the best results is confirmed, but the effect slightly drops when we consider the top performers (Tab. 12 and Tab. 13). More in details; looking at model 2, northern pupil, controlling with all other variables, has twice the chance of obtaining a sufficient level of competence compared to a pupil of the Centre and about triple that of one in Southern Italy. This proportion decrease when we consider a competence level at least ok “4”.

Other general interesting results regards pupils’ origin and background. After having purified the effect of the other variables, pupil’s origin has significative odds-ratios for the 1st model, while for the 2nd and 3rd model they’re significative only related to Italian language. That is, compared to native students, I and II generation foreigners are on average 34% and 43% less likely to reach grade 13 without being rejected (from grade 10 grade 13) but then, if the career is regular, they have the same probability to reach a good competence level in Mathematics, and they’re only slightly less likely to get a good competence level in Italian language (in particular for the II generation students, their probability of getting a “top” competence level in Italian language is only 14% less than native ones). The regularity of the course of study (in lower secondary school) shows an odds-ratio of repeating students which is half that of regular students, and constitutes an important variable in the model as it allows us to evaluate the other predictors net of any repetition of school year in middle school (an event generally linked to very complex situations). The family background, has pretty similar odds-ratios in both models, so the effect doesn’t change considering a threshold competence level of “3” or “4”; its values, although significant, are not the strongest ’effect’ that we usually find in in other analyses related to competences. It should also be emphasized, again with reference to the ESCS, that its influence in both models is greater in Italian language than in Mathematics.

ISBN 9788835123330

Tab. 12 – Model 2. Logistic regression coefficients with Y2 outcome variable: competence level at least 3 out of 5 (sufficient level)

Variable	*Categories*	*Ita.*	*Mat.*	*Ita.*	*Mat*	*Ita.*	*Mat.*	*Ita.*	*Mat.*
		B		*S.E.*		*Exp(B)= oddsratio*		*π*	
Constant		-1.098**	-0.808**	0.02	0.02	0.33	0.45	0.25	0.31
Geographical area	North-West	0.627**	0.668**	0.02	0.02	1.87	1.95	0.38	0.47
	North-East	0.607**	0.725**	0.02	0.02	1.84	2.07	0.38	0.48
	Centre					1.00	1.00		
	South	-0.459**	-0.353**	0.02	0.02	0.63	0.70	0.17	0.24
	South-Islands	-0.694**	-0.650**	0.02	0.02	0.50	0.52	0.14	0.19
2014 8th grade INVALSI test score	1° quartile					1.00	1.00		
	2° quartile	0.638**	0.644**	0.02	0.02	1.89	1.90	0.39	0.46
	3° quartile	1.630**	1.507**	0.02	0.02	5.10	4.51	0.63	0.67
	4° quartile	2.784**	2.751**	0.02	0.02	16.20	15.66	0.84	0.87
ESCS	1° quartile					1.00	1.00		
	2° quartile	0.203**	0.174**	0.02	0.01	1.22	1.19	0.29	0.35
	3° quartile	0.310**	0.280**	0.02	0.02	1.36	1.32	0.31	0.37
	4° quartile	0.531**	0.381**	0.02	0.02	1.70	1.46	0.36	0.39
Family origin indicator	Italian native					1.00	1.00		
	Foreigner I g.	-0.159**	0.042	0.03	0.03	0.85	1.04	0.22	
	Foreigner II g.	-0.128**	-0.009	0.03	0.03	0.88	0.99	0.23	
Desired qualification (10th Grade Qst)	No Univ. degree					1.00	1.00		
	Univ. degree or more	1.029**	0.824**	0.01	0.01	2.80	2.28	0.48	0.50
Repeating student at lower secondary school	Not repeating					1.00	1.00	0.25	
	Repeating*	-0.791**	-0.772**	0.03	0.03	0.45	0.46	0.13	0.17
Gender	Male					1.00	1.00		
	Female	0.104**	-0.624**	0.01	0.01	1.11	0.54	0.27	0.19

ISBN 9788835123330

Tab. 13 – Model 3. Logistic regression coefficients with Y2 outcome variable: competence level at least 4 out of 5 (top performers)

Variable	*Categories*	*Ita.*	*Mat.*	*Ita.*	*Mat.*	*Ita.*	*Mat.*	*Ita.*	*Mat.*
		B		*S.E.*		*Exp(B)= odds ratio*		π	
Constant		-2.465**	-1.843**	0.02	0.02	0.09	0.16	0.08	0.14
Geographical area	North-West	0.547**	0.556**	0.02	0.02	1.73	1.74	0.13	0.22
	North-East	0.544**	0.605**	0.02	0.02	1.72	1.83	0.13	0.22
	Centre					1.00	1.00		
	South	-0.405**	-0.356**	0.02	0.02	0.67	0.70	0.05	0.10
	South-Islands	-0.588**	-0.642**	0.02	0.02	0.55	0.53	0.05	0.08
2014 8th grade INVALSI test score	1° quartile					1.00	1.00		
	2° quartile	0.112**	0.428**	0.02	0.02	1.12	1.53	0.09	0.20
	3° quartile	1.112**	1.313**	0.02	0.02	3.04	3.72	0.21	0.37
	4° quartile	2.444**	2.650**	0.02	0.02	11.52	14.15	0.49	0.69
ESCS	1° quartile					1.00	1.00		
	2° quartile	0.222**	0.167**	0.02	0.01	1.25	1.18	0.10	0.16
	3° quartile	0.349**	0.268**	0.02	0.02	1.42	1.31	0.11	0.17
	4° quartile	0.587**	0.384**	0.02	0.02	1.80	1.47	0.13	0.19
Family origin indicator	Italian native					1.00	1.00		
	Foreigner I g.	-0.238**	-0.013	0.04	0.03	0.79	0.99	0.06	0.14
	Foreigner II g.	-0.153**	0.006	0.03	0.03	0.86	1.01	0.07	0.14
Desired qualification (10th Grade Qst)	No Univ. degree					1.00	1.00		
	Univ. degree/more	0.972**	0.832**	0.01	0.01	2.64	2.30	0.18	0.27
Repeating student at lower secondary school	Not repeating					1.00	1.00	0.08	
	Repeating*	-0.699**	-0.722**	0.04	0.04	0.50	0.49	0.04	0.07
Gender	Male					1.00	1.00		
	Female	-0.026**	-0.734**	0.01	0.01	0.97	0.48	0.08	0.07

ISBN 9788835123330

The most influential variables of model 2 and 3, that is, with the highest odds-ratios, even more than the ESCS, are the "desired qualification" expressed in grade 10 tests, the 8th grade test score before entering the high school and the dummy indicating a repeating student at lower secondary school (as indeed it was emerged in Tab. 2 descriptive statistics with respect to gender): their effect could be observed by the odds ratios or looking at the Beta coefficients which are the biggest. A student who declares at grade 10 test that he/she wants to achieve at least a tertiary degree has more than twice the chance of achieving a level of competence in Maths at least sufficient or among the best compared to those declaring a lower qualification (odds-ratio equal to 2.3 in both model 1 and model 2), even almost triple if we consider Italian language (odds-ratios 2.8 and 2.6 respectively). The 8th grade test score, on the other hand, for model 2, confirms that a pupil with a score in the second quartile has almost double the chances of achieving a sufficient level of competence compared to pupils with a score in the first quartile (odds-ratio 1.8 for both Italian language and Mathematics); this effect stands but it is much lower considering model 3 (odds-ratios 1.1 and 1.5 respectively). Considering the subsequent quartiles, the values are even much higher. It is a sign that a good score in the 8th grade test is an excellent indicator of the level of competence that a pupil can achieve after 5 years (not repeating any school year), and this result is also interesting if we keep in mind the different construction of the INVALSI tests between the two administrations considered: the grade 8 test of 2014 *was paper & pencil* and built differently than that of grade 13 of 2019, actually the second year of the CBT (Computer Based Test) technique. Last of the three main effect is having repeated a year in the lower secondary school, which has similar odds in both model 2 and 3 and both subjects (Italian or Maths).

5. Conclusions and discussion

The important conclusion that emerges from this work with respect to gender is the following: starting from the descriptive statistics that confirm the best performance of females in Italian language and of males in Maths, when we get into the model to check the effect controlling for other variables, the higher effect of females in Italian language almost disappears, while that in favour of males in Mathematics remains strong: after controlling for the other 6 variables of the model, the probability for a male student to achieve a level of competence in Math at least sufficient or at least equal to "4" is double that of females. The odds ratio reported in the crosstabs and in the models are here synthesized in Tab. 14.

ISBN 9788835123330

Tab. 14 – Comparison between crosstabs and logistic models odds ratios

Dependent variable	*Odds ratio*		*Conclusions from the models*
	Crosstabs	*Models*	
Y1	1.61	1.63	Positive gap in favour of females
Y2 ITA	1.37	1.11	Decrease to a slight positive gap in favour of females
Y2 MAT	0.63	0.54	Increase of the positive gap in favour of males
Y3 ITA	1.21	0.97	Decrease to a no difference between males and females
Y3 MAT	0.57	0.48	Increase of the positive gap in favour of males

As an interpretation of these results, when we considered Y1, females are at least 60% more likely than males to reach from grade 8 the grade 13 without repeating or dropping out. When we consider only the subset of students reaching grade 13 in 5 years, females are 37% more likely to get a competence level “3” in Italian language and 21% more likely to get a level “4”. On the other hand, they are 37% less likely to get a competence level “3” in Mathematics and 43% less likely to get a level “4” (“crosstabs” column in the table). These values change while we obtain odds ratios from logistic regression net of all the other predictors included: for Italian language the previous the 37% more likely decrease becomes an 11% and the 21% becomes a 3% less likely, while for Mathematics the gap increase in favour of males.

There’re some other results emerged about the other predictors. About the geographical area, students from the south are more likely to reach the 13th grade without repeating or dropping out; but when we consider only those reaching the last grade regularly in five years, the highest competence of students from the north emerged, even net of the other predictors included in the models.

Students with a foreign family origin are less likely to reach grade 13 without being rejected or dropped out but then, if we consider only those with a regular career, they have the same probability to reach a sufficient or high competence level in Mathematics, and they’re only slightly less likely to get a sufficient/high competence level in Italian language.

The ESCS indicator has a slight positive effect in reaching the 13th grade with a regular career of five years. Its effect, despite higher when we consider as a response the competence level, is not strong as expected and anyway lower than other predictors with the highest coefficients, which are the “desired qualification” expressed in grade 10 tests, the 8th grade test score before entering the high school and the dummy indicating a repeating student at lower secondary school.

The desired qualification express at the 10th grade questionnaire, has a really high coefficient for those students declaring a tertiary degree: the asso-

ISBN 9788835123330

ciation is stronger in particular with the Italian language competence level. Finally, the 8th grade score test is the predictor showing the strongest association with the competence level at the 13th grade test (which was in a CBT format): it's also a proof of the reliability of the test INVALSI carried out since many years. Last, the variable indicating a repeating student at lower secondary school, involves only a little part of the students and has a strong negative effect.

Finally, it is interesting to consider that the three aforementioned variables that have the greatest effect in the three models are aspects related to the performance or aspirations of the pupils, while all the others are personal "attributes" (gender), or of the family context (ESCS and origin) and territorial (Geographical area). These most influential variables in the model are, however, the most differentiated with respect to gender, so we can presume that, at least for Italian language, the gender gap is reduced because a large part of this variability is absorbed by the presence in the model of the "desired educational qualification", the "8th grade test score" and the dummy "repeating student at lower secondary school", three variables that are highly characterized by gender, as described at the bottom of Tab. 2. On the contrary, in Math, this effect of cancelling the gap is not noticed: males, net of the score in the 8th grade test and their conviction on the qualification to be achieved, still reach levels of better grade competence (the odds ratio of females in Mathematics is in fact equal to 0.5 in both models 2 and 3).

These results are to be considered with some considerations and warnings. First of all, the validity of the dependent variables while Y2 and Y3 are a precise measure of the competence level, that is over or under an exact competence level in the Italian and Mathematics test, Y1 is more ambiguous, because the "not success" option (i.e. grade 13 test not reached) includes different situations (rejected students, drop out, private and vocation training students) which can influence data. That could be the reason why for the model 1 the R-square is not sufficient and the percentage of correct classification is very poor for those students not reaching the 13th grade exam without repeating (as showed in Tab. 9 and Tab. 10).

Another limit of these results is that the three models (as showed in Tab. 1) they are based on a portion of the 8th grade students of 2013-14, respectively 55,1% for model 1 and 46,9% for models 2 and 3, and that's a warning to be always reported when showing these results.

Further possible developments of this work may be the following:

– in the case of models 2 and 3, since the test scores are continuous variables, we could consider the Probit model, which requires the normality of the distribution of the dependent variable, and which therefore would

ISBN 9788835123330

consider the whole range of response values, taking into account more information detailed with respect to the dichotomous variables considered here; it would also be possible to insert ESCS and scores on the test of G08 as continuous explanatory variables;
- again in the case of models 2 and 3, considering the levels of competence (from 1 to 5, therefore ordinal and not dichotomous), a "variable proportional odd model for ordinal response" could be set, thus exploiting the entire scale of competences for each student;
- for a better analysis on the possibility of the 8^{th} grade 2014 students' cohort to finish regularly the high school carrier (i.e. the aim of model 1), one could try to identify the individual reasons for mismatch and create a cleaner dichotomous dependent variable, excluding vocational training and other confounding cases;
- to separate the analysis according to the school type: high schools, technical institutes and vocational institutes;
- to repeat the same models using the new data available after the 2021 INVALSI tests.

References

Hosmer D.W., Lemeshow S. (2000), *Applied Logistic Regression*, Wiley, New York.

Tabachnick B.G., Fidell L.S. (2007), *Using Multivariate Statistics*, Pearson, Boston, 7^{th} ed.

Cicerone P.E. (2004), "Educati a parte. Mente e cervello", *La Rivista di Psicologia e Neuroscienze*, 11, pp. 36-41.

Davies P.G., Spencer S.J. (2005), "The Gender-Gap Artifact: Women's Underperformance in Quantitative Domains Through the Lens of Stereotype Threat", in A.M. Gallagher, J.C. Kaufman (*eds.*), *Gender differences in mathematics: An integrative psychological approach*, Cambridge University Press, Cambridge, pp. 172-188.

Else-Quest N.M., Hyde J.S., Linn M.C. (2010), "Cross-national patterns of gender differences in mathematics: a meta-analysis", *Psychological Bulletin*, 136, pp. 103-127.

Bleeker M.M., Jacobs J.E. (2004), "Achievement in math and science: Do mothers' beliefs matter 12 years later?", *Journal of Educational Psychology*, 96, pp. 97-109.

INVALSI (2016a), *Indagini IEA 2015 TIMSS IV anno di scolarità: sintesi dei risultati degli studenti italiani in matematica e scienze*, INVALSI, Roma.

INVALSI (2016b), *Indagini IEA 2015 TIMSS VIII anno di scolarità: sintesi dei risultati degli studenti italiani in matematica e scienze*, INVALSI, Roma.

ISBN 9788835123330

INVALSI (2016c), *Indagine Ocse Pisa 2015. I risultati degli studenti italiani in scienze, matematica e Lettura*, INVALSI, Roma.

INVALSI (2017), *Indagine IEA 2016 PIRLS. Rapporto Nazionale*, INVALSI, Roma, retrieved on Dicember, 30, 2020, from https://stats.idre.ucla.edu/other/mult-pkg/faq/general/faq-what-are-pseudo-r-squareds/.

Poliandri D., Cardone M., Muzzioli P., Romiti S. (2011), "A rating scale model for a scale of text anxiety in Italy", *INVALSI working paper*, 11/2011, retrieved on Dicember, 30, 2020, from https://papers.ssrn.com/sol3/papers.cfm?abstract_id=1856110.

Steele C.M. (1997), "A Treat in the Air. How Stereotypes Shape Intellectual Identity and Performance", *American Psychologist*, 52, pp. 613-629.

Sternberg R.J., Williams W.M. (1997), "Does the Graduate Record Examination predict Meaningful Success in the Graduate Training of Psychologist? A Case Study", *American Psychologist*, 52, pp. 630-641.

ISBN 9788835123330

4. Math gender gap according to socio-economic background in Italy: the better the conditions the larger the gap?

by Patrizia Giannantoni, Veronica Pastori, Cecilia Bagnarol

Most of the studies on disadvantaged categories and their educational achievements focus the attention on students' gender and socio-economic background taken separately. In line with some researchers (Tinklin *et al.*, 2003; Strand, 2014) our aim is to consider these two sources of inequality together.

In particular, the research questions are: 1) does gender gap increase in low SES (socio-economic status) families and decrease in more favourable social context? 2) if it is so, is this effect constant over time? 3) how much does the family economic and educational status worsen the basic gender gap?

We used INVALSI data of Mathematics for students in 5^{th} grade of the Italian school system.

In the first part of the paper, we analysed the gender gap over time considering the last ten school years (from 2009/2010 to 2018/2019, the last available) and focus our attention on SES, mothers' education, cognitive "Parts" of INVALSI test (partitions of the whole test pertaining to the same cognitive domain: Numbers, Statistics, Shapes and Figures, Relations and Functions) and geographical macro-area where the schools are located.

The second part of our study was centred on 2017/2018 data with the aim to deepen the study of gender gap considering other variables that could influence achievement, as nursery school frequency, anxiety before test and Math score in 2^{nd} grade.

The main result, observed in descriptive analysis and corroborated in nested regression models, can be summarized by the expression "the better the conditions the larger the gap". In fact, it seems that the gender gap is wider where the conditions are more favourable (high socio-economic level, Norther regions in Italy, nursery school frequency) while in disadvantaged conditions the gap between girls and boys decreases.

ISBN 9788835123330

1. Theoretical framework

Educational studies have extensively documented the existence of several factors that continue to determine persistent inequalities in school performances and learning trajectories. There is a large chunk of disadvantaged pupils that do not perform academically as well as their advantaged peers (Reardon, 2011). The term "disadvantaged" includes many aspects re-calling a variety of situations; disadvantage factors studied by social scientists are: socio-economic status (parental income, educational level and family resources available), parental involvement in academic achievement, teacher's expectations, living area and neighbourhood, ethnicity minority status, speaking school language as a second language, immigrant status, perceived discrimination, self-concept, motivation (Banerjee, 2016).

A consistent part of these studies focused the attention on another important characteristic: student's gender. Together with the other factors named above, some researchers (Fryer and Levitt, 2010) have highlighted disadvantages between boys and girls at the end of primary school: in particular, in Reading for boys and in Mathematics for girls. These inequalities are consistent across countries and over time (EACEA P9 Eurydice, 2010).

The literature on the subject have proposed many explanations about gender gap in achievement.

A first strand is oriented to biological factors, which cannot be affected by societal and cultural factors (Naour, 2001). Another group of explanations emphasizes psychological and societal factors. Boys and girls grow up in structured sets of beliefs about personal attributes that characterized them: for example, Math is not important, useful, doable, or part of the identity of a girl (Wilder and Powell, 1989). The acquisition of so-called "gender-appropriate" preferences, skills, personality, attributes, behaviours and self-concepts depend on the process of gender typing (Kollmayer, Schober and Spiel, 2016). Together with social and cultural factors, attitude, self-confidence, anxiety, self-esteem, motivation, interest in matter, stereotypes and the expectations of parents and teachers (Lubienski *et al.*, 2013) are also widely considered as potentially important in determining school results, especially in the perspective of gender gap.

The gender attitudes and expectations stem not only from society as a whole, but especially from the most important socializing agent: parents. Parents influence gender schemas and gender-stereotyped cognitions, but they act also as models. In this way, they, with their examples, can encourage to do better.

Parent's expectations are strictly linked with family resources. Socio-economic background is probably the most important agent in determining stu-

ISBN 9788835123330

dents' achievement at school. However, when it comes to gender differences, the role of familiar cultural and economic resources is still unclear. Some studies (Entwisle, Alexander and Highes, 2007) showed how cultural and economic context can have an effect both in terms of stronger parental support and expectations, and in terms of less gender stereotypes. Other research found out an opposite, counter-intuitive effect, of economic and cultural resources on gender gap, showing that inequalities increase as the SES level improves (Contini, Di Tommaso and Mendolia, 2017; Lubienski *et al.*, 2013). A deeper analysis of these aspects is needed especially in Italy, one of the OECD countries displaying the largest differentials between boys and girls, according to PISA assessment (OECD, 2015).

Beyond social justice, the study of gender together with socio-economic background, is important if we think about cumulative advantage. In literature, this term is used to indicate that an advantage of one individual or group over another accumulates over time (DiPrete and Eirich, 2006): a cumulative advantage process can «magnifying small differences over time and makes it difficult for an individual or group that is behind at a point in time in educational development, income or other measures to catch up» (272).

Talking about inequalities, several scholars and researchers (Tinklin *et al.*, 2003; Strand, 2014) argue about the need to consider not only one source of inequality because it doesn't allow to have a complete vision about the phenomenon. Low performances do not associate with a single student's characteristics: it is rather the combination and accumulation of factors and experiences lived by students (OECD, 2016).

Our paper contributes to the existing literature showing over the entire Italian population of 5th grade and in a time interval of ten years the interaction of socio-economic background and gender in determining Math achievement. Furthermore, this research takes into account other external "beneficial" factors that can affect Math performance, observing a general phenomenon that we synthetized as "the better the conditions, the larger the gender gap".

2. Design of research

2.1. Objectives and hypotheses

Starting from the importance to analyse more than one source of inequality, our research was directed to explore differences in gender gap in 5th grade of primary school in Italy, according to the different socio-economic background of students.

ISBN 9788835123330

We wanted to test whether the gap between boys and girls could be larger for students coming from low SES, as they experience a lack of home support and stronger cultural influences reinforcing gender stereotypes or, on the contrary, it could be larger for students with a high SES for reasons that literature has not clearly identified yet.

Furthermore, we observed the trend of these differences over a long-time interval, the decade from 2009/2010 to 2018/2019, to detect whether the trend is directed towards a reduction of these gaps or rather to an increase.

We finally aimed at gaining a better understanding of the effect of family resources in moderating the disadvantage of gender.

Research questions we intended to answer are:

- does gender gap increase in low SES families and decrease in more favourable social contexts?
- if it is so, is this effect constant over time?
- how much does the family economic and educational status worsen the basic gender gap?

2.2. Data

We used INVALSI population data in Math test (both percentage and WLE[1] scores) for students in 5th grade of Italian school system in order to estimate the gap in performance between girls and boys. INVALSI database of standardized tests for 5th grade includes always information about socio-economic status (SES) of students based on ESCS index[2] and mother's educational attainment.

We performed a classification of students according to SES status based on two indicators: quartiles of index ESCS and alternatively mother's educational attainment grouped into three categories: lower secondary/high school/ university or higher.

With the aim to identifying whether there is a trend over the time, in the first part of analysis, we covered the period from academic year 2009/2010 to 2018/2019, while in the second part we focused our attention on the year 2017/2018 because it is the year with the largest differences between girls and boys in Mathematical results.

[1] Percentage score corresponds to the percentage of correct answers over total items of the tests; WLE score is built on the basis of Rasch's Model taking into account students' ability and difficulty of items.

[2] The index is built considering educational level and occupational condition of parents and some family's resources available (Campodifiori *et al.*, 2010).

ISBN 9788835123330

Percentage scores are available over the whole decade, whereas WLE scores have been calculated by INVALSI since school year 2011/2012. Thus, the temporal trend has been observed with a different starting year for the two indicators. Both scores have been corrected for cheating[3].

For each school year in the analysis we created an "ad hoc" dataset with specific data cleaning procedure composed of these steps:

– exclusion of all students with missing gender variable;
– exclusion of students with missing scores (either percentage or WLE);
– exclusion of students with missing ESCS or mother's educational attainment variables.

Eventually, in order to create a homogeneous reference population, we selected only native[4] students with a regular course of studying.

Tab. 1 reports percentage reduction of reference population after data-cleaning procedure, which is always about 30%[5] of original population. However, giving the census nature of the data the overall size of population after cleaning procedure is always large enough for statistical analysis.

Tab. 1 – Total population and percentage reduction after data-cleaning procedure for different years of interest

Year of interest	*Before data-cleaning*	*After data-cleaning*	*Lost of information (%)*
2009/2010	498,382	333,127	33
2010/2011	515,104	336,334	35
2011/2012	489,581	297,565	39
2012/2013	483,921	323,410	33
2013/2014	477,944	331,612	31
2014/2015	412,743	286,642	30
2015/2016	482,761	354,151	27
2016/2017	489,343	346,332	29
2017/2018	499,353	355,336	29
2018/2019	489,189	344,377	30

[3] *Cheating*, defined as set of anomalies that alter the test results, has been identified according to the procedure of Quintano, Castellano and Longobardi (2009). The procedure is a multistage method, combining the factorial analysis with a fuzzy clustering approach. This procedure estimates the cheating considering the following items: percentage of correct answers, variability within the class, index of answers' homogeneity and class non-response rate.

[4] Native students are defined as: born in Italy from parents born in Italy.

[5] The high percentage for the school year 2011/2012 is due a high rate of invalid or missing responses related to the mother's education.

ISBN 9788835123330

In the year 2017/2018, where we performed a more in-depth analysis of gender gap, we used also a geographical variable: region and province where school attended by student is located. In addition, only for 2017/2018 we matched Mathematical test data with INVALSI Student's questionnaire data. This questionnaire collects information about students and their family background together with questions about approach to school and study (e.g. feeling anxiety before Math test, self- confidence in learning skills) as well as parental support.

In the development of this work we considered the following variables taken from INVALSI Student's questionnaire: attending nursery school (yes/no) and feeling anxiety before Math test (A lot/Enough/Few/ Not at all).

Finally, as INVALSI data have a longitudinal key (variable: SIDI INVALSI) to follow the students over time, the final dataset was also linked with previous scores of the same students in INVALSI Math test three years before, in 2nd grade of school in the year 2014/2015. This kind of information allowed us to take into account the "baseline Math skills" of the students, in order to perform a more accurate analysis comparing students with the same "starting ability" in Math. Including this information in the analysis means that we could look more specifically to what operates on gender gap in the period between grade 2nd and 5th, net from the gap already existent at the beginning of primary school.

2.3. Methodological choices

The attention on 5th grade students was motivated by the presence of differences in gender-oriented self-concepts and attitudes since primary school (Becker and McElvany, 2018). At the beginning of primary school, the differences between boys and girls would seem not to emerge, but at the end the gap is more evident. These results appear moderated by socio-economic status, in particular into low status group (Entwisle, Alexander and Olson, 2007).

In this way, the situation in 5th grade can became important considering the cumulative advantage process above cited. In fact, the results can constitute a red flag for policy makers interested in contrasting educational inequalities.

As indicator of socio-economic background we considered the index ESCS, but also one of its components, the variable mother's educational level, which has been shown as a very important factors explaining inequalities in school attainment for children (Checchi, Fiorio and Leonardi, 2007).

ISBN 9788835123330

In the last sixty years, the relationship between parents' educational level and student's achievement remained stable and as a result of a strong correlation between the first one and family income (Reardon, 2011), the choice to consider also educational level as an alternative to the complex index are justified.

The reason of our focus also on mother's educational attainment (separately from the family one) is motivated by the fact that mother is yet the main caregiver in Italy.

Mothers' education is positively associated with some personality's characteristics that emerge as significant factors of advantage in achievements, as openness and extroversion. In addition, these variable influences expectations of children's educational attainment at the beginning of school and this may contribute to a virtuous circle at later stages (Parveen and Alam, 2008).

Several studies highlight that the academic performances of students whose mothers have higher educational level is better than the students whose mothers are less educated or uneducated and that this trend increases with the increasing of mother's educational qualification (Englund *et al.*, 2004). Harding, Morris and Highes (2015) compare mother's educational attainment to a "black box" because it's useful to understand the complexity of the statistical positive relationship with achievements. Educational level is associated to the income that, in turn, guarantees several resources (in terms of cultural, social and economic availabilities).

2.4. Methods

This research followed different steps of analysis.

Gender gap in Math was always calculated as difference between girls and boys average scores both using percentage and WLE scores of INVALSI test.

In the first part we outlined temporal trend of gender gap in Math over the decade from 2009/2010 to 2018/2019.

Temporal trend was then observed according to a stratification of students in groups based on their socio-economic status. Stratification was created by means of two indicators: ESCS index and mother's educational attainment. Category of ESCS were divided into quartiles (< 25%, 25-50%, 50-75%, > 75%) whereas mother's educational attainment was categorized as follows: lower secondary/high school/university or more.

As we found a persistent increase in gender gap as ESCS increases, which was counter-intuitive with respect to our hypothesis, we decided to develop the research in different directions to shed more light on this result.

ISBN 9788835123330

The results from temporal trend by socio-economic conditions seemed to indicate that consistently over time the better the socio-economic resources the larger the girls' disadvantage. This result can be the effect of either:

- cultural factors that operate in a opposite direction with respect to what initially expected (e.g. the influence of the cultural stereotype that sees girls "not suited" for mathematics could be stronger at high levels of ESCS, where higher results can be reached but with the requirement of more motivation and self-esteem;
- or that biological/cognitive factors (i.e. cognitive strategies and process and mental approach between boys and girls) have an effect, that become more and more evident as the student's environment conditions improve.

We decided to deep the analyses into different directions considering the following aspects: "Parts" of Math test (partitions of the whole test pertaining to the same cognitive domain), geographical areas (region and province) and some students' characteristics.

The study of gender gap by Parts is intended to answer to this research hypothesis: if cultural factors play the major role in determining gender gaps, then different Parts of Math test should all show the same magnitude in gender gap, because cultural factors operate indistinctly on all the mathematical skills.

On the contrary if different cognitive strategies between boys and girls have the highest influence on gender gap, this will be reflected in varying levels of gender gap according to different Parts that activate different cognitive domains.

The temporal trend of gender gap by Parts was built up with difference of mean scores between girls and boys, calculated and plotted separately for:

- Numbers;
- Statistics;
- Shapes and Figures;
- Relations and Functions.

Parts were included in INVALSI Math scores since school year 2015/2016.

The study of geographical trend was projected with the aim to test the following hypothesis: if it is true that biological/cognitive differences exist and enhance in context with more favourable conditions (higher mean scores), we should observe a reversed trend in gender gap also at territorial level (higher differences in more favourable conditions, such as North East, lower gender gap in South).

Geographical analysis was built up estimating gender differences in mean scores at provincial and regional levels, represented in a chart with different intensity or colouring a different magnitude of gender gap using Tableau software (v. 2020.1).

ISBN 9788835123330

Nested regression models allowed us to verify the hypothesis of "the better the conditions the larger the gap": girls' disadvantage in Math performance increases in the conditions of highest advantage, for example among students of well-educated groups in northern regions.

The models referred only to year 2017/2018, because it is the year where the maximum gender gap was observed.

The outcome variable for all the models was WLE scores. We selected WLE scores because they consider jointly the ability of the student and the difficulty of the items, thus they represent a more sophisticated measure of Math skills.

Gender is the first covariate in the model. Successively, by means of nested models, we introduced all the other covariates one by one, studying both the effect of the new variable variation of the coefficient of all the other variables already included in the model.

Furthermore, introducing both primary (simple) and secondary (interaction with gender) effects of all the confounders we could distinguish the positive direct effect of the covariates (e.g. higher ESCS higher score) from the lowering or even the reverse effect of these same variables when they refer to girls (positive effect of ESCS is gradually lower for girls, and it is minimum for girls in the top category of ESCS).

3. Results

Observing the gender gap over the time (Fig. 1), we noted a consistent disadvantage for girls over the whole time-period: the vertical axes represents the difference between the average score of boys and the average score of girls, for each school year. The two scores, percentage and WLE[6], have the same trend, but differences more pronounced when we consider jointly the ability of students and difficulty of items (WLE score). In fact, focusing the attention on WLE score, we found in 2012/2013, 2014/2015 and especially in 2017/2018 the highest differences between girls and boys. In the last year, 2018/2019, there is a drastic reduction even if the average score in Math is still about 5 points lower for girls.

[6] WLE score wasn't calculated by INVALSI in the first two years (2009/2010 and 2010/2011).

ISBN 9788835123330

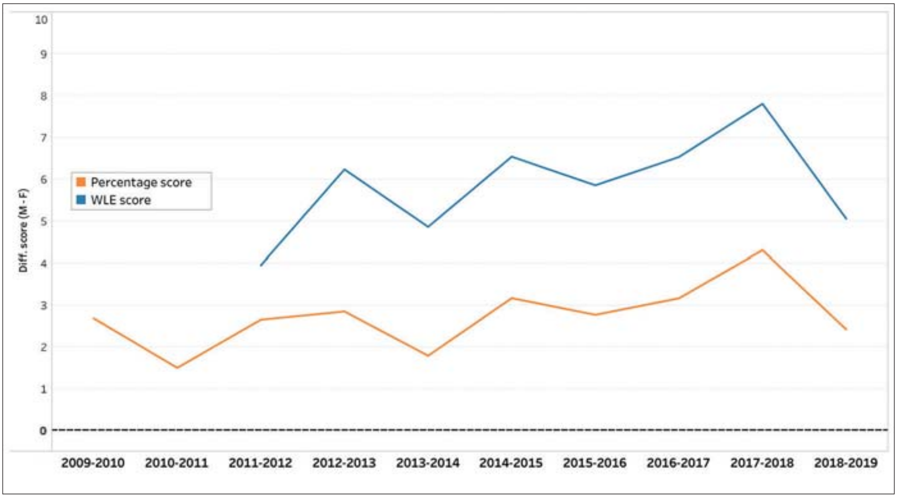


Fig. 1 – Gender Gap in Mathematics from 2009/2010 to 2018/2019 (percentage score and WLE score)

After these general results, we decided to study in-depth gender differences considering family's resources. In contrast to our original hypothesis data showed that the gender gap increases with the better conditions[7] (Fig. 2). The trend was directly proportional: the more the increase of ESCS and mother's education the higher the differences between boys and girls, and this result is consistent over the entire decade.

Basically, the results from descriptive analysis seem to indicate that boys take advantages from favourable conditions more than girls.

We looked whether this result was confirmed also at territorial level, with higher gender differences in regions more economically developed and with better school outcomes, traditionally in the North of Italy.

Furthermore, geographical area is an important and discriminant variable to study gender gap. In international research, as well as PISA tests, the hypothesis is that differences in achievements could decrease in countries characterized by a better gender equality system.

[7] For 2009/2010 and 2010/2011 information about ESCS and mother's education weren't recorded.

ISBN 9788835123330

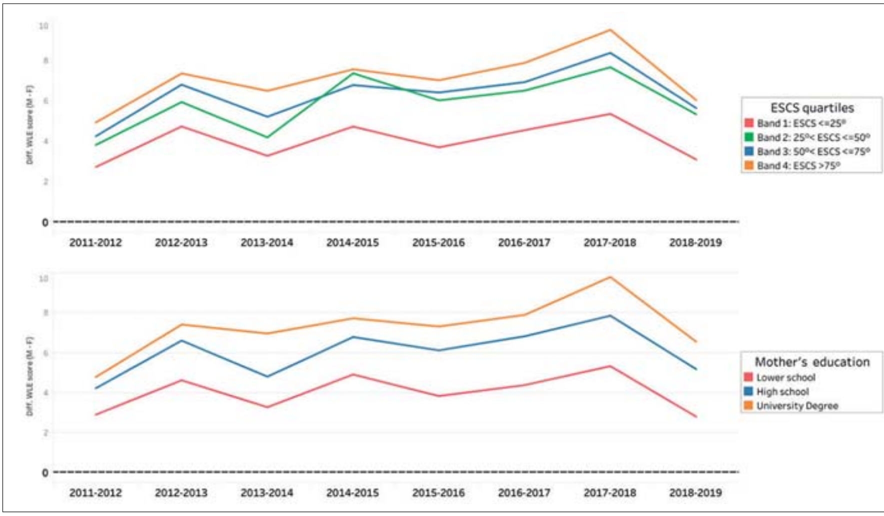


Fig. 2 – Gender Gap by ESCS quartiles and Mother's education from 2011/2012 to 2018/2019 (WLE score)

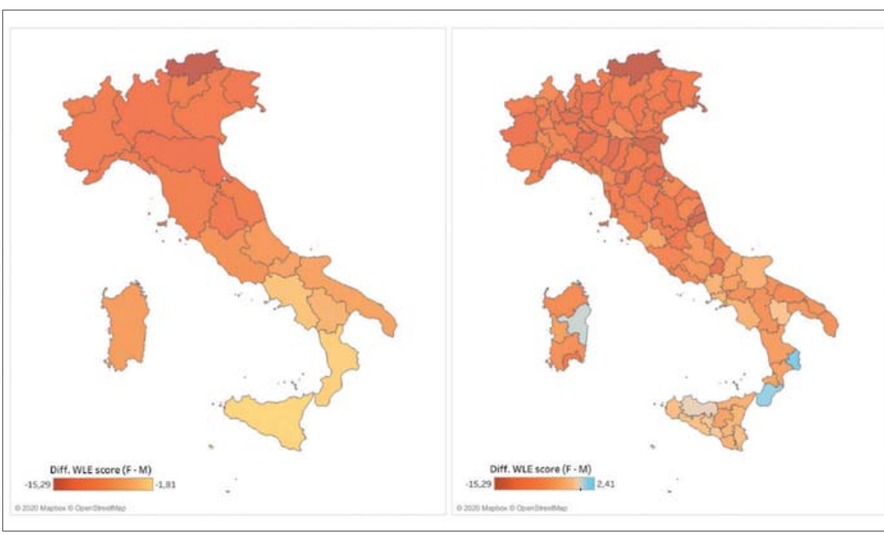


Fig. 3 – Female disadvantage by Regions and Provinces in 2017/2018 (WLE score)

ISBN 9788835123330

Again, our results went against the common belief (Fig. 3). In fact, in 2017/2018[8] we observed a counter-intuitive gender gap trend in more advantaged areas and, in general, in the North Italy.

In terms of mean scores, the substantial gender gap in more favourable conditions was found both if we consider socio-cultural-economic aspects and geographical ones.

In better contexts and whatever the advantage conditions, the differences between boys and girls increases.

With the aim to deepen this hypothesis, we analysed the gender gap by Parts of test[9] (Fig. 4) and the results corroborate the existence of difficulties related to specific competences in Mathematics. In fact, we noted a clear gender gap in Numbers and Relations/Functions parts and a lower one in Statistics and, especially, in Shapes/Figures parts (in this last case the difference is near to 0).

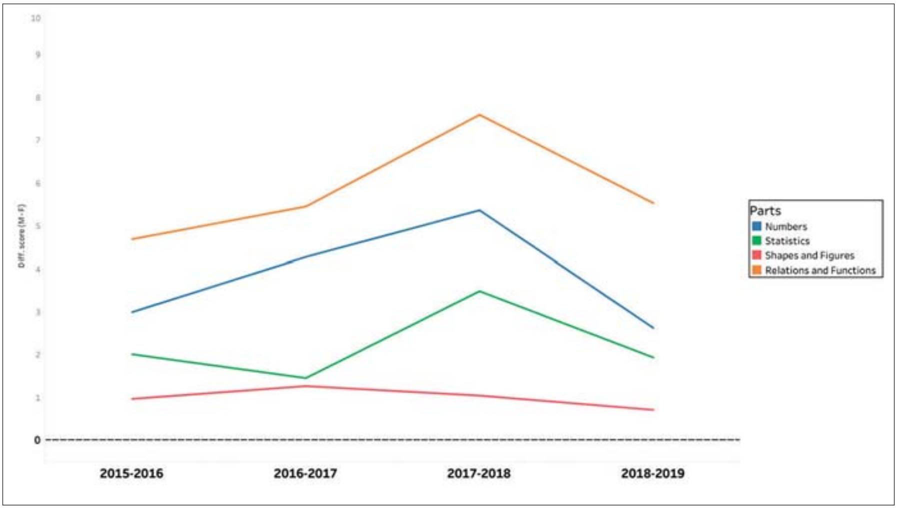


Fig. 4 – Gender Gap by Parts from 2015/2016 to 2018/2019 (percentage scores)

In this way, it seems that differences in performance are influenced also by the type of mathematical requested activity and by cognitive strategies used by boys and girls.

[8] Our decision to considering 2017/2018 is justified by the higher differences that characterize this school year. Robustness check was performed using year 2018/2019 with the same results (not shown).

[9] We considered from school year 2015/2016 because only since this year onwards the Parts of INVALSI test remained with this fixed structure.

ISBN 9788835123330

In a final step of analysis of this study, we elaborated nested linear regression models for WLE scores in year 2017/2018 with the intention of better understanding and reinforcing the results observed in the descriptive analysis. Gender is the key independent variable in these models, thus the coefficient associated with "female" is the reduction in WLE scores for girls with respect to boys.

This approach allowed us to evaluate the robustness of the main result of this work, i.e. the observation that "the better the conditions the larger the gender gap", net from possible confounding factors.

In addition, this analysis consented to look at the effect of each factors (ESCS, macro-area, nursery school, anxiety) separated between boys and girls. All the results are summarized in Tab. 2.

For each covariate with an interaction effect the interpretation of coefficient is as follows: the general effect is estimated only among boys, whereas the effect of the interaction (e.g. ESCS * Female) is the variation of the effect for girls.

For each factor of advantage (e.g. ESCS, geographical area, etc.) the positive effect observed for boys (primary effect) is decreasing for girls (secondary/interaction effects) as we move from less advantaged categories to the most advantaged ones. The gender gap in fact has its maximum within the top categories: highest quartile of ESCS, North-East regions, attended nursery school.

In Model 1, for example, being in the second quartile of ESCS increases WLE scores of 11.8 for boys and 10.3 for girls (coefficient of girls is obtained as 11.8-1.5 of the interaction effect); being in the upper quartile of ESCS increases WLE scores of 23.7 points for boys, but only 20.6 (23.7-3.1) for girls. The same trend is observed for geographical area where the school is located, with the highest difference in the effects observed in the North-East regions. The same trend is also found for the variable "have attended nursery school".

The regression results confirm what has already been shown in the first section of descriptive analysis. It seems that girls experience a sort of *plateau* in their performance, beyond which they cannot rise (a sort of "glass ceiling"), whereas boys reach more easily the top performances.

ISBN 9788835123330

Tab. 2 – Nested linear regression models in 2017/2018

	(1)	*(2)*	*(3)*	*(4)*	*(5)*	*(6)*
Variables	*ESCS + Interaction*	*Geo Area + Interaction*	*Nursery + Interaction*	*WLE_MAT_2ND Grade*	*Anxiety*	*Anxiety + Interaction*
Gender (reference: Male)						
Female	-7.1***	-3.6***	-3.4***	-2.4***	-1.1***	1.7***
ESCS (reference: <25%)						
ESCS 25-50%	11.8***	11.4***	11.2***	6.9***	6.7***	6.6***
ESCS 50-75%	16.7***	16.2***	15.9***	9.7***	9.4***	9.3***
ESCS > 75%	23.7***	23.2***	22.7***	14.2***	13.9***	13.7***
ESCS * Gender						
ESCS 25-50% * Female	-1.5***	-0.8*	-0.7	-0.7**	-0.7*	-0.6
ESCS 50-75% * Female	-2.1***	-1.4***	-1.2***	-1.1***	-1.1***	-0.9**
ESCS > 75% * Female	-3.1***	-2.4***	-2.2***	-2.0***	-2.0***	-1.7***
Geographical Area (reference: South)						
Centre		3.1***	3.0***	-0.9***	-1.2***	-1.3***
Nord-East		3.6***	3.4***	-0.6***	-1.0***	-1.1***
North West		3.5***	3.2***	0.6**	0.4	0.3
Geographical Area * Gender						
Centre * Female		-3.9***	-3.8***	-3.3***	-3.2***	-3.1***
North-East * Female		-5.8***	-5.7***	-4.4***	-4.3***	-4.1***
Nord-West * Female		-6.2***	-6.1***	-4.8***	-4.7***	-4.5***

(to be continued)

ISBN 9788835123330

Tab. 2 – Nested linear regression models in 2017/2018

	(1)	*(2)*	*(3)*	*(4)*	*(5)*	*(6)*
Variables	*ESCS + Interaction*	*Geo Area + Interaction*	*Nursery + Interaction*	*WLE_MAT_2ND Grade*	*Anxiety*	*Anxiety + Interaction*
Nursery School (reference: no)						
Yes			2.1***	1.4***	1.3***	1.3***
Yes * Female			-1.1***	-0.6**	-0.6**	-0.5**
Wle score at Grade 2 (continuous)				0.4***	0.4***	0.4***
Anxiety before Test (ref: A lot)						
Enough					5.4***	6.9***
Few					8.9***	11.3***
Not at all					10.8***	12.6***
Anxiety before Test * Gender						
Enough * Female						-2.3***
Few * Female						-3.7***
Not at all * Female						-5.3***
Constant	201.1***	199.0***	198.6***	119.2***	114.1***	112.4***
Observations	230,598	230,598	230,598	230,598	230,598	230,598
R-squared	0.065	0.066	0.067	0.310	0.320	0.320

*** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$

ISBN 9788835123330

To go more in-depth on these results we inserted in Model 4 the variable "WLE Math score at grade 2^{nd}" obtained by linking for each student his/her score three years before, in grade 2^{nd} in the year 2014-2015. This variable allowed us to capture the basic skills in Math at the beginning of primary school. The hypothesis here was that if a "cognitive" difference exists between boys and girls, we should see this difference already in grade 2^{nd} and, after the introduction of WLE score in grade 2^{nd} in Model 4 the significance of coefficient of gender in grade 5^{th} should decrease.

Results of Model 4 with basic ability in Math in grade 2^{nd} show that the coefficient of the females drops again of 1 point passing from -3.4 to -2.4 (demonstrating that part of the difference was already present in grade 2^{nd}) however it does not become completely insignificant, indicating that:

- either there are cultural factors that act between grade 2^{nd} and grade 5^{th} (two students, one male and one female, who start from the same level in grade 2^{nd} on average are seen detached in grade 5^{th} for the benefit of the male student);
- or biological/cognitive factors become more prominent as the test becomes more difficult.

Obviously the two hypotheses can coexist.

We also included anxiety, first as the only primary effect without interactions (Model 5). Subsequently in the last model (Model 6) we kept the same variable with the addition of interaction effects, i.e. the differential effect of anxiety on boys and girls.

In Model 5, the overall effect of anxiety does not explain entirely the difference in Math scores between girls and boys, in fact the gender coefficient remains significant with a mean reduction of WLE scores of -1.6 points for girls with respect to boys. If we include also the interaction effects of anxiety with gender, in Model 6, we can see the general trend observed for all the "positive factors" on school achievement, corroborating our hypothesis "the better the conditions the larger the gender gap". In fact, we saw one more time that in the most favourable context (no anxiety) the gap between boys and girls is the largest (the benefit of +13.6 points shrinks of 5 points for girls, going to +8 points). Surprisingly, in the last model (Model 6) the effect of gender, which is calculated for the "reference category" that summarizes all the less favourable conditions (lowest ESCS, school in the South, no nursery school attendance, high anxiety levels), changes its direction and becomes for the very first time in favour of girls (+1.78 points of WLE).

Summarizing the main results, it seems that in difficult conditions, girls perform better than boys, but this advantage reverses as we move to more favourable conditions, and it worsens gradually reaching the largest gap in

ISBN 9788835123330

favour of boys in the best contexts, i.e. when both, male and female students, have the highest mean scores.

This can be confirmed also by the ratio of male/female students among the top performers. Although in the original population the ratio of males over females is perfectly balanced, if we look among the 10% top performing students the proportion of male students is 62.8%, over 10 points above the half.

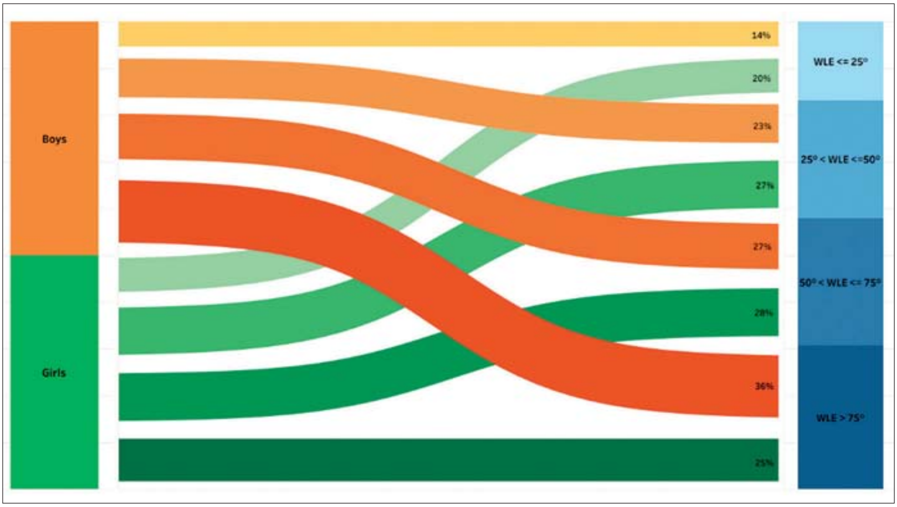


Fig. 5 – Students in the top quartile of ESCS: distribution into WLE levels according to gender in 2017/2018

Fig. 5 represents the distribution of only students of the highest socio-economic level (4th quartile of ESCS) into the four quartiles levels of Math WLE scores (in blue on the right) according to their gender (in green and orange on the left). We can have an illustration of the disproportion of boys and girls coming to the highest socio-economic group into the levels of Math scores (< 25%, 25-50%, 50-75% and > 75%).

Among students with the highest availability of economic and cultural resources girls' scores are basically equidistributed through the 4 classes of performance, whereas boys' scores are more concentrated in the top category. In fact, 36 over 100 boys with the highest ESCS status reach the top quartile of performance (WLE > 75%), whereas only 25 over 100 girls have the same result. This result seems to corroborate the hypothesis that gender differences are especially pronounced among students that benefit of the most favourable conditions: in such situations boys outperform girls, reaching top-level results more than girls do.

ISBN 9788835123330

4. Discussion

Girls disadvantage in Mathematical results at school has been extensively documented over time and space (EACEA P9 Eurydice, 2010; Fryer and Levitt, 2010).

The debate about factors affecting Mathematical gender gap sees the contraposition of two perspectives: the "internal" determinants, i.e. biological/cognitive factors (Naour, 2001) versus "external" determinants, particularly social/cultural background of the student, that are manly found associated with gender gap at country level (Guiso *et al.*, 2008). The study of gender gap according to socio-economic level of the student's family gives the opportunity to have some insight on the prevalence of cultural factors over biological ones (socio-economic status is strictly linked to cultural implications) and at the same time it responds to the indications of studying different sources of inequality jointly (OECD, 2016).

This study adopted the perspective of cumulative disadvantage and is oriented to gain a better understanding of factors influencing Mathematical achievement in grade 5th of schooling in Italy, through a population level analysis over an entire decade, from 2009/2010 to 2018/2019.

The general trend over the decade showed a consistent girls' disadvantage in Mathematical performance independently from the scoring method used (percentage or WLE score), reaching the largest gap in the year 2017/2018. By looking to the same trend, splitting students into four categories according to their socio-economic background we found an increasing differential between boys and girls as the ESCS conditions improve.

Results of this kind have already been shown in literature (Contini, Di Tommaso and Mendolia, 2017; Fryer and Levitt, 2010; Lubienski *et al.*, 2013); however, a clear interpretation of this inverse relationship has not been found yet.

We decided to better investigate this result by extending the concept of "favourable conditions". We looked at territorial differences in gender gap across Italian macro-areas and found that regions with the highest average Math scores, traditionally those in the North, showed also the highest level of gender inequality (Giofrè *et al.*, 2020).

Findings from socio-economic and territorial analyses seemed to give the same indications: when the external conditions improve the differences in the performance between boys and girls increase. Thus, we named such a phenomenon as "the better the conditions the larger the gender gap". A possible reason for this effect could lie on different cognitive strategies and abilities between girls and boys that become more evident when the average

ISBN 9788835123330

performance improves, i.e. among top performing students. In such group boys manage in reaching the highest scores more than girls.

We finalized the research through an analysis with nested linear regression models, in order to look at our hypothesis in a more systematic way, considering the effect of all the available potential determinants of gender inequalities jointly and in interaction with gender. These models highlighted how girls perform better than boys in difficult conditions, whatever the typology of disadvantage (socio-economic, territorial, absence of pre-schooling activities, etc.), however, this trend reversed as the conditions improve and worsen gradually until it reaches the maximum disadvantage for girls in the most favourable conditions.

Contini, Di Tommaso and Mendolia (2017) put forward the hypothesis that girls benefit less from family resources. However, with our analysis, we showed that this effect is not limited to family resources, but it spread to other forms of external advantage (social, territorial, attitudinal). In all of these "better conditions" girls systematically underperform boys.

A "socio-cultural" interpretation could be that the influence of the cultural stereotype that sees girls "not suited" for Mathematics, is greater at high levels of performance (which require more motivation and self-efficacy) than at low-medium levels. In other words, when circumstances make a student in the condition of reaching high results, boys express more their full potential than girls restrained by social stereotypes.

Our hypothesis is that if girls benefit less from "favourable conditions" it could also be because biological characteristics and cognitive strategies are at work, differentiating boys and girls in specific skills particularly evident at high level of average performance.

In fact, it has been largely demonstrated in neuroscience literature that brain development and functioning is different between men and women already from the first years of life: e.g. girls outperform boys in fine motor coordination, and vice-versa males have superior visuospatial skills, detectable already among 3-month-old infants (Goldman, 2017).

This hypothesis finds support in results from Cascella, Giberti and Bolondi (2020) about different performances between males and females according to specific items of the INVALSI test. For items that involve either interpreting decimal numbers or estimating a measurement, authors illustrated the highest differences in percentage of correct answers between boys and girls, indicating that specific cognitive abilities penalize mostly girls' performance. A similar indication can be derived from our analysis of gender gap by Parts of the Math test, where specific abilities, such as Relationship and Functions, seemed to be more discriminating between girls and boys.

ISBN 9788835123330

What our results display is, in fact, that girls overperform boys in difficult conditions, where skills such as precision, perseverance, concentration count the most. On the contrary, in the most favourable conditions, girls' performance reaches a sort of *plateau*, while boys in the same condition go beyond that *plateau* reaching average scores at the very top levels, probably thanks to cognitive and meta-cognitive skills such as greater intuition and higher propensity to new methods and alternative approaches.

Differences between the two groups (boys and girls) are thus mostly evident in contexts with the most favourable conditions, because there we find the highest mean scores, where boys show a superiority in reaching top scores. Thus, gender gap is maximum among the top performers, and, thus, indirectly among the categories with the most favourable conditions (high socio-economic level, Norther regions in Italy, pre-school attenders, and so on).

This result needs to be further corroborated by future research in different contexts, however, if it comes out to be reliable, it will indicate that cognitive skills differentiating boys and girls have an important impact on Math achievement, already in the first years of schooling.

Nevertheless, it does not mean that girls are destined to lower performances. Although biological factors are innate and immutable, the way in which mathematical is proposed, the teaching methods, the selection of items for the test can promote the use of specific cognitive skills, more easy-accessible and dominant for girls, and consequently reduce the gender gap (Boaler, 2009).

Some studies have shown that when mathematical teaching is centred upon collaborative work, proactivity, investigative approach and treats mistakes as an opportunity for learning (Boaler, 2013) the gender gap decreases. In a very recent study about Italy, Di Tommaso *et al.* (2020) designed and tested an innovative teaching method promoting the active participation of students through peer interactions, sharing of ideas and problem posing. The authors demonstrated that this kind of educational strategy has causal effects in narrowing gender gaps.

5. Conclusion

The effects of ESCS on gender gap are debated. With this study we confirm a counter-intuitive effect of high socio-economic background on gender gap, with girls disadvantaged particularly at higher level of socio-economic status. This effect is not limited to ESCS but is observed also with other external "beneficial" factors, in a wide-spread phenomenon we named "the better the conditions the larger the gender gap".

ISBN 9788835123330

We tried to give an interpretation to these consistent findings about girls' disadvantage in more favourable conditions, stressing biological factors and cognitive strategies that characterized girls and boys. Cognitive strategies in fact seem to advantage boys especially in the most advantageous conditions, where they can express their maximum potential, reaching very top levels of performance.

Part of these differences can be reinforced or weakened based on the educational methods adopted, changing teaching strategy in a more proactive and cooperative way opens the field to potential significant reductions of gender gap and should be encouraged and promoted at all levels of schooling.

References

Becker M., McElvany N. (2018), "The Interplay of Gender and Social Background: A Longitudinal Study of Interaction Effects in Reading Attitudes and Behavior", *British Journal of Educational Psychology*, 88, pp. 529-549.

Boaler J. (2009), *The Elephant in the Classroom: Helping Children Learn and Love Math*, Souvenir Press, London.

Boaler J. (2013), "Ability and Mathematics: the Mindset Revolution that is Reshaping Education", *Forum*, 55, 1, pp. 143-152.

Campodifiori E., Figura E., Papini M., Ricci R. (2010), "Un indicatore di status socio-economico-culturale degli allievi della quinta primaria in Italia", *INVALSI Working Paper*, 2.

Cascella C., Giberti C., Bolondi G. (2020), "An Analysis of Differential Item Functioning on INVALSI Tests, Designed to Explore Gender Gap in Mathematical Tasks", *Studies in Educational Evaluation*, 64, pp. 1-12.

Checchi D., Fiorio C.V., Leonardi M. (2006), "Sessanta anni di istruzione in Italia", *Rivista di Politica Economica*, 96, 7-8, pp. 285-318.

Contini D., Di Tommaso M.L., Mendolia S. (2017), "The Gender Gap in Mathematics Achievement: Evidence form Italian Data", *Economics of Education Review*, 58, pp. 32-42.

DiPrete T.A., Eirich G.M. (2006), "Cumulative Advantage as a Mechanism for Inequality: A Review of Theoretical and Empirical Developments", *Annual Review of Sociology*, 32, pp. 271-297.

Di Tommaso M.L., Contini D., Piazzalunga D., De Rosa D., Bernardi M. (2020), *Tackling the Gender Gap in Math with Active Learning Teaching Practices*, retrieved on December, 30, 2020, from https://drive.google.com/file/d/1tn-qzP1iJ-gIzfikCNDp9LLgCifEck460/view.

EACEA P9 Eurydice (Education, Audiovisual and Culture Executive Agency) (2010), *Gender Differences in Educational Outcomes: Study on the Measures Taken and the Current Situation in Europe*, retrieved on December, 30, 2020,

ISBN 9788835123330

from https://op.europa.eu/en/publication-detail/-/publication/40271e21-ca1b-461e-ba23-88fe4d4b3fd4.

Englund M., Luckner A., Whaley G., Egeland B. (2004), "Children's Achievement in Early Elementary School: Longitudinal Effects of Parental Involvement, Expectations, and Quality of Assistance", *Journal of Educational Psychology*, 96, 4, pp. 723-730.

Entwisle D.R., Alexander K.L., Olson L.S. (2007), "Early Schooling: The Handicap of Being Poor and Male", *Sociology of Education*, 80, pp. 114-138.

Fryer R.G., Levitt S.D. (2010), "An Empirical Analysis of the Gender Gap in Mathematics", *American Economic Journal: Applied Economics*, 2, 2, pp. 210-240.

Giofrè D., Cornoldi C., Martini A., Toffalini E. (2020), "A Population Level Analysis of the Gender Gap in Mathematics: Results on over 13 Million Children Using the INVALSI Dataset", *Intelligence*, 81.

Goldman B. (2017), "Two minds. The cognitive differences between men and women", *Sex, gender and medicine*, *Stanford Medicine*, Spring, retrieved on December, 30, 2020, from https://stanmed.stanford.edu/2017spring/how-mens-and-womens-brains-are-different.html.

Guiso L., Monte F., Sapienza P., Zingales L. (2008), "Culture, Gender, and Math", *Science*, 320, 5880, pp. 1164-1165.

Harding J.F., Morris P.A., Hughes D. (2015), "The Relationship between Maternal Education and Children's Academic Outcomes: A Theoretical Framework", *Journal of Marriage and Family*, 77, 1, pp. 60-76.

Kollmayer M., Schober B., Spiel C. (2018), "Gender Stereotypes in Education: Development, Consequences, and Interventions", *European Journal of Developmental Psychology*, 15, 4, pp. 361-377.

Lubienski S., Robinson J., Crane C., Ganley C. (2013), Girls' and Boys' Mathematics Achievement, Affect, and Experiences: Findings from ECLS-K, *Journal for Research in Mathematics Education*, 44, pp. 634-645.

Naour P. J. (2001), "Brain/behavior Relationships, Gender Differences, and the Learning Disabled", *Theory into Practice*, 24, pp. 100-104.

OECD (2015), *The ABC of Gender Equality in Education: Aptitude, Behaviour, Confidence*, PISA, OECD Publishing, retrieved on December, 30, 2020, from https://www.oecd.org/pisa/keyfindings/pisa-2012-results-gender-eng.pdf.

OECD (2016), *Low-Performing Students. Why They Fall Behind and How to Help Them Succeed*, OECD Publishing, retrieved on December, 30, 2020, from https://www.oecd-ilibrary.org/docserver/9789264250246-en.pdf?expires=1587738001and id=idand accname=guestand checksum=EE3C251AFA67AA31027778C522B22059.

Pallavi A.B. (2016), "A Systematic Review of Factors Linked to Poor Academic Performance of Disadvantaged Students in Science and Math in Schools", *Cogent Education*, 3, 1, pp. 1-17.

Parveen A., Alam M. T. (2008), "Does Mothers' Education Influence Children's Personality Factors and Academic Achievement?", *Bulletin of Education and Research*, 30, 2, pp. 1-6.

ISBN 9788835123330

Quintano C., Castellano R., Longobardi S. (2009), *A Fuzzy Clustering Approach to Improve the accuracy of Italian Students' Data. An Experimental Procedure to Correct the Impact of the Outliers on Assessment Test Scores*, retrieved on December, 30, 2020, from https://www.invalsi.it/invalsi/ri/sis/documenti/022013/longobardi_paper.pdf.
Reardon S. (2011), "The Widening Academic Achievement Gap Between the Rich and the Poor: New Evidence and Possible Explanations", in G. Duncan, R. Murnane (*ed.*), *Whither Opportunity?: Rising Inequality, Schools, and Children's Life Chances*, Russell Sage Foundation, New York.
Tinklin T., Croxford L., Ducklin A., Frame B. (2003), "Inclusion: A Gender Perspective", *Policy Futures in Education*, 1, 4, pp. 640-652.
Strand S. (2014), "School Effects and Ethnic, Gender and Socio-Economic Gaps in Educational Achievement at Age 11", *Oxford Review of Education*, 40, 2, pp. 223-245.
Wilder G.Z., Powell K. (1989), *Sex Differences in Test Performance: A Survey of the Literature*, College Board Report n. 89-3, College Entrance Examination Board, New York.

ISBN 9788835123330

5. Immigrant performance towards reading in OECD PISA 2018

by Paola Giangiacomo, Valeria F. Tortora

The number of immigrant students has grown considerably in the past 20 years in most countries. According to the OECD, around 4.8 million immigrants arrived in OECD countries in 2015, a wave that reinforced a long and steady upward trend (OECD, 2018). In OECD countries, between 2000 and 2009, the percentage of students of immigrant origin grew by an average of three percentage points, in Italy the percentage of immigrant students grew by almost 4.5% (4.1% second generation students, 0.3% first generation students).

The way in which schools and education systems respond to the challenges and opportunities that arise with migratory flows has profound economic and social implications.

The key to maintaining social cohesion during these population movements is to encourage the integration of immigrants and their families in the countries of adoption; education can be an important lever to achieve this.

In Italy, since the beginning of the migration phenomenon, the school has taken shape as a place of integration, of cultural exchange, of meeting with languages and stories that are worth knowing and enhancing in common educational spaces. Most of the time, the inclusion of one or more foreign students has led to a change in the organization of teaching in various aspects; for example the preparation of spaces dedicated to language laboratories, the acquisition of specific aids for linguistic needs and information on the origin and personal and scholastic history of each. In other cases, the school has made use of external resources, operators organized by the local authority, indigenous mediators with more or less defined professional skills, adopting response methods ranging from delegation to other integrations of project resources.

In most countries, as demonstrated by international surveys sponsored by the OECD and the IEA, native students achieve results well above their first and second generation immigrant colleagues.

ISBN 9788835123330

Through international surveys, a strong correlation between the performances of immigrants and their socio-economic conditions has been demonstrated (OECD, 2019).

1. Introduction

Europe has always underlined the importance of promoting the integration of children and young people from migrant backgrounds in schools and facilitating their integration into society through education. There are many political initiatives that the Union has developed over the years to address the various challenges, the most recent of which are: the “2016 European Commission’s Action Plan on the Integration of Third Country Citizens” and the “Communication on Protection of children in migration” in 2017.

The first document highlights how education and training are among the most powerful tools for integration; the second suggests actions aimed at the protection of all migrant children in the various stages of the process, also including assessing the needs of each in the shortest time possible from the moment of arrival, and an immediate access to education. Finally, on this basis is inserted the “2018 Council Recommendation” on the promotion of common values, inclusive education and a European dimension of teaching which underlines the importance of ensuring fair and effective access to education with the necessary support for all pupils, including those from migrant backgrounds.

On September 2015, the 193 ONU member countries adopted the 2030 Agenda for sustainable development. The 2030 Agenda represents the new global reference framework for national and international efforts aimed at finding common solutions to the great challenges of the planet, such as extreme poverty, climate change, environmental degradation and health crises. It has been in force since 2016 and has 17 Sustainable Development Goals (SDGs), ONU member states have declared themselves willing to achieve these goals together by 2030. The SDG 4 aims to “ensure an inclusive and equitable quality education and promote lifelong learning opportunities for all”. Quality education is the basis for improving people’s lives and achieving sustainable development. Important results have been achieved regarding the increase in access to education at all levels and the increase in enrollment levels in schools.

The way in which schools and educational systems respond to the challenges and opportunities that arise with migratory flows has profound economic and social implications: a well-integrated student, both at school and

ISBN 9788835123330

at social level, has more opportunities to achieve their potential and contribute to making the new country more competitive both economically and socially. The skills acquired by students in the host country will positively influence his chances of successfully entering as a citizen and worker in that country, school success can be interpreted as one of the first indicators of the integration of the immigrant population in the host country.

The key to maintaining social cohesion during population movements is to encourage, in the best way, the integration of immigrants and their families in the countries of adoption; education can be an important lever to achieve this. In most school systems, access to education and training, in the compulsory school age, is the same as that of their native companions, while in 13 education systems, among which Italy is not included, for young migrants, no longer of compulsory school age, it is possible not to have the right of access to education (European Commission/EACEA/Eurydice, 2019).

The creation of an egalitarian educational system, able to offer all students the same learning opportunities, has always been a central node in the debates related to the effectiveness of the national school systems of the OECD countries (Davoli, 2016). In several European countries, where there is already a stronger history of immigration, a complete integration of students of foreign origin into the system has not yet been achieved, as is also shown in the investigations promoted by the OECD (2020).

In Italy, since the beginning of the migration phenomenon, the school has taken shape as a place of integration, of cultural exchange, of meeting with languages and stories that are worth knowing and enhancing in common educational spaces. Most of the time, the inclusion of one or more foreign pupils has led to a change in the organization of teaching in various aspects; for example, the preparation of spaces dedicated to language laboratories, the acquisition of specific aids for linguistic needs and information on the origin and personal and scholastic history of each. In other situations, the small number of foreign pupils has led to underestimate the phenomenon and to rely only on the children's ability to adapt. The individual schools and teachers involved, based on the resources available, the availability of other bodies, responded differently to the presence of foreign students.

In most countries, as demonstrated by international surveys sponsored by the OECD and the IEA, native students achieve results well above their first and second-generation immigrant colleagues. Among the factors traditionally used by international literature to explain the educational disadvantage of first and second-generation immigrant students, elements of a socio-economic, institutional and nature linked to the migratory history of their family play a fundamental role. Among the most important explanatory factors of

ISBN 9788835123330

the school gap between natives and students of immigrant origin, a strong impact is given by the variable *language spoken at home* on the performance of students in general and even more on the performance in reading. In fact several studies show that speaking a language other than that of the test always involves, other things being equal, a decrease in the score obtained (Isphording and Otten, 2014).

The study is articulated in this way: in the following paragraph it illustrates the literature characteristis of immigrants, the third paragraph shows the objective and research hypotheses, the fourth introduces the target population of this study, the data and methodology are present in the fifth paragraph and the last part is dedicated to the results and discussion.

2. Background

In this perspective, it is of fundamental importance to investigate the differences in learning between foreign students, both first and second generation, and their native counterparts[1] (INVALSI, 2019). Thus, the analysis of INVALSI data allows us to shed light on those aspects in which educational institutions fail in the integration process: knowing where to intervene to reduce the *gap* in results means having the opportunity to find solutions.

To integrate fully into the new country, mastery of the language of the host community is essential if immigrant: it is one of the major challenges that immigrant students face (Isphording and Otten, 2014) and many immigrants never reach an adequate knowledge of the language of the host country (Isphording, 2015).

The performance and the relationship with socio-economic background is one of the enduring issues in educational research. Socio-economic background is usually considered a strong predictor of a student's score (Agasisti and Cordero-Ferrera, 2013; Willms, 2006; Woessmann, 2007).

The effects of socio-economic status on student achievement have been extensively studied and research has shed light on specific mechanisms that link economic, social and cultural heritage in the family context to student achievement (Jæger and Breen, 2016). Further studies show that students whose parents have higher educational levels and more prestigious and bet-

[1] By Native student we mean a child born in Italy to parents born in Italy, a First-generation foreign student means a child born in a foreign state and arrived in Italy at school age, and a Second generation foreign student means a child born in Italy to parents born in a foreign state.

ISBN 9788835123330

ter-paid jobs generally benefit from a wider range of financial (e.g. private tutoring, computers, books), cultural (e.g. time in active parenting) and social resources (e.g. networks) that make students' school success easier, compared to peers who come from families with lower education levels or who are affected by chronic unemployment, jobs low income or poverty.

3. Objective and Research Hypotheses

The aim of the study was to investigate what reading skills the first and second-generation immigrant students, who participated in the OECD PISA 2018 international survey, obtained, assessing whether and to what extent the school participated in integrating them into the school context. Furthermore, the OECD survey PISA, through a questionnaire completed by the students participating to the tests, collected data on a large set of characteristics of the families including the socio-economic and cultural level of the students' parents, the status of immigrant, the language spoken at home and family structure. By crossing these data with the results obtained in the tests, an attempt will be made to analyze their impact on the students' results. Specifically, it will be interesting to analyze the impact of the language spoken at home on the results of immigrant students, the involvement of parents and the disciplinary climate of the class. Being able to arrive to a deeper and more realistic examination of the factors that influence the levels of competence of immigrant students is useful for preparing a better development of intervention tools, both at the level of immigrant management in schools and more precise school policies.

The research questions of this study are the following: is the language barrier overcome if the test language is spoken at home? To explain the educational disadvantage of first and second-generation immigrant students, is it necessary to analyse the role played by the socio-economic aspects of parents? Are the territorial gaps also evident between immigrant students from the North and those from the South?

4. Target population

In order to investigate the integration of immigrant students we used OECD PISA 2018. The surveyed students are representative samples of the population of grade 8, 9, 10 and 11 students; 11,785 Italian students participated in PISA 2018.

ISBN 9788835123330

Immigrant students are divided by type of school into Lycei 2.5% first-generation immigrant and 3.5% second, in technical institute 4.7% first-generation immigrant and 6.7% second, in vocational institutions 4.7 % of first-generation immigrant and 6.7% of second, and finally vocational training institution 8.9% of first-generation immigrant and 9.2% of second. Already within the type of school, there are differences in the choice of the course of study: the presence of immigrant students within Lycei is clearly less than that of the natives, on the other hand, in technicians, professionals and vocational training their presence is decidedly higher. Even among the 15-year-old students present in the secondary school, first generation immigrants are definitely more than their second generation immigrant colleagues (43.1%). These data somehow communicate to us that foreign students prefer vocational courses over more Lyceum ones and, therefore, it seems that they tend to ensure a school path that can guarantee their insertion into the world of work.

Taking into account, however, the geographical area and adding the first and second-generation immigrant students, the percentage rises to 13.7% in the North West, 16.4 in the North East and 3.7% in the South. As for the type of school, 9% of foreign students are enrolled in vocational training courses, a percentage that rises to 11.4% for technical institutes and 18.2% for vocational institutes while it falls to 6% for Lycei. Italian students report average reading scores significantly 31 points higher than second-generation immigrant students and 43 points higher than first-generation immigrant students.

It should be noted that in the analysis data from lower secondary schools will not be commented on in this analysis because of the high standard errors indicating inaccurate estimates.

Tab. 1 – Percentage of students per type of schools

	Non-immigrant		*Second-generation immigrant*		*First-generation immigrant*	
	%	*S.E.*	*%*	*S.E.*	*%*	*S.E.*
Lyceum	93.99	0.54	3.52	0.38	2.50	0.33
Technical institute	88.57	0.90	6.70	0.77	4.73	0.40
Vocational institute	81.78	2.05	9.25	1.62	8.96	1.04
Lower secondary education	38.88	7.76	17.99	5.78	43.13	9.33
Vocational training*	90.88	2.04	4.61	0.98	4.50	1.41

* Lower secondary education means Scuole secondarie di I grado and Vocational training means Center of Vocational Education.

Source: OECD, PISA 2018 database

ISBN 9788835123330

This figure is also confirmed by the greater presence in the two Italian areas of lower unemployment and higher industrialization, North East and North West (respectively 8.4% of 2nd generation and 8% of 1st, and 7.6% and 6% of first-generation immigrant). These data inform us not only that families have settled in the places of greatest possibility of employment for them but also in perspective for their children. In the southern areas the attendance rates are very low.

Tab. 2 – Percentage of students per Geographic Area

	Non-immigrant		*Second-generation immigrant*		*First-generation immigrant*	
	%	*S.E.*	*%*	*S.E.*	*%*	*S.E.*
Italy	90.00	0.50	5.50	0.40	4.60	0.30
North West	86.32	1.32	7.64	0.88	6.04	0.73
North East	83.64	1.10	8.38	1.07	7.98	0.63
Center	89.76	1.03	5.48	0.76	4.76	0.62
South	96.34	0.50	2.14	0.40	1.52	0.45
South Islands	94.05	1.07	3.51	0.89	2.44	0.46

Source: OECD, PISA 2018 database

5. Data and methodology

For the purpose of this report, we included some variables of the OECD reports when presenting and discussing PISA 2018, cross-country differences and similarities in the performance and characteristics of students with immigrant background. We included three variables (epistemological beliefs, truancy) which have been shown to be major predictors of PISA performance in the recent PISA 2018 round (Hippe, Jakubowski, and Araújo, 2017; OECD, 2016) and which are also susceptible to be relevant drivers of expected dropout rates.

Economic, Social and Cultural Status (ESCS) is an OECD index measuring student socio-economic background. PISA measures ESCS with an extensive set of questions related to parent occupation, education and household cultural, educational and economic resources[2]. The index is standard-

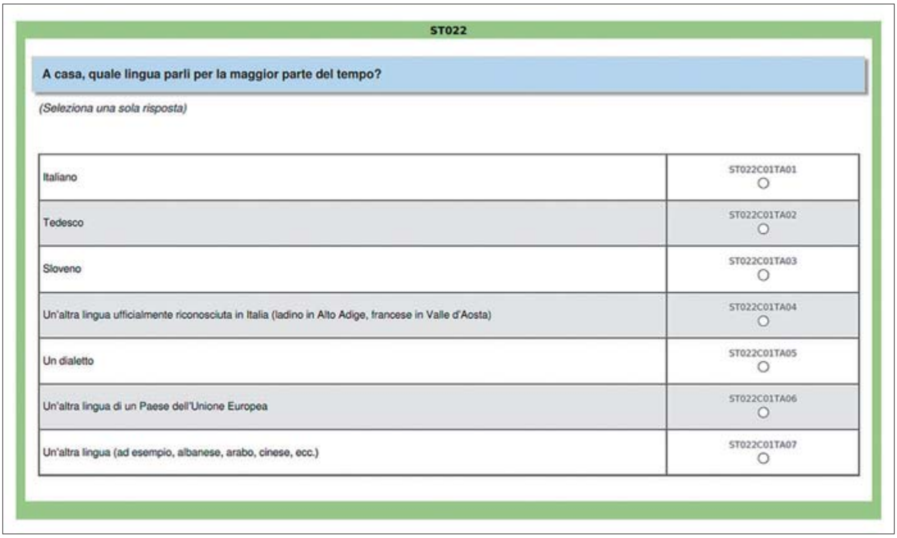
[2] The ESCS index is calculated by the OECD through the use of three indicators: the highest employment level of parents (HISEI); the highest level of parental education measured in years of schooling (PARED); home-owned assets (HOMEPOS), which in turn include all the items of the family wealth indicator (WEALTH), home-owned cultural assets (CULT-

ISBN 9788835123330

ised to have a mean of 0 and a standard deviation of 1 across OECD countries (weighting each country equally).

Immigrant background refers to first or second-generation immigrant students. As first and second-generation immigrant students have different immigration backgrounds, the results may, in many cases, vary according to each group. In consequence, we include this distinction among immigrant students.

Language spoken at home controls for the fact that the language taught at school and the language spoken at home may be different. In consequence, a student may have more difficulty in the subjects at school and in socialising with schoolmates when the same language is not practiced in both home and school environments. Thus, not speaking the language at home has also been shown to be negatively associated with PISA scores. In the questionnaire, the students have answered the question about language, in particular which language speak most at home (see Fig. 1). To analyze the students' answers to this question, the variables were re-coded in two, language of test and other language.


ST022

A casa, quale lingua parli per la maggior parte del tempo?

(Seleziona una sola risposta)

Italiano	ST022C01TA01 ○
Tedesco	ST022C01TA02 ○
Sloveno	ST022C01TA03 ○
Un'altra lingua ufficialmente riconosciuta in Italia (ladino in Alto Adige, francese in Valle d'Aosta)	ST022C01TA04 ○
Un dialetto	ST022C01TA05 ○
Un'altra lingua di un Paese dell'Unione Europea	ST022C01TA06 ○
Un'altra lingua (ad esempio, albanese, arabo, cinese, ecc.)	ST022C01TA07 ○

Fig. 1 – Question from the Student Questionnaire on the language spoken at home administrated in Italy

POSS) and owned educational resources (HEDRES), such as possession of books, requested through four categories: 0-10 books, 11-25 or 26-100 books, 101-200 or 201-500 books, more than 500 books.

ISBN 9788835123330

International surveys such as PISA report student performance through plausible values.

In order to take into account the evaluation errors deriving from the fact that not all the questions have been submitted to all the students, a set of ten plausible values has been calculated for each student with which an estimated probability is associated. The average of the population of Italian 15-year-old students is stimulated using the average of ten plausible values. The two scale marks were standardized with an average of 500 and a standard deviation of 100 (OECD, 2017b).

We have used first of all the simple descriptive analysis of the differences between native and immigrant students, in order to highlight the extent of the two groups gap.

Making use of Linear regression models we will be able to understand how student proficiency is influenced by immigrant and socio-economic status, and to estimate to what extent the interaction of these 2 variables has a cumulative effect. The linear regression used allowed to verify the effect of a particular predictor (bn), net of the variability share of the dependent variable explained by the other explanatory variables. Therefore, the partial regression coefficient associated with each independent variable represents the average change that we can expect to observe in correspondence to a unit increase of a predictor, keeping all the other predictors constant.

Data preparation was done with the IEA IDB Analyzer and IBM SPSS Statistics.

6. Results

PISA 2018 shows that first-generation immigrant students who spent more time in the destination country tend to perform better than those who spent less time in the country; whereas second-generation immigrant students tend to perform better than first-generation immigrant students but worse than their non-immigrant peers; and that the most vulnerable immigrant students tend to be those who arrive late at school, who have a limited command of the evaluation language in the host country and who come from a country where education standards are weaker (OECD, 2019b).

ISBN 9788835123330

6.1. Students' performance in reading

The results reported in this work concern reading literacy, as the main domain of PISA 2018.

The Figure 2 shows descriptive statistics, broken down by non-immigrant, first and second-generation immigrant and for all students in reading performance.

Non-immigrant students obtain the best reading score within each geographical area compared to their immigrant colleagues. But looking at the graph by analysing the performance of the second-generation immigrant students, it can be seen that in the North East the second-generation immigrant students reach averages statistically significant above those of the native students in the South and Islands: 462 in the North East compared to the natives of the South and the islands 445. It is interesting to note, too, that the "non-immigrant" students of the Northern regions achieve statistically significant higher results, with deviations of 50 points on average, compared to students in the areas of Southern Italy (512 in the North East and 445 in the South and Islands).

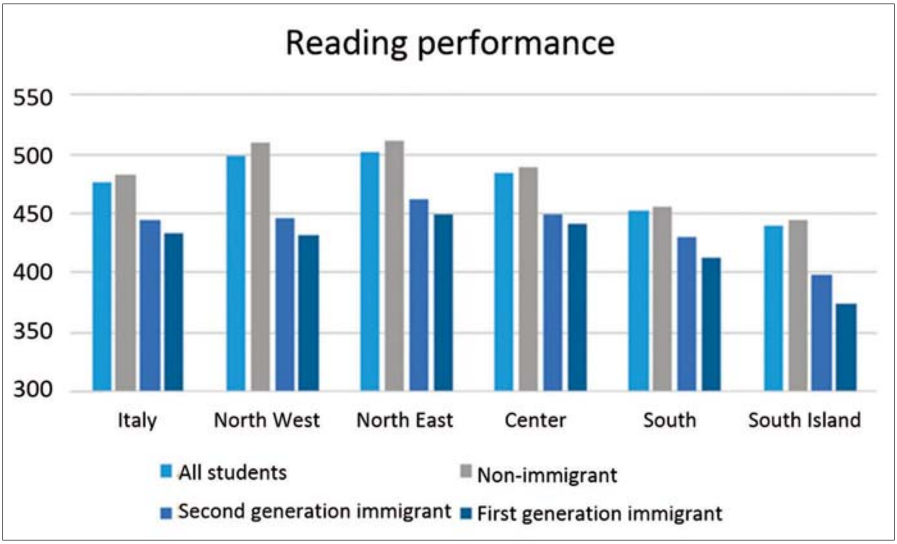


Fig. 2 – Performance in reading in PISA 2018

The following graphs show the results of the students (natives, first and second-generation immigrant students) divided by type of school attended and by geographical origin. The graphically represented data confirm a clear

ISBN 9788835123330

territorial gap and also show how first and second-generation immigrant students, in some geographic areas and in the same type of school, perform better than students without an immigrant background.

In reading the graphs it is immediately deduced that the second-generation immigrant students who attend Lycei in the North East (543 points on average) have significantly higher performances than the natives who attend Lycei in the South and Islands (490 points on average). If the analysis moves between the first-generation immigrants and the natives, it is interesting to note that the first-generation immigrant students of the North East achieve performances of 546 average points compared to the native South colleagues who instead have 503 points, therefore a difference of 43 points. The data also show the same trend in the other types of school: in the Southern and islands Technical Institutes, second-generation immigrant students obtain 394 average points, first-generation immigrant students 363 unlike North West students who obtain 459 average points.

Fig. 3 – Performance in reading: students non-immigrant attended the different types of school and divided by geographical area

ISBN 9788835123330

Fig. 4 – Performance in reading: second-generation immigrant attended the different types of school and divided by geographical area

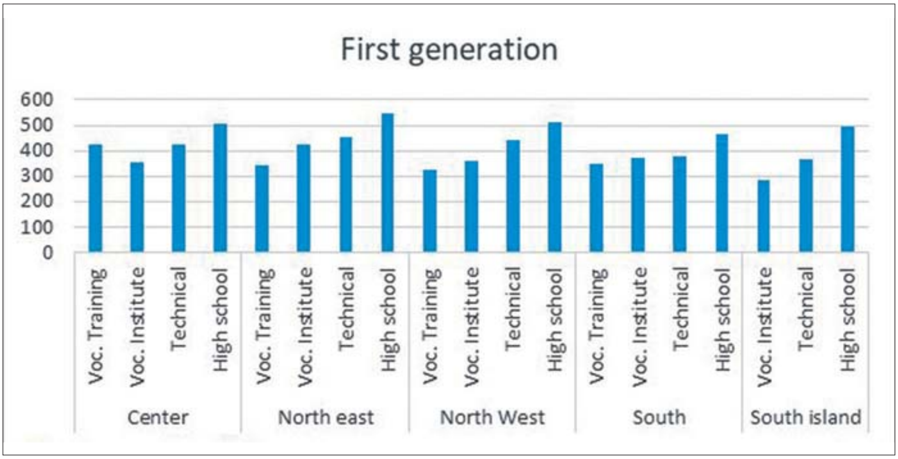


Fig. 5 – Performance in reading: first-generation immigrant attended the different types of school and divided by geographical area

6.2. Language spoken at home

Through some questions of the student questionnaire it was examined whether natives and immigrants students routinely used the language of the test or a different language at home.

ISBN 9788835123330

The Table 3 below shows that in Italy 18.5% of students speak at home a language different from that used in the PISA tests; the percentage rises for first and second-generation immigrant students respectively 70.4% and 55.4%.

Tab. 3 – Percentage of students who speak mainly another language at home – Italy

	Percentage of immigrant students		*Percentage of students who speak mainly another language at home*					
			All students		*Second-generation immigrant*		*First-generation immigrant*	
	%	*S.E.*	*%*	*S.E.*	*%*	*S.E.*	*%*	*S.E.*
Italy	10.0	(0.5)	18.5	(0.6)	55.4	(2.4)	70.4	(3.1)

If we analyze the impact of the language spoken at home on the results of students in reading tests, it can be seen that it has a very significant impact. Students who do not speak, at home, the language spoken at school tend to have a lower reading performance: the difference in average reading between immigrant students who speak the test language and those who speak a different language, in Italy, is 44 points and is statistically significant.

Tab. 4 – Average performance in reading of students who speak mainly another language at home – Italy

Immigrant students who speak mainly another language at home		*Immigrant students who speak the language of assessment at home*		*Difference in reading performance between immigrant students who speak the language of assessment at home and those who do not*	
Mean score	*S.E.*	*Mean score*	*S.E.*	*Score dif.*	*S.E.*
423	*(6.1)*	*467*	*(7.3)*	*44*	*(9.4)*

Note: statistically significant score point differences are marked in a darker tone.

The same data were analyzed by dividing the students by geographical area; Table 5 presents the data of first and second-generation immigrant, non-immigrant students, and the result of all students.

The difference between reading performances is significant between non-immigrant and second-generation immigrant in the North East, i.e. between the first immigrant students and the second there is a difference of 45 points on average, therefore the possibility of practicing the test language at home is very important to obtain good results. Furthermore, there is a significant difference also between non-immigrant and first-generation immigrant

ISBN 9788835123330

in the North East and in the Center, with a high difference also in these two cases, respectively of 43 and 40 points on average.

Tab. 5 – Average performance in reading of students who speak mainly another language at home – For Geographic Area

	All students		*Non-immigrant*		*Second-generation immigrant*		*First-generation immigrant*	
	Media	*E.S.*	*Media*	*E.S.*	*Media*	*E.S.*	*Media*	*E.S.*
Italy	432	4.2	439	4.9	424	7.5	422	8.1
North West	431	8.1	446	13.4	423	10.0	418	14.9
North East	465	9.5	485	11.2	440	15.0	442	13.6
Center	451	7.8	466	8.4	432	15.3	426	14.1
South	407	10.0	410	10.7	384	22.6	395	21.5
South Islands	388	10.4	396	10.6	382	26.8	344	31.2

Tab. 6 – Average performance in reading of students who speak mainly language of test at home – For Geographic Area

	All students		*Non-immigrant*		*Second-generation immigrant*		*First-generation immigrant*	
	Media	*E.S.*	*Media*	*E.S.*	*Media*	*E.S.*	*Media*	*E.S.*
Italy	487	2.8	489	2.8	471	8.90	459	12.8
North West	511	6.1	515	6.3	468	15.0	465	21.3
North East	515	5.6	517	5.7	503	17.6	469	24.6
Center	493	4.9	493	4.9	469	14.5	488	20.4
South	463	4.8	464	4.9	478	21.7	432	30.8
South Islands	450	8.5	454	8.2	415	21.1	408	49.9

Table 7 shows the results of students who speak mainly another language at home by type of school. The difference is significant between non-immigrant and first-generation immigrant students in technical institutes.

Tab. 7 – Average performance in reading of students who speak mainly another language at home – For type of school

	All students		*Non-immigrant*		*Second-generation immigrant*		*First-generation immigrant*	
	Media	*E.S.*	*Media*	*E.S.*	*Media*	*E.S.*	*Media*	*E.S.*
Lyceum	494	5.2	499	6.6	470	15.2	504	12.6
Technical institute	439	5.4	447	5.5	432	10.8	423	10.4
Vocational institute	373	8.1	373	8.1	374	14.6	375	18.8
Vocational training	372	10.9	369	11.2	374	23.5	388	32.8

ISBN 9788835123330

Tab. 8 – Average performance in reading of students who speak mainly the language of test at home – For type of school

	All students		*Non-immigrant*		*Second-generation immigrant*		*First-generation immigrant*	
	Media	*E.S.*	*Media*	*E.S.*	*Media*	*E.S.*	*Media*	*E.S.*
Lyceum	525	3.2	525	3.3	530	9.1	521	13.6
Technical institute	464	4.1	464	4.2	462	10.6	473	14.3
Vocational institute	406	10.6	408	10.7	434	24.1	398	23.8
Vocational training	418	5.6	422	5.7	358	34.6	341	30.3

6.3. Socio-economic status

The Table 9 shows the mean of index SES, Socio-economic status, and the belonging to the geographical area of the students.

The PISA socio-economic and cultural status index (ESCS) provides a measure of the family context and is based on the indices of material well-being and possession of educational and cultural resources, including the number of books at home, the occupation of the parents (considering the higher employment status between the two parents) and the level of education of the parents (considering the parent with the title of higher study). The value of this index can be referred to the student, to the school (as an average of the values of the index of students who attend it) or to the education systems (as an average of values of the index of all students in the country). The index is standardized with zero average and standard deviation 1 among the OECD countries. A low index value corresponds to a disadvantaged socio-economic and cultural situation.

As can be seen in Table 9, Non-immigrant students from the North and Center have an average level of socio-economic background higher than the national average, while those of the South and South Islands are below this level.

Tab. 9 – Mean of index of SES for Geographic Area

	Non-immigrant students		*Second-generation immigrant students*		*First-generation immigrant students*	
	Mean index	*S.E.*	*Mean index*	*S.E.*	*Mean index*	*S.E.*
Italy	-0.16	-0.02	-0.68	-0.06	-0.82	-0.06
North West	-0.07	0.04	-0.58	0.09	-0.81	0.1
North East	-0.04	0.05	-0.68	0.14	-0.8	0.12
Center	-0.01	0.05	-0.77	0.08	-0.94	0.14
South	-0.33	0.05	-0.66	0.13	-0.49	0.17
South Islands	-0.33	0.06	-0.86	0.13	-0.91	0.12

ISBN 9788835123330

Second and First generation immigrant students have an average level of socio-economic and cultural index lower than their non-immigrant colleagues.

The regression model presented in Table 10 shows a regression analysis in which the origin variable and the average score obtained for reading tests were inserted: the immigrant condition has a significant effect on the results of students in cognitive tests in reading, in particular being a second generation immigrant student means having 37 points less in the reading score than native students and 49 points less for first-generation students. These data are both significant. The ESCS index was included in the following model, considered statistically significant both in the literature and in Table 9.

Tab. 10 – Regression model to estimate the effect of immigrant status on reading performances

CNTRYID	*EqVar*	*b*	*b.t*
Italy	(CONSTANT)	482.4	185.5
Italy	IMMIG_2_generation	-37.4	-6.5
Italy	IMMIG_1_generation	-49.29	-6.43

In the regression model presented in Table 11, the ESCS index was inserted to assess its incidence: by introducing the ESCS, the difference in reading score of 30 points is explained. So, if you consider the ESCS, the advantage of the natives in the score decreases (22 points for the second generation students and 29 for the first generation students) but it still remains significant.

Tab. 11– Regression model to estimate the effect of immigrant status and socio-economic status on reading performances

CNTRYID	*EqVar*	*b*	*b.t*
Italy	(CONSTANT)	487.3	205.6
Italy	ESCS	29.8	15.9
Italy	IMMIG_2_generation	-22.1	-3.9
Italy	IMMIG_1_generation	-28.9	-3.9

Note: statistically significant score point differences are marked in a darker tone.

ISBN 9788835123330

7. Conclusions

With this study we have tried to show how it is possible with appropriate analysis models and empirical data to get a deeper and more realistic view of the factors that influence the outcomes of foreign students, also for the purpose of better development of intervention tools, both at the level of daily management in schools and in schools general policies

The data from this study show that students with immigrant origins differ considerably, both by geographic area and by type of school. This work also confirms some of the difficulties faced by immigrant students while settling in their new communities and new schools. Speaking a language other than the test language has a significant negative impact on obtaining a reading result in accordance with the averages of the native students.

Reading the international data of PISA 2018 it is clear that the educational success of immigrant students differs so widely between different countries and countries pursue such different policies and practices in exploiting the potential of immigrant children. So the data point out that countries among them can learn a lot from each other.

These results are also mostly in line with the literature. The impact of immigrant status on PISA scores in Italy was discussed. The results indicate, also, that students with an immigrant background are often associated to negative score gaps within the geographic area. Education systems, schools and teachers have a decisive role in helping students of immigrant origin to integrate into their own community, overcome language barrier and build their competencies scholar, social, emotional and motivational.

Further studies show that many of these differences are also related to students' perceptions of the disciplinary climate in the classroom and the support their teachers give them. The data also show that parents help make a big difference on school outcomes and effective placement in the school setting. In further study will be interesting to investigate these issues.

To improve the performance of immigrant students it is important to provide additional support for disadvantaged students and schools, ensure availability and participation in extracurricular activities, involve parents in extracurricular activities, make available language training, training a teaching staff aware of the diversity that is able to support everyone the students.

Migration flows have changed a lot in recent years, this makes it absolutely necessary to monitor the integration of non-native students.

It is necessary to have datasets and information on immigrant students available, the INVALSI and ISTAT data allow us to analyze the situation of the students in depth.

ISBN 9788835123330

The Sustainable Development Goals (SDGs) allow us to have a large dataset that is constantly updated. This allows to analyze the data in depth and for each goal of the 2030 agenda to obtain interpretation of the results useful for the world of school and the scientific community. Specifically for objective 4, the possibility of integrating data from various sources (as in this work with the OECD PISA 2018 databases, we refered to student questionnaire and cognitive data) provided a very complete interpretative picture of the situation of immigrant students in Italy.

The variable examined, the choice of upper secondary school, represents a crucial step in the Italian educational path, it is strongly associated with skills and competencies. Furthermore, the national test results (INVALSI, 2019) show that the distribution of students between the different upper secondary school tracks is decidedly asymmetric: technical institute and vocational institute are characterized by a percentage of presence of immigrant students to a much greater extent than Lyceum, highlighting a school system that tends to segregation rather than inclusion.

Unfortunately, OECD-PISA is a cross sectional survey, the individual context variables still explain a limited portion of the variability in the secondary school attended.

Having the opportunity to integrate different surveys will allow us to overcome this limit and deal with the issue of immigrants in depth.

References

Agasisti T., Cordero-Ferrera J.M. (2013), *Educational disparities across regions: A multilevel analysis for Italy and Spain Journal of Policy Modeling*, 35, 6, pp. 1079-1102.

Agasisti T. *et al.* (2018), "Academic resilience: What schools and countries do to help disadvantaged students succeed in PISA", *OECD Education Working Papers*, 167, OECD Publishing, Paris, retrieved on January, 21, 2020, from http://dx.doi.org/10.1787/e22490ac-en.

Baker D., Goesling B., Letendre G. (2002), "Socio-economic Status, School Quality and National Economic Development: A Crossnational Analysis of the 'Heyneman-Loxley Effect' on Mathematics and Science Achievement", *Comparative Education Review*, 46, 3, pp. 291-312.

Beaumont-Walters Y., Soyibo K. (2001), "An Analysis of High School Students' Performance on Five Integrated Science Process Skills", *Research in Science & Technological Education*, 19:2, pp. 133-145.

Boyden J, Mann G. (2005), "Children's risk, resilience, and coping extreme situations", in M. Ungar (*ed.*), *Handbook for Working with Children and Youth: Path-*

ISBN 9788835123330

ways to Resilience Across Cultures and Contexts, Sage, Thousand Oaks (CA), pp. 3-26.
Buchmann C. (2002), “Measuring family background in international studies of education: Conceptual issues and methodological challenges”, in A.C. Porter, A. Gamoran (*eds.*), *Methodological Advances in Cross-National Surveys of Educational Achievement*, National Academy Press, Washington DC, pp.150-197.
Cerna L. (2016), *Immigration Policies and the Global Competition for Talent*, Palgrave Macmillan, Basingstoke.
Commissione europea/EACEA/Eurydice (2019), *Integrazione degli studenti provenienti da contesti migratori nelle scuole d’Europa: politiche e misure nazionali. Rapporto Eurydice*, Ufficio delle pubblicazioni dell’Unione Europea Lussemburgo.
Crane J. (1996), “Effects of home background, SES, and maternal test scores on mathematics achievement”, *Journal of Educational Research*, 89, 5, pp. 305-314.
Davoli M. (2016), “Scelte scolastiche degli immigrati: fattori socio-economici o di identità etnica? Un’analisi dei dati PISA 2012”, in L. Palmerio (a cura di), *OCSE PISA 2012. Contributi di approfondimento*, FrancoAngeli, Milano.
European Commission (2014), *Immigrant background and expected early school leaving in Europe: evidence from PISA*, retrieved on January, 21, 2020, from https://publications.jrc.ec.europa.eu/repository/bitstream/JRC109065/jrc109065_techbrief_migesl_180202final.pdf.
Feinstein L., Duckworth K., Sabates R. (2008), *Education and the Family: Passing Success across the Generations*, Routledge, London.
Ferreira F.H.G., Gignoux J. (2014), “The measurement of educational inequality: Achievement and opportunity”, *The World Bank Economic Review*, 28, 2, pp. 210-246.
Heath A., Kilpi-Jakonen E. (2012), “Immigrant Children’s Age at Arrival and Assessment Results”, *OECD Education Working Papers*, 75.
Hippe R., Jakubowski M., Araújo L. (2018), “Regional inequalities in PISA: the case of Italy and Spain”, *JRC Technical Reports*, retrieved on January, 21, 2020, from https://publications.jrc.ec.europa.eu/repository/bitstream/JRC109057/techbrief_pisareg_esit_final_1.pdf.
Hanushek E.A., Woessmann L. (2007), ”The Role of Education Quality for Economic Growth”, *Policy Research Working Paper*, 4122, World Bank, Washington DC.
INVALSI (2019), *Le prove INVALSI 2019*, retrieved on January, 21, 2020, from https://www.invalsiopen.it/risultati/rapporto-prove-nazionali-invalsi-2019/.
IOM (2015), *IOM and UNICEF Data Brief: Migration of Children to Europe*, International Organization for Migration, retrieved on January, 21, 2020, from www.iom.int/ sites/default/files/press_release/file/IOM-UNICEF-Data-Brief-Refugee-and-Migrant-Crisis-in-Europe-30.11.15.pdf.
Isphording I.E. (2015), “Language and Labor Market Success”, in *International Encyclopedia of the Social & Behavioral Sciences (Second Edition)*, Elsevier, Amsterdam, pp. 260-265.
Isphording I.E., Otten S. (2019), “Linguistic Distance and the Language Fluency of Immigrants”, *Ruhr Economic Paper*, 274, pp. 5-37.

ISBN 9788835123330

Jæger M.M., Breen R. (2016), "A dynamic model of cultural reproduction", *American Journal of Sociology*, 121, 4, pp. 1079-1115.

Masten A.S. (2013), "Risk and resilience in development", in P.D. Zelazo (*ed.*), *Oxford handbook of developmental psychology*, Oxford University Press, New York, pp. 579-607.

Masten A.S., Powell J.L., Luthar S.S. (2003), "A resilience framework for research, policy, and practice", in S. Luthar (*ed.*), *Resilience and Vulnerability: Adaptation in the Context of Childhood Adversities*, Cambridge University Press, New York, pp. 1-25, retrieved on January, 21, 2020, from https://doi.org/10.1017/CBO9780511615788.003.

Masten A., Garmezy N. (1985), "Risk, vulnerability, and protective factors in developmental psychopathology", in B. Lahey, A. Kazdin (*eds.*), *Advances in Clinical Child Psychology*, Plenum Press, New York, Vol. 8, pp. 1-52.

Martin M.O., Mullis I.V.S., Foy P., Stanco G.M (2012), *TIMSS 2011 International Results in Science*, TIMSS & PIRLS International Study Center, Boston College, Chestnut Hill (MA).

Mullis I.V.S., Martin M.O., Foy P., Arora A. (2012), *TIMSS 2011 international results in mathematics*, TIMSS & PIRLS International Study Center, Boston College, Chestnut Hill (MA).

OECD (2007), *Education at a Glance. OECD Indicators*, OECD, Paris.

Turmo A. (2007), "Scientific Literacy and Socio-economic Background among 15-year-olds. A Nordic Perspective", *Scandinavian Journal of Educational Research*, 48, 3, pp. 287-305.

OECD (2012), *Untapped Skills: Realising the Potential of Immigrant Students*, OECD Publishing, Paris.

OECD/European Union (2014), *Matching Economic Migration with Labour Market Needs*, OECD, Paris, retrieved on January, 21, 2020, from http://dx.doi.org/10.1787/9789264216501-en.

OECD/European Union (2015), *Indicators of Immigrant Integration 2015: Settling*, OECD Publishing, Paris, retrieved on January, 21, 2020, from http://dx.doi.org/10.1787/9789264234024-en.

OECD (2016), *International Migration Outlook*, OECD Publishing, Paris, retrieved on January, 21, 2020, from http://dx.doi.org/10.1787/migr_outlook-2016-en.

OECD (2017a), *Scaling Procedures and Construct Validation of Context Questionnaire Data. Ch. 16*, OECD Publishing, Paris.

OECD (2017b), *PISA 2015 Technical Report*, OECD Publishing, Paris.

OECD (2019a), *PISA 2018 Assessment and Analytical Framework*, PISA, OECD Publishing, Paris, retrieved on January, 21, 2020, from https://dx.doi.org/10.1787/b25efab8-en.

OECD (2019b), *PISA 2018 Results (Volume II): Where All Students Can Succeed*, PISA, OECD Publishing, Paris, retrieved on January, 21, 2020, from https://doi.org/10.1787/b5fd1b8f-en.

OECD (2020), "Immigrant students' attitudes and dispositions", in *PISA 2018 Results (Volume II): Where All Students Can Succeed*, OECD Publishing, Paris.

ISBN 9788835123330

Sandoval-Hernandez A., Cortes D. (2012), *Factors and Conditions that Promote Academic Resilience: A Cross-Country Perspective*, Paper presented at the annual meeting of the 56th Annual Conference of the Comparative and International Education Society, Caribe Hilton, San Juan, Puerto Rico.
Scharenberg K., Rudin M., Mueller B., Meyer T. (2014), *Education Pathways from Compulsory School to Young Adulthood: The First Ten Years*, TREE, Basel.
Suarez-Orozco C., Suarez-Orozco M.M. (2001), *Children of immigration*, Harvard University Press, Cambridge (MA).
Sutton A., Soderstrom I. (1999), "Predicting elementary and secondary school achievement with school-related and demographic factors", *Journal of Educational Research*, 92, 6, pp. 330-338.
UNESCO, EFA Global Monitoring Report (2015), *Education for All: Achievements and Challenges: 2000-2015*, UNESCO, Paris.
UNICEF (2016), *Uprooted: The Growing Crisis for Refugee and Migrant Children*, UNICEF, New York.
White K.R. (1982), "The relation between socioeconomic status and academic achievement", *Psychological Bulletin*, 91, 3, pp. 461-481.
Willms J.D. (2006), "Learning Divides: Ten Policy Questions about The Performance and Equity of Schools and Schooling Systems", *UIS Working Paper*, 5, UNESCO Institute of Statistics, Montreal, Canada.
Wong D.F.K. (2008), "Differential impacts of stressful life events and social support on the mental health of mainland Chinese immigrant and local youth in Hong Kong: A resilience perspective", *British Journal of Social Work*, 38, 2, pp. 236-52.

ISBN 9788835123330

6. INVALSI tests and the Italian territory: a comparison between native and foreign students of grade 8

by Jana Kopecna, Francesca Leggi, Maria Carmela Russo

In September 2015, the governments of the 193 UN member countries signed an action plan Agenda 2030, adopting a set of 17 goals as the blueprint for a global sustainable development. Eradicating extreme poverty, fighting inequalities and injustices are some of the goals that states have committed to achieve by 2030. Among these, goal number 4 aims to provide quality, equitable and inclusive education, and lifelong learning opportunities for all. The goal is then concretized into a sub-goal: to eliminate gender inequalities in education by 2030 and to ensure equal access to all levels of education and vocational training of protected groups, including persons with disabilities, indigenous populations and children in vulnerable situations.

The ISTAT's monitoring of the Agenda objectives (2020b) places Italy among the last positions in Europe according to the number of graduates, dropout rates and skills. The separating line between Southern and Central-Northern Italy concerns a wide range of phenomena and represents an emblematic case throughout the European Union. Undoubtedly, it constitutes a leitmotif of economic and social research in the country.

The INVALSI National Learning Surveys point out that in the South there are not only less satisfactory results, but also a lower ability to ensure the same educational opportunities for all students.

The variability of results among schools and between classes, an indicator of the equity of the educational system, is higher in the South and the Islands since primary school.

In addition to the analyses on the divergence dynamics among the Italian macro-regions, firstly along the North-South axis and then through analyses on Third Italy. In recent years, the attention has returned also to the development trajectories within each region and macro-region in more analytical terms.

ISBN 9788835123330

In this regard, the concept of inner areas seems to be an effective tool for understanding territorial differences. These areas represent the part of the Italian territory that is characterized by the significant distance from the centres of the essential services supply, and often it is affected by a long process of demographic decline and economic marginalization.

Within this scenario, the growing presence of foreign students on the national territory reveals a condition of inequality attributable not only to linguistic difficulties, but also to the socio-economic background of the families of origin. These differences vary geographically, with a greater incidence of the foreign population in the most dynamic and developed areas of the country. Our research focuses on the gap between the learning levels of immigrants and Italian students. In this paper, we will study differences between native and foreign students in the last year of lower secondary school (grade 8) in the 2018/19 academic year with the aim not only to photograph the phenomenon as a whole, but to highlight the geographical areas of possible intervention.

1. Introduction

To provide quality, fair and inclusive education, the school system must be able to provide all students with the same learning opportunities and benefits, regardless of their initial condition.

In contrast, the national and international literature (INVALSI, OECD PISA, IEA PIRLS) is rich in studies that highlight significant differences in school results linked to different geographical contexts, gender, ethnicity and background of students.

Being able to reduce inequalities and provide equal access to educational services means setting the stage for future citizens who will be able to actively participate in the life of their country.

The school is an agent of socialization in which future citizens acquire knowledge and skills and education contributes decisively to self-realization and it is closely related to the national social and economic level (UNESCO, 2015).

In this context, the paper aims to photograph the learning results of foreign students, a component that has recently become an important part of the Italian school system. By analyzing both Italy-born and recently immigrated students, it highlights the characteristics of the phenomenon on the national territory. Nevertheless, the distribution of foreign students and the performance of the education system appear to vary regionally along the classic

ISBN 9788835123330

North-South axis. Thus, it is important to examine the geographical dimension of learning and its local characteristics.

The migration processes that have affected the country over the past 25 years have brought an exponential increase of students with non-Italian citizenship. The proportion on the total number of students grew from 6.4% in the academic year 2007/2008 to 9.7% in the academic year 2018/2019 (ISMU, 2019). The latest available data of the Ministry of Education (MIUR, 2019) report 842,000 foreign students enrolled from kindergarten to upper secondary education.

Hence, the multicultural evolution of the Italian school system is a fertile and relatively recent field of study to understand the educational experience of foreign students from both the school performance and integration point of view. In fact, it can be said that one of the goals for the success of a migratory process is the success of the young children of immigrants (Crul and Vermeulen, 2003; Simon, 2003; Zhou, 1997).

For this reason, the analysis focuses on academic results in Italian language, available from the INVALSI standardized tests, since the competence in the language of the host country is fundamental in the perspective of integration. Mastering the local language does not only mean learning words. It entails a more important meaning: it allows to have relationships, to orientate and act effectively in different situations. Moreover, it represents a tool of meeting with other people to overcome stereotypes and prejudices, and to actively exercise their citizenship rights.

In the study we used the results of Italian tests carried out by students at the end of the lower secondary education (hence grade 8). From the school and social integration point of view, examining grade 8 is particularly significant because the student has to decide if continue studying or not. An important decision that will certainly influence his/her future. Not only the academic future, but also the labour one and the future as a citizen. If it is true that the family plays a fundamental role in this decision, the achieved academic results represent one of the starting points for the decision to be taken.

In this perspective, it is important to investigate the learning differences between foreign students, both first and second generation, and their native companions[1] (INVALSI, 2019). Thus, the analysis of INVALSI data allows us to shed light on those aspects in which educational institutions fail in the

[1] By Native student we mean a child born in Italy to parents born in Italy, a First-generation foreign student means a child born in a foreign state and arrived in Italy at school age, and a Second generation foreign student means a child born in Italy to parents born in a foreign state.

ISBN 9788835123330

integration process. To know where to intervene to reduce the gap in the results means having the opportunity to find solutions.

The contribution is articulated in this way: the second paragraph illustrates the concept of inner areas; the third paragraph allows to explore quantitatively and qualitatively the presence of foreign students on the Italian territory; and the following paragraphs are those dedicated to the part of the data, the research methods and the results achieved.

2. North-South axis and Inner areas: two territorial reading keys on students' learning

The different findings on the Italian education system describe Italy as a country furrowed by profound geographical differences. The analysis of the territorial gaps, especially following the comparison between central-northern regions and regions of the South, represents a *leitmotiv* of the official statistics related to a variety of socio-economic phenomena. In fact, ISTAT's monitoring of Objective 4 of the Agenda 2030 (ISTAT, 2020b) places Italy in the last positions in Europe by the number of graduates, drop-out rates and mathematical, scientific and reading skills and, in the national context, highlights wide disparity of success in the educational path. Save the Children (2020) highlights that this condition has even worsened following the effects of the Covid-19 emergency. The analysis of the educational risk index calculated on the basis of socio-economic and educational development or deprivation indicators shows a highly bipolarized territorial situation where the greatest vulnerability affects the South and the Islands (Save the Children, 2020).

By monitoring the general trend of the level of performance in Italian detected by the INVALSI tests, the territorial differences emerge as the school grade increases. In detail, in the second grade of primary education, the scores at the regional level are fairly uniform, with the results above the national average in regions of Valle d'Aosta, Umbria, Marche, Molise and Basilicata. At grade 5, some differences begin to emerge between the macro-areas, with the scores above the national average in the North-West and significantly lower scores in the South and Islands. This results variability expands further at grade 8, highlighting even more visible regional differences with positive results for the northern part of the country (especially for all the regions of the North-East and for the regions of Valle d'Aosta and Lombardy in the North-West). The negative results are registered in the southern part of the country, particularly in Campania, Calabria, Sicily and

ISBN 9788835123330

Sardinia. At the secondary education level, the regional gap becomes even sharper, with scores gradually decreasing from North to South at grades 10 and 13. Analyzing the results by type of school, they follow the trend of the geographical divisions (*ibid.*).

In the last decade, along with the classic interpretation of the macro-areas, the interest in the classification of the inner areas in the analysis of the educational system has increased. Since the Italian territory is highly diversified geographically, and from the socio-economic and demographic point of view, the inner areas offer a complementary perspective to the classic North-Central-South division. Thus, the areas of the country at the greatest risk of marginalization are identified and concrete actions can be taken.

The Italian inner areas:

– are significantly distant from the main centres offering essential services (education, health and mobility);
– have important environmental resources (water resources, agricultural systems, forests, natural landscape and human resources) and cultural resources (archaeological heritage, historical settlements, abbeys, small museums and trade centres);
– are a deeply diversified territory, as a result of the dynamics of the various and differentiated natural systems and the peculiar and century-old processes of anthropization (Barca, Casavola and Lucatelli, 2014).

This mapping of the territory is based on the identification of centres offering various services that are surrounded by other areas with different levels of peripherality. Along with the definition of these centres, it is possible to distinguish four other areas according to the level of accessibility (that are calculated in terms of travel minutes from the nearest pole centre). Hence, there are *pole belt* areas (less than 20 minutes from the nearest pole municipality), *intermediate* areas (between 20 and 40 minutes from the pole), *peripheral* areas (between 40 and 75 minutes from the pole) and *ultra-peripheral* areas (over 75 minutes from the pole). The last three categories form the inner areas which represent 53% of Italian municipalities (4,261) and 23% of the Italian population (approximately 13,540 million inhabitants) distributed over 60% of the national area (*ibid.*).

The main demographic dynamics concerned by these areas include the decrease of the population in the poles (in the peripheral and ultra-peripheral municipalities) and also the increase of the population over 65 (especially in the inner areas of the Centre-North). The presence of the foreign population equalizes these trends, because of their ongoing growth (*ibid.*; Luisi, 2018).

In this context, a school in the inner areas plays a fundamental role since it should consolidate its connection with the reality of the territory, respond-

ISBN 9788835123330

ing in a targeted manner to the needs of students. Specifically, it is desirable to implement the inclusive educational strategies, because they can strengthen the sense of community, and are being particularly important for students with a migration background (Osservatorio povertà educative, 2019).

This contribution has two main objectives:

– analyze the differences between Italian and foreign students in the results of the INVALSI Italian tests, in a perspective of the geographical areas, according to the North-South trajectory;
– investigate these differences by taking into account the classification of the Inner areas.

3. Foreign students in the Italian school and on the territory

The presence of foreign students in the Italian educational system is a consolidated, but not one-sided reality (see par. 1). Under the concept of first- and second-generation immigrants there is not only the different birthplace, the homeland or the place chosen by their parents, but an experiential background related also to different scholar needs.

First-generation foreigners are children who often arrive in Italy because of the family reunion and enter the school system at an advanced age. Language is the biggest obstacle they meet, causing the school performance differences when compared to their classmates. The data published in the report of the Ministry of Education *Pupils with non-Italian citizenship* (MIUR, 2019), referring to the academic year 2017/2018, showed that about 63% of foreign students were born in Italy belonging, therefore, to the second-generation foreigners. They are children who speak Italian as their first language, attend the same educational facilities as their peers, socialize with peers and live their own experiences. Although, the data relating to school performance are not reassuring. In fact, the disadvantage of the children of immigrants compared to natives is widely documented in literature (Heath and Brinbaum, 2007; Heath, Rothon and Kilpi, 2008; INVALSI, 2019).

The Report of MIUR allows us to frame the phenomenon of the presence of foreign students living in Italy from two other points of view: the nationality of origin and the distribution on the territory.

From more than 200 present nationalities, most of them come to Italy from areas of historical emigration such as Romania, Albania and Morocco, respectively with 18.8%, 13.6% and 12.3% of students with non-Italian citizenship. Over the previous year, Chinese students registered the most significant increase (+8%).

ISBN 9788835123330

However, the distribution of foreign students across the territory is not uniform. The concentration is very high in the North-West 38.3%, in the North-East 25.7% and in the Centre 23.1% and drops below the national average in the South and the Islands with 9% and 3.9%, respectively.

Lombardy is the region with the highest number of non-native students: it represents a quarter of the total present in Italy (about 213 thousand). Emilia Romagna, on the other hand, is the one with the highest percentage of foreign students of the total of the regional school population (16.1%), followed by Lombardy (15.1%), Tuscany (13.8%), Umbria (13.7%), Veneto (13.3%) and Piedmont (13.2%). Among the southern regions, the percentages of the presence of foreign students range from 2.5% in Sardinia to 4.3% in Calabria.

Looking at the provincial data, the highest presence of foreign population is in Milan with 10.6%, followed by Rome and Turin with 7.4% and 4.7%, respectively. The top ten provinces collect 41.4% of the total. If we consider the incidence in relation to the local school population, at the first position is Prato with 26.1%, followed by the provinces of Piacenza (22.2%), Mantua (18.4%), Cremona, Brescia, Asti, Parma (17.8%).

The Report suggests that at the municipal level some communities are particularly rooted in certain areas. In some municipalities of Lazio, such as Tivoli and Guidonia, 65-70% of foreign students are students of Romanian citizenship. In some municipalities of Emilia Romagna (Sassuolo, Vignola, Cento and Imola) the percentage of Moroccan students varies between 24% and 40% of the total of foreign students. The Chinese are present massively (between 40% and 60% of foreign students) in various Tuscan municipalities, including Campi Bisenzio, Prato, Fucecchio. Finally, some municipalities in Lombardy stand out for the strong presence of Egyptian students (Milan, Sesto San Giovanni, Cinisello Balsamo, Vigevano, Pioltello and Cologno Monzese).

Another crucial aspect concerns foreign students' distribution between schools, within schools and between classes. MIUR has set the percentage of students with non-Italian citizenship and with reduced knowledge of the Italian language enrolled in each class at 30%. However, schools cannot refuse the registration of a minor because of exceeding this quota. Starting from kindergartens, there are schools exceeding this threshold (mainly in Lombardy, Emilia Romagna and Veneto); and others, 729, in which foreign pupils exceed 50%. There is also 18.6% of schools that have no foreign student while 60% of schools have a percentage of students with non-Italian citizenship up to 15%.

At a regional level, the schools that exceed the expected quota are those that are characterized by the high percentage of foreign presence (Lombardy, Emilia Romagna and Veneto).

ISBN 9788835123330

4. Data and methods

Within the presented scenario, the study was conducted at the national level. The analysis was based on the results obtained in the Italian test carried out by the student population of grade 8 that participated in the national surveys INVALSI in the school year 2018/2019. The dataset is composed of 542,626 cases.

The database has been set up to conduct the multinominal logistic regression at the individual level, and the spatial analysis at the territorial level.

The first analysis measures the effect of each independent variable on the student's performance at the standardized test. The independent variable was the score in Italian Test (WLE ITA)[2], origin, gender, indicator of the socio-economic and cultural status of the student – ESCS[3], Geographic area and Internal areas, used as control variables.

Logistic regression is based on logit, the natural logarithm of the probability ratio between the two modes of the dependent variable:

$$Logit(P) = 1\text{n}(P/1 - P)$$

After the logistic transformation the equation becomes:

$$1\text{n}(P/1 - P) = \alpha + b_1x_1 + b_2x_2 + \ldots\ b_ix_i + b_nx_n$$

In the models, the change of the logarithm of the probability ratio (logit) is estimated as a result of the change of a unit in the independent variables.

The change of the logarithm of the probability ratio is equivalent to the logarithm of the association ratio (odds ratio) (Di Franco, 2017, pp. 242-244).

In this research we used four different techniques of spatial analysis:1) Multivariate Clustering method; global and local indicators of spatial autocorrelation, specifically 2) Global Moran's I and 3) Local indicator of spatial association (LISA); and 4) an Ordinary Least Square (OLS) regression model.

To introduce the geographical perspective, the k-means clustering method was applied on the selected characteristics. This method look for a fixed number of clusters *k* in a data we aim to analyze. We used the tool of the

[2] WLE is a numerical score on a quantitative scale that estimates skills according to the Rasch model (INVALSI, 2019).

[3] The calculation of the ESCS is based on discrete indicators such as the level of education, the employment of parents and the presence of the material conditions in which the student lives outside the school (Campodifiori *et al.*, 2010).

ISBN 9788835123330

spatial software package “Multivariate Clustering” that permit us to have a brighter view on the whole territory.

Global Moran’s I Index is one of the oldest and most widely used indicators of spatial autocorrelation (Moran, 1950). It compares the value of the variable in one unit with the value in all other units. The formula is as follows:

$$I_{ij} = \frac{n}{\sum_i \sum_j w_{ij}} \frac{\sum_i \sum_j w_{ij} (x_i - \bar{x})(y_j - \bar{y})}{\sum_i((x_i - \bar{x})) \times \sum_j \left((y_j - \bar{y})\right)}$$

where:

x_i	is the value of the observed variable x in unit i;
y_j	is the value of the observed variable y in unit j;
w_{ij}	is the spatial weight between a unit i and j (a measure of spatial contiguity between the units i and j);
n	is the total number of units;
$(x_i - \bar{x})$	is the deviation of a variable value for unit i from its mean.

The range of values of global Moran’s I is between -1 (perfect dispersion) and 1 (perfect correlation).

The concept of the bivariate local Moran’s $I_{B,i}$ is very similar to its global counterpart (Anselin, 1995). Essentially, it captures the relationship between the value for one variable x in unit i, and the average of the adjacent values for another variable y (its spatial lag $\Sigma_j w_{ij} y_j$). The statistic is a product of x_i with the spatial lag of y_i, with both variables standardized (their means are zero and variances equal one):

$$I_{B,i} = c x_i \Sigma_i w_{ij} y_i$$

For the correct working of the spatial autocorrelation methods, it is necessary to construct a weight matrix that defines a neighbourhood of the observed unit. We used the arc-distance.

To understand factors associated with spatial clustering of the score in Italian, we applied OLS regression to explain the global relationship between dependent and independent variables. The OLS regression model can be expressed as:

$$y = \beta_0 + \sum_{i=1}^{k} \beta_i x_i + \varepsilon, \quad \varepsilon \sim N(0, \delta^2)$$

ISBN 9788835123330

where:
x_i and y_i are, respectively, the independent and dependent variables;
k is the number of independent variables;
β_0 is the intercept;
β_i is the parameter estimate (coefficient) for the independent variable x_i;
δ is the error term.

All analyses and mapping in this study were implemented using ArcGIS and GeoDa software packages.

5. Data analysis

5.1. Logistic regression

Firstly, we aim to analyze the school performance (based on the INVALSI Italian tests at grade 8) on an individual level using a logistic regression model. Four variants of the model have been developed, considering as predictors both the students' individual characteristics (gender, origin, ESCS, regularity) and, subsequently, the territorial variables (classification of the inner areas and geographical divisions).

In this way, it is possible to observe the relative likelihood of a student with certain characteristics (included in the model) to get an Italian test score above or below the national average[4].

The overall results of the models confirm findings from the literature (among all: Heath, Rothon and Kilpi, 2008; Murineddu, Duca and Cornoldi, 2006; Schmid, 2001): between the natives and foreigners persist the constant disadvantage of the latter in achieving high levels of learning in Italian.

In the first model, we note how the probability of obtaining a better result in the Italian test is the prerogative of native female students. If we add the ESCS variable (model 2), the relative advantage of native students get slightly lower and at the same time, the disadvantage for foreigners of the first generation decreases.

The indicator of social, economic and cultural status plays an important role in school performance. To understand how much the school system helps in achieving certain results, it is necessary to consider the starting point of the students. Its importance confirms also the decision to implement the

[4] The national average for grade 8 Italian tests carried out in 2019 is set at 199 (INVALSI, 2019).

ISBN 9788835123330

Intervention Plan for the reduction of territorial gaps in education[5] an intervention program launched by the Ministry of Education in collaboration with INVALSI, INDIRE, the Offices Regional Scholastics and other local authorities in January 2020.

The transition from model 2 to model 3 shows the impact of the inner areas variable: there is a moderate advantage in residing in areas belonging to the pole or belt, showing almost equal values (around 1.3). It should be noted again that the weight of the ESCS is considerable, indicating for the *high* modality a very high probability of the student obtaining a score above the national average.

In the last model, in addition to the individual variables, are involved also the variables related to the territory, inner areas and geographic areas. In this case, the geographical area variable has a very strong impact, showing a high probability for North West students to achieve a good performance in the INVALSI test. At the same time, in terms of advantage, the percentage difference between a native student and a foreign student of the second generation increases.

Thus, it can be said that assuming a territorial perspective of interpretation, a student living in the Inner Areas has a lower probability of having a score above the national average when compared to the other students living in less peripheral areas. This disadvantage intensifies moving from North to South and it is even more evident for foreign students.

Table 1 shows the percentage of correctness and the Nagelkerke R^2 of the four logistic regression models. The first value is a measure of the predictive capacity of the models, that is, the ability to assign correctly the modality of the dependent variable assumed by each case. Nagelkerke R^2, on the other hand, is useful for observing the ability to reproduce conditional probabilities and measure the variance of dependent variables explained by predictors.

Tab. 1 – Parameters to evaluate models' quality

	Percentage of correctness	*Nagelkerke R^2*
Model 1	56.7	0.043
Model 2	63.8	0.139
Model 3	64.2	0.150
Model 4	65.1	0.169

[5] In orig. *Piano di Intervento per la riduzione dei divari territoriali in istruzione*.

ISBN 9788835123330

The inclusion of the control variables has produced good results, going from 56.7% of the first model correctness up to 65.1% of cases classified in the fourth model.

Table 2 shows the values of the parameter Exp (B) obtained from the logistic regression models. This parameter can be used for the interpretation of the relationship between the variables: it expresses the variation of the dependent variable given by the student's performance in the INVALSI Italian test, as a function of variations of the independent variables.

Tab. 2 – Multivariate logistic regression analysis

Above the National average					
Intercept		-,888	-1,471	-1,460	-1,891
Variables	*Modality*	*Model 1*	*Model 2*	*Model 3*	*Model 4*
Gender	Female	1.531***	1.563***	1.530***	1.540***
	Male				
Origin	Native	2.242***	1.640***	1.541***	1.858***
	Generation I foreigner	0.544***	0.611***	0.758***	0.773***
	Generation II foreigner				
ESCS	High		5.257***	4.698***	4.282***
	Medium-high		2.862***	2.589***	2.392***
	Medium-low		2.096***	1.925***	1.800***
	Low				
Regularity	Non regular			0.389***	0.370***
	Anticipator/Regular				
Urban classif.	Pole areas			1.247***	1.119***
	Belt areas			1.241***	1.055***
	Inner areas				
Geographic area	North West Italy				2.020***
	North East Italy				1.963***
	Central Italy				1.685***
	South Italy				1.095***
	Insular Italy				

Note: ***, **, or * indicates p-value < 0.01, p-value < 0.05 or p-value < 0.10 respectively; reference category in italic.

Source: our processing on INVALSI data.

ISBN 9788835123330

5.2. Spatial analysis

As shown above, the academic performance of students is influenced by various factors concerning the individual and family background, the previous educational experience and the characteristics of the territory in which they live. In the previous paragraph (see paragraph 5.1), the academic performance in Italian INVALSI test was analyzed from the individual point of view, that is from the perspective focused on the micro-level. To complete the picture, we move also to the macro-level, where the territorial dimension of learning and its local characteristics are investigated. In this section, therefore, the attention is focused on the context in which students are situated, and the characteristics on the school level are explored through different techniques of spatial analysis.

First of all, the k-means clustering was applied, using as entry parameters the score in Italian test of grade 8, the percentage of total foreign students, and the ESCS at school level. As a result, four clusters were identified.

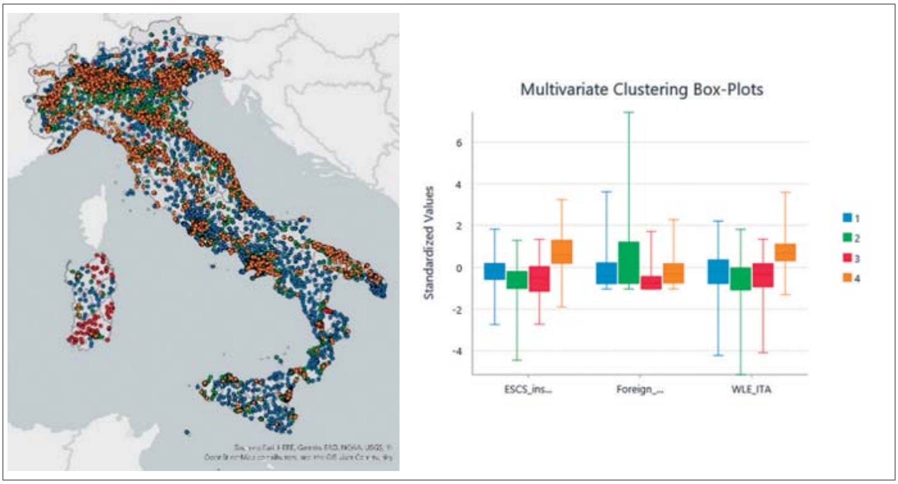


Fig. 1 – The map of Clusters and Box-plot identified by Multivariate Clustering

The first cluster (Cluster 1) is composed of schools with all the values slightly below the national average, reproducing visually the presence of the inner areas across the national territory. The second group (Cluster 2) is represented by schools with rather low ESCS and scores, and with a percentage of foreign students above the national average. The school belonging to this cluster are represented especially in northern Italy. The third cluster (Cluster 3) is composed of schools with low ESCS, worse performance in Italian test

ISBN 9788835123330

and a low proportion of foreign students, and they are located in the Islands and southern Italy. The last group (Cluster 4), is composed of schools with a low percentage of foreign students, and with the ESCS and scores above the Italian average. These schools are present mainly in the North and Centre of Italy, but some clusters are identified also in Apulia, Campania.

Consequently, the spatial pattern of the WLE score in Italian test across the Italian territory was analyzed by using global and local methods of spatial autocorrelation. In particular, we applied Moran's *I* Index and LISA. The univariate global Moran's I[6] of the score in Italian was calculated to evaluate its the spatial distribution, and the value $I \approx 0.359$ and Z-score[7] of 110.495 confirmed the presence of significant (p-value < 0.001) clustering of positive values of the score across the Italian territory.

To investigate if exists, for an observed unit (in our case a school unit), a relationship between the value of one variable and the value of another variable in an adjacent unit, the bivariate Moran's I was applied. We analyzed the spatial dependence between the score and different characteristics at school level – the ESCS ($I \approx 0.214$), the percentage of foreign students of the first generation ($I \approx 0.167$), the percentage of foreign students of the second generation ($I \approx 0.262$), the percentage of regular students ($I \approx -0.057$) and the percentage of females ($I \approx -0.002$). The indicator revealed a negative significant spatial association in the relationship between the score and the proportion of regular students and females. All other values suggest the existence of significant and positive spatial autocorrelation.

Since the global Moran's I does not provide the information on the significance of the identified spatial clusters, the univariate and the bivariate LISA[8] were applied. The results of the LISA analysis, measuring spatial clustering

[6] If the values in the dataset tend to cluster spatially, i.e. high values cluster close to other high values or low values cluster near other low values, then the global index will be positive (showing positive spatial autocorrelation). In the case the high values tend to be near low values, the indicator will be negative (showing negative spatial autocorrelation). Basically, the spatial distribution of high values and/or low values in the dataset is more spatially clustered than would be expected if underlying spatial processes were random.

[7] The Z-score is used to test the significance of any spatial autocorrelation, and it is simply a standard deviation. For each unit it is calculated the deviation from the mean and consequently, these deviation values for all adjacent units are multiplied together, resulting in the formation of a cross-product. Z-score greater than 2.58 and p-value = 0.01 indicate that there is a less than 1% likelihood that the clustered pattern could be the result of random chance.

[8] Spatial clusters identified by bivariate LISA include High-High clusters (high values of a variable x in a high-value y neighbourhood) and Low-Low clusters (low values of a variable x in a low-value y neighbourhood). Moreover, LISA can reveal the presence of High-Low/Low-High outliers – i.e. high/low values of x in a low/high y neighbourh.

ISBN 9788835123330

and heterogeneity of one variable (the score) or among the pairs of variables, are displayed in Figure 2.

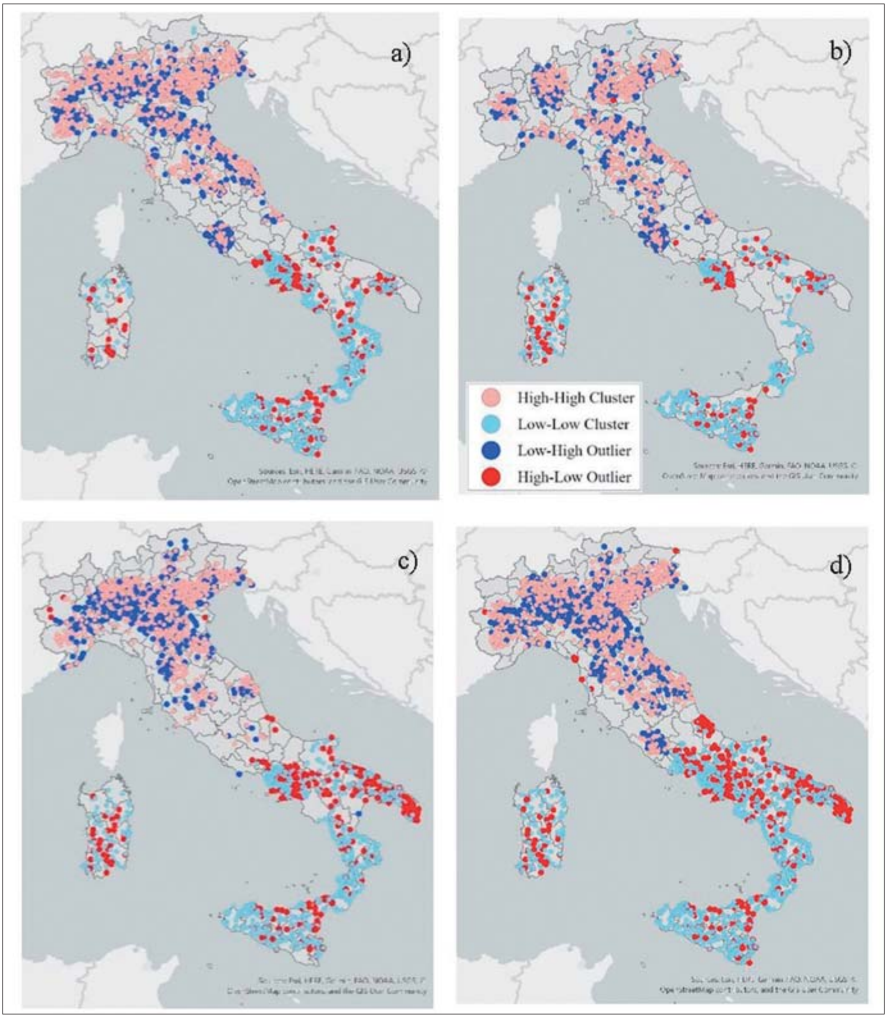


Fig. 2 – Univariate LISA Clusters and Outliers of a) the score WLE ITA; and Bivariate LISA Clusters and Outliers of the score WLE and: b) ESCS; c) the percentage of foreign students of I generation; and d) the percentage of foreign students of II generation

Since the values of Global Moran's I were very low for the pairs including the percentage of female and regular students and the results of bivariate LISA were prevalently not significant, we did not visualize the correspond-

ISBN 9788835123330

ing maps. Only the schools that had statistically significant values of spatial autocorrelation are visualized.

The High-High clusters indicate schools with high values of the score in Italian which are surrounded by schools whose value of another variable is above the national average. The Low-Low clusters indicate schools with low values of the score which are adjacent to schools with the value of another variable under the average. The Low-High outliers are schools with a low value of the score surrounded by schools with higher values of another variable. Finally, the High-Low outliers are schools with high values of the score which are adjacent to schools with the values of another variable below the national average.

The results confirm the existing territorial division described in detail in the literature (*ibid.*). While, in the northern and central Italy is undeniable the presence of clusters of schools with better performance in Italian test surrounded by schools with the scores above the national average, in the South and Islands prevail clusters of school having lower values of the score in Italian that are adjacent to schools with the scores below the Italian average. Similarly, the outliers with low values of the score neighbouring with schools with high values are located mainly in the Centre-North of Italy. On the other hand, the clusters with low values of the scores neighbouring with low values of the observed variable, and the outliers with high values of the score surrounded by schools with low values are situated in the regions of Southern Italy and Islands.

For the ESCS and the percentage of foreign students of the first and the second generation, similar patterns to results of spatial autocorrelation of the score in Italian are detected.

To better understand factors associated with spatial clustering of observed scores we applied multivariate linear regression to verify the statistical relationship between dependent variable (the score in Italian test) and selected independent variables. The calculation of the OLS models was realized by using the ArcGIS software package. In Model 1 the indicator ESCS, percentage of foreign students of I generation and percentage of foreign students of II generation were included, and in Model 2 we added also the percentage of regular students and the percentage of females. The VIF values in both models vary from 1.032 to 1.790, indicating that OLS estimations are not biased from multicollinearity. The Model 1 explains 52.502 % of the variance in the score, and in the second model this value increases to 53.222%, indicating slightly better goodness of fit.

Tab. 3 shows the results of two global regression models. In the first model (Model 1) all dependent variables are significant and only the percentage

ISBN 9788835123330

of foreign students of first generation has a negative impact on mean score in Italian. This explanatory variable is also the only one that is not significant in the Model 2. Except for the percentage of foreign students of I generation, all the variables result statistically significant. In both models, we can observe that the ESCS has a great weight, but relatively high is also a weight of the percentage of foreign students of second generation, indicating highly positive impact on the mean score.

Tab. 3 – Ordinary least squares (OLS) estimation

Variables	*Model 1*	*Model 2*
Intercept	197.410***	182.095***
ESCS	21.675***	20.813***
% of foreign students I generation	-14.951***	-8.735
% of foreign students II generation	19.424***	21.306***
% of female students		12.873***
% of regular students		10.098***
R-Squared	0.525	0.532
Adjusted R-Squared	0.524	0.532
AICc	39,665.034	39,584.867

Note: ***, **, or * indicates significance at 1%, 5% or 10% levels respectively; AICc = Akaike Information Criterion.

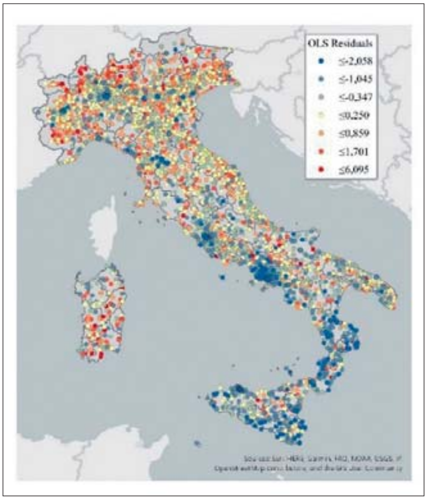


Fig. 3 – Residual maps of OLS model

ISBN 9788835123330

To visualize spatial clustering of the over- and under-predictions evident in residuals from the OLS regression, we used the univariate LISA analysis. Fig. 3 shows the residuals from OLS model (model 2) and the result of the LISA analysis created by plotting residuals of OLS regression model.

Being a global model, the OLS provides a general view of the relationships between the scores in Italian and other variables. The disadvantage of its use with the spatial data is that it considers all places with the same weight as if all places shared the same location. For this reason, we are aware of the fact that the geographically weighted regression would be a better tool to deepen the overall view we presented in this paper, but for now, we consider it as an indication for the future research directions.

6. Conclusions

The results, both of the logistic regression at the student level and the spatial regression at the school level, confirm what has been highlighted in the theoretical introduction. As expected, the worse performance results for the foreign students, who live in Southern Italy and have a low ESCS (the result of the logistic regression).

The macro-level perspective (the spatial analysis) also shows the peninsula cut in half: in the North, there are those schools in which the average Italian test score is above average, while in the South the most representative clusters are from schools with a score below the national average. A photograph that remains unchanged with the different control variables, for example, the ESCS and the origin.

Undoubtedly, the ESCS has an active role in determining school outcomes. Although, the results obtained in the logistic regression and the OLS show that it is not possible to summarize in a single variable the answer to why students' results differ or how important for the results' correct interpretation is the socio-economic background as well as the geographical residence.

Altogether, these non-cognitive variables are important because they help contextualise school results and allow to better identify groups for intervention purposes.

The MIUR project (see paragraph 5.1) makes use of this type of studies to achieve the objective of the equity in the education system and has launched a plan of interventions that is currently aimed at schools in difficulty in Calabria, Campania, Puglia, Sardinia and Sicily.

The school must be a social lift, regardless of the geographical location and the type of students enrolled.

ISBN 9788835123330

The empirical evidence presented up to this point allows to formulate some conclusive considerations on the school performance of grade 8 students in standardized tests. As we have observed, different factors play a role in the school performance differences between native and foreign students. They are related to the student's educational path, his family of origin, and the territorial characteristics. It should be emphasized that there is not only a gap in the INVALSI test scores between Italians and foreigners, but, more specifically, also a gap between 1st and 2nd generation of foreigners. In fact, the latter have characteristics that make them closer to their Italian peers.

A first dimension that characterizes the difference in school results between natives and foreigners concerns the language. The linguistic advantage of those born in the host country or who arrived in early childhood draws a line of distinction compared to those who arrived at school age or in adolescence. This type of knowledge plays a fundamental role in the educational path, promoting cultural comparison and personal enrichment (Heath, Rothon and Kilpi, 2008). In fact, thanks to an unstable identity profile, the student can adapt the spoken language according to the communicative context in which he/she finds himself (Bolognesi, 2008). Thus, we can speak of "multiple identities" (Valtolina and Marrazzi, 2006), considering their sense of belonging on the edge between the culture of the country of origin of the parents and the culture of the country in which they live.

A second interpretative hypothesis involves the element of stability represented by the family of origin. It could be assumed that a family unit that has been living in the host country for some years has consolidated its position in the labour market and improved its relationship network, extending it to subjects of different nationalities respect to the origin (Bertolini, Lalla and Toscano, 2010). The elements of discontinuity that distinguish the second from the first generation also bring with them a system of expectations that differentiates the children of immigrants from their parents. Consequently, on the one hand, young people can develop desires for social mobility, rejecting the modalities of subordinate integration to which their parents have often had to adapt. On the other hand, these ambitions can cause frustration because they are not always achievable due to the different economic and cultural endowments of the families of origin (Ambrosini and Molina, 2004). A central point for the educational/labour path of foreign students is the families' investment in their future. It is the secondary school choice that emphasizes the diversity of orientation respect to Italian students. In fact, while the latter often proceeds with the choice of high school to continue their studies even in the presence of not brilliant school performances, the children of foreigners, on the contrary, enroll in technical or vocational institutes (ISTAT, 2020).

ISBN 9788835123330

The issues discussed to date are intertwined with the characteristics of the school system. As the spatial analysis has shown, there are significant North-South territorial differences that often interfere with the school's intention to be a vehicle for socio-cultural integration (Barban *et al.*, 2008). Living in a socio-economically disadvantaged context or a peripheral territory far from the main services, such as the inner areas, can represent an additional burden for foreign students to achieve adequate levels of performance. In this context, the Italian school should remove these obstacles and strengthen inclusive education and intercultural learning paths, contributing consequently to the construction of the additional human capital essential for students from families with lower socio-economic endowments.

References

Ambrosini M., Molina S. (a cura di) (2004), *Seconde generazioni: un'introduzione al futuro dell'immigrazione in Italia*, Fondazione Giovanni Agnelli, Torino.

Anselin L. (1995), "Local indicators of spatial association – LISA", *Geographical Analysis*, 27, 2, pp. 93-115.

Barca F., Casavola P., Lucatatelli S. (a cura di) (2014), *Strategia nazionale per le aree interne: definizione, obiettivi, strumenti e governance*, *Materiali Uval*, 31, Roma.

Barban N., Dalla Zuanna G., Farina P., Strozza S. (2008), "I figli degli stranieri in Italia fra assimilazione e disuguaglianza", *Working Paper Series*, 16, Dipartimento di Scienze Statistiche, Università di Padova, retrieved on December, 30, 2020, from http://www.stat.unipd.it/ricerca/wp.

Bertolini P., Lalla M., Toscano V. (2010), "L'inserimento scolastico degli studenti stranieri di prima e seconda generazione in Italia", *Materiali di discussione*, 631, Dipartimento di Economia Politica, Modena.

Bolognesi I. (2008), "Identità e integrazione dei minori di origine straniera Il punto di vista della pedagogia interculturale", *Ricerche di Pedagogia e Didattica*, 3, *Educazione Sociale, Interculturale e della Cooperazione*.

Campodifiori E., Figura E., Papini M., Ricci R. (2010), Un indicatore di status socio-economico-culturale degli allievi della quinta primaria in Italia, *INVALSI Working Paper*, 2.

Crul M., Vermeulen H. (2003), "The second generation in Europe", *International Migration Review*, 37, 4, pp. 965-986.

Di Franco G. (2017), *Tecniche e modelli di analisi multivariata*, FrancoAngeli, Milano.

Luisi D. (2018), *Popolazione straniera nelle aree interne: argine allo spopolamento e risorsa per territori fragili*, retrieved on December, 30, 2020, from https://www.forumdisuguaglianzediversita.org/popolazione-straniera-nelle-aree-interne-argine-allo-spopolamento-risorsa-territori-fragili/.

ISBN 9788835123330

Heath A., Brinbaum Y. (2007), "Explaining ethnic inequalities in educational attainment", *Ethnicities*, 7, 3, pp. 291-305.
Heath A., Rothon C., Kilpi E. (2008), "The second generation in Western Europe: education, unemployment and occupational attainment", *Annual Review of Sociology*, 34, pp. 211-235.
INVALSI (2019), *Rapporto prove INVALSI 2019*, retrieved on December, 30, 2020, from https://INVALSI-areaprove.cineca.it/docs/2019/Rapporto_prove_INVALSI_2019.pdf.
ISMU (2019), *XXV Rapporto Ismu sulle migrazioni*, retrieved on December, 30, 2020, from https://www.ismu.org/wp-content/uploads/2018/10/Comunicato-Stampa-XXV-Rapporto-Ismu-sulle-Migrazioni-29_11_19-ok.pdf.
ISTAT (2020a), *Identità e percorsi di integrazione delle seconde generazioni in Italia*, Istituto Nazionale di Statistica, Roma.
ISTAT (2020b), *Rapporto SdGs 2020. Informazioni statistiche per l'agenda 2030 in Italia*, Istituto Nazionale di Statistica, Roma.
MIUR (2019), *Gli alunni con cittadinanza non italiana. A.s. 2017/18*, Ufficio Statistica e Studi, Roma.
Moran P.A.P. (1950), "Notes on continuous stochastic phenomena", *Biometrica*, 37, pp. 17-23.
Mullis I.V., Martin M.O., Sainsbury M. (2016), *PIRLS 2016 reading framework*, PIRLS, pp. 11-29.
Murineddu M, Duca V., Cornoldi C. (2006), "Difficoltà di apprendimento scolastico degli studenti stranieri", *Difficoltà di apprendimento*, 12, 1, pp. 49-70.
OCSE PISA (2018), *I risultati degli studenti italiani in lettura, matematica e scienze*, INVALSI, Roma.
OECD (2019), *PISA 2018 Results (Volume II) Where All Students Can Succeed*, OECD Publishing, Paris.
Osservatorio povertà educative (2019), S*cuole e asili per ricucire il paese*, *Openpolis*, retrieved on December, 30, 2020, from https://www.conibambini.org/wp-content/uploads/2018/10/Report-2019.pdf.
Save the Children (2020), *L'impatto del Coronavirus sulla povertà educativa*, retrieved on December, 30, 2020, from https://www.savethechildren.it/cosa-facciamo/pubblicazioni/impatto-del-coronavirus-sulla-poverta-educativa.
Schmid C.L. (2001), "Educational achievement, language-minority students, and the new second generation", *Sociology of Education*, 74, pp. 71-87.
Simon P. (2003), "Challenging the 'French model of integration': discrimination and the labour market case in France", *Studi Emigrazione*, 152, pp. 717-745.
UNESCO (2015), *Education for All 2000-2015: Achievements and Challenges*, UNESCO.
Valtolina G.G., Marrazzi A. (a cura di) (2006), *Appartenenze multiple. L'esperienza dell'immigrazione nelle nuove generazioni*, FrancoAngeli, Milano.
Zhou M. (1997), "Growing Up American: The Challenge Confronting Immigrant Children and Children of Immigrants", *Annual Review of Sociology*, 23, pp. 63-69.

ISBN 9788835123330

7. *Targeting students with high risk of dropping out of school: a latent profile analysis*

by Giuseppina Le Rose, Chiara Sacco

The Goal 4 of the 2030 Agenda for Sustainable Development launched by UN Member States in 2015 aims to "ensure an inclusive and equitable quality education and promote lifelong learning opportunities for all". In the last years, a worrying problem that requires an appropriate and effective political response is the significant number of early school leavers. It is well known that the transition between lower and upper secondary school is a difficult and delicate step of the student's life and it is essential to monitor the students in this phase in order to prevent the school dropout. Dropout prevention programs are intervention designed to keep students in school and encourage them to complete the upper secondary school. Usually these programs target student at-risk and provide them with several support services and activities. An efficient and early identification of the student at-risk it is necessary to target students destined to dropout prevention programs. This paper allows monitoring the transition of the students between the last year of the lower secondary school and the upper secondary school, using INVALSI data. We focus on the identification of the principal characteristics of the students at high-risk of dropping out, hypothesizing the presence of several high-risk students' profiles in terms of self-efficacy, demotivation, future expectations, relationship with parents and with peers and school performance. A latent profile approach is applied to assess and to model the different student's profiles and a multinomial model is used to examine the relationship between the profile membership and some demographic and socio-economic variables. The findings of this work could be useful for efficiently targeting students destined to dropout prevention programs. A wrong characterization of the students at-risk could lead to include in the intervention programs who would not have dropped out, missing who eventually do drop out and, so, could benefit of an intervention. For this rea-

ISBN 9788835123330

son, the results of this work could be an important help to design effective dropout prevention programs.

1. Introduction

The present study focuses on Goal 4 of 2030 Agenda for Sustainable Development launched by UN Member States in 2015. The SDG 4 aims to “ensure an inclusive and equitable quality education and promote lifelong learning opportunities for all”. Goal 4 is divided into ten targets that belong to three dimensions: equal opportunity to access to all levels of instruction (preprimary, primary, secondary and tertiary), possession of knowledge and skills for employment and sustainable development, elimination of gender inequalities, monitoring of school structures, so that they are suitable for everyone’s needs (ISTAT, 2019).

The process of monitoring and verifying the Development Goals represents a fundamental phase for the concrete implementation of the 2030 Agenda as it allows transforming SDGs in a national, regional and local policy management tool. For this reason, the United Nations Statistical Commission has set up the Inter Agency Expert Group on SDGs (UN-IAEG-SDGs) with the aim of identifying a common statistical information framework for all member countries in order to monitor progress towards Agenda objectives.

In Italy, the Italian National Institute of Statistics (ISTAT) plays a crucial role as referral agency for the production of the indicators for monitoring the SDG. In the latest report, ISTAT provided an updated set of 123 UN-IAEG-SDG indicators composed by 303 national statistical measures, of which 273 are different from the international ones. There is not a univocal correspondence between the indicators defined at international level and the measures identified for Italy.

In addition to the list of indicators proposed by the UN-IAEG-SDGs, the Bertelsmann Foundation and the Sustainable Development Solution Network (UN-SDSN) have created the Global SDG Index and the dashboards to summarizes countries’ current performance and trends on the 17 SDGs as agreed by the international community in 2015. All 17 goals are weighted equally in the Index: the score identifies the country’s position between the worst (0) and the best or target (100) outcomes. Italy’s overall Index score (75.8) suggests that the country is on average at 75.8% of the way to the best possible outcome across the 17 SDGs (for further details see https://sdsna.github.io/2019GlobalIndex/2019GlobalIndexRankings.pdf).

ISBN 9788835123330

Compared to the average of the other states in the OECD area, Italy has a 2.4% lower score. The SDG dashboards (available at https://dashboards.sdg-index.org/profiles/ITA) highlight strengths and weaknesses of each country on the 17 SDGs. They indicate the rating, which is represented by the colour of the SDGs boxes, and the trend of current Development Goals, which is depicted by the arrows next to the Objectives. For Italy, the orange colour of goal 4 boxes indicates that there are still significant challenges to be faced. The yellow colour of the arrow indicates that the score relative to the indicators of that goal has moderately increased but is still insufficient for its achievement.

One of the crucial indicators of Goal 4 concerns the proportion of the early leavers from education and training (ELET), defined as a person aged 18 to 24 who has completed at most lower secondary education and is not involved in further education or training. In Europe, although the ELET rates in most member states have dropped since the introduction of the EU 2020 strategy, there remain large disparities between gender, social and ethnic groups (Nouwen and Clycq, 2020). In Italy, in 2018, the early school leavers rate is instead gone back to 14.5%, returning to 2015 levels, with marked territorial differences (ISTAT, 2019).

The first critical element of the Italian school system and of the school dropout seems to concern the transition from lower secondary school to upper secondary school. Among the indicators reached in Goal 4 there is the lower secondary completion rate, that is about 99.8%. On the other hand, Ballarino and Schadee (2006) demonstrated that the probability of completing the lower secondary school is higher than the one of completing the upper secondary school.

In this study, we aim to monitor the transition of the students between the last year of the lower secondary school (8^{th} grade) and the upper secondary school in order to improve the efficacy of the dropout prevention program. A crucial question for planning an efficient prevention program is: which are the characteristics that accurately identify students more likely to drop out? We exploited the Student's Questionnaire administered by INVALSI in the school year 2017/2018 at the end of lower secondary school. The questionnaire collects information about the student's cultural background and explores the relationship of the student with the school from different perspectives. To identify the students at risk of dropping out, we combined the Student's Questionnaire of grade 8 with the INVALSI administrative data of student enrolment in the second year of upper secondary school (10^{th} grade) in the school year 2019/2020. Our hypothesis is that the students that have compiled the questionnaire of grade 8 in 2017/2018 and are not enrolled on

ISBN 9788835123330

the 10th grade in the school year 2019/2020, are those with a higher risk of dropping out. Moreover, we hypothesized that the students at risk of dropping out are not a homogenous group but is composed by different latent subpopulations characterized by different profile in terms of self-efficacy, demotivation, future expectations, relationship with parents and with peers and school performance. In order to improve the understanding of the phenomenon and to target student at-risk, we identified three different profiles of students at-risk of dropping out of school using a latent profile analysis. The cluster membership is examined with respect to different socio-demographic student's characteristics.

The efficient and early identification of the student at-risk is a powerful tool necessary to target students destined to dropout prevention programs. A wrong characterization of the students at risk could lead to include in the intervention programs who would not have dropped out, missing who eventually do drop out and, so, needs an intervention. For this reason, the findings of this work could be an important help to design effective dropout prevention programs.

2. Theoretical background

The dropout risk factors, connected to large disparities between gender, social and ethnic groups, can be found throughout international literature on ELET (Lamb and Markussen, 2011; Rumberger and Lim, 2008) and their continuing impact shows how resistant these systematic features seem to be. Even the few Italian studies that have analyzed this phenomenon highlight the impact of the socio-economic background factor (Ballarino *et al.*, 2010; Checchi and Flabbi, 2006; Mocetti, 2012) and the gender difference (Borgna and Struffolino, 2017). Data from Italian national student registry, collected annually by the Italian Ministry of Education, University and Research (MIUR), confirms that male students who live in the South of the country are at higher risk of dropout. The highest percentage of school dropouts is registered in the upper secondary school's study branches called IeFP, vocational training courses that are organized at regional level and last from two to four years, followed by vocational education, technical education, and lyceum (MIUR, 2019).

In addition to these dropout risk factors, called status (i.e. parental education, socioeconomic status, gender, family structure), there are the risk factors called alterable, such as academic failure, lack of attendance and motivation, misbehaviour correlated to dropout (Malmberg-Heimone *et al.*,

ISBN 9788835123330

2018). Numerous studies have highlighted that students at high dropout risk show lack of expectations of staying in school (Esch *et al.*, 2014; Sagatun *et al.*, 2014; Hawkins *et al.*, 2013; Hascher and Hagenauer, 2010; Hadjar and Lupatsch, 2010; Finn, 1989; Fine and Rosenberg, 1983); they report also unsatisfactory relationships with teachers and classmates, expectations of failure, poor motivation and low commitment (Renaud-Dubé *et al.*, 2015; Lessard *et al.*, 2014; Fortin *et al.*, 2013). In all the mentioned cases there is a high proportion of inadequate intrinsic motivation and low profit.

According to these works, in our study we include as status variables the economic, social and cultural status indicator, the immigrant status and the gender, and as alterable factors the student's self-efficacy, the student's demotivation and the student's future expectation, the parent's support and the relation with peers, and the cognitive skills measured using the student's performance in Italian Language and Mathematics at the INVALSI tests. Students differ in their motivations, preferences and perception of their abilities and, although it is a complex issue, it is important to identify the individual differences and understand and model the complex interplay of all the variables. In this work, we tried to outline the profiles of the students at risk of dropout answering the following research question: which are the characteristics that accurately identify students more likely to drop out?

3. Methods

3.1. Participants

This study is based on the analysis of data from INVALSI national large-scale program. INVALSI, yearly, in Italy carries out standardized tests to measure the student's achievements at different scholastic grades and in different subjects. In addition, INVALSI required each student to compile a questionnaire (Student's Questionnaire) after the standardized tests. The first part of the questionnaire is administered after the Italian Language tests and the second part after the Math tests. This survey aims to collect information about the socio-demographic characteristics of students, such as parental status and educational qualification, gender, citizenship. Moreover, annually, school' secretarial offices provide to INVALSI information about the students enrolled in the scholastic grades of INVALSI standardized tests.

In this work, we exploited data obtained by two different sources: the results at INVALSI test and Student's Questionnaire administered by INVALSI at the end of lower secondary school (8^{th} grade) in the school year

ISBN 9788835123330

2017/2018 and administrative data of the inscription of all the students of 10th grade at the INVALSI test of 2019/2020. The administrative data contains the list of all the students enrolled on a lyceum, a technical institute or a vocational institute. From 2018, INVALSI standard test is compulsory at the end of the lower secondary school for the students' admission at the state final exam. The test compulsoriness allows collecting data on the entire population. In 2017/2018, INVALSI administered the complete Student's Questionnaire to a population of 51,0821 students in 5,795 schools. The 2017/2018 Student's Questionnaire is characterized by the presence of 37 multiple choices that examine in-depth 8 domains of the students' life: the student's motivation for studying Italian Language and Math, the student's view of the school, parent's support, student's self-efficacy, student's relationships with the peers, student's future expectations and student's qualification expectations (for more information about the theoretical framework see: https://invalsi-areaprove.cineca.it/docs/file/QdR_Questionari.pdf).

In our study, we focused on the students who compiled the items belonging to 5 out of 8 domains (demotivation, self-efficacy, parent's support, relationships with peers and future expectations) and for whom the demographic and the socio-cultural information are available. Thus, the exploited sample is composed of 418,861 students.

3.2. Variables

To monitor the students after the lower secondary school, we used the administrative data of student enrolment at grade 10 to compute the binary indicator of drop-out risk matching the records at student level. The dropout risk indicator assumes value 1 if the 8th grade student is enrolled in 2019/2020 on the second year of the upper secondary school and 0 otherwise. Thus, the group of students identified by the null value of the indicator of dropout risk are those with a higher risk of dropping out. For sake of simplicity, we labelled the students with the indicator of dropout risk equal to 0 "High-risk group" and those with the drop-out risk indicator equal to 1 "Low-risk group".

To account for the socio-demographic background of the students we include in our analysis the following variables: the index of economic, social and cultural status (ESCS) computed by INVALSI following the OECD's standard (OECD, 2005); the indicator of the immigrant status (native, first-generation immigrant or second-generation immigrant); the gender of the students; a binary indicator of the late-enrolled students (i.e. students

ISBN 9788835123330

enrolled at least one year after the age of 6 or repeated one or more years). The ESCS was coded in a categorical variable with 4 different levels: low, medium-low, medium-high and high.

The Student's Questionnaire explores the student's demotivation through a set of seven items. All items begin with "Think about your experience at school…" and are scored using a five-point Likert scale ranging from 1 = "Definitely False" to 5 = "Definitely True". Three statements are related to a negative perception of the school, as a waste of time, an effort or tedium, two items represent the school as a positive place where the student has the possibility to do interesting things. The other two items are strictly related to the student's demotivation to attend the school and the intention of dropping out of school as soon as possible. The student's self-efficacy is characterized by six items in a six-point agreement Likert scale (from 1 = "Strongly Disagree" to 6 = "Strongly Agree"), investigating the level of the student belief in his/her ability. In particular, the questionnaire aims to analyse how much the student believes in his/her ability to think, wondering if he/she is able to think fast, to have a good idea and if the student has the perception to be smart. Moreover, to measure the self-efficacy, the questionnaire asks if the student considers himself/herself able to communicate his/her own opinions and if he/she puts more effort into dealing with obstacles. To study the parent's role in the school life of the student's, the questionnaire asks the student how much they are motivated from the parents to be committed to study and how much they are encouraged from the parents to be confident: four items in a six-point agreement Likert scale were administered. The relationships with the peers are investigated, using five items in a six-point Likert scale ranging from 1 = "Not at all True" to 6 = "Completely True", from two different points of view: peers as friends, persons that have your trust, and the class as a safe place where you have the perception of being accepted for your scholastic skills. Finally, the future expectations of the students were assessed by a five-item measure to investigate the degree to which the students believe a series of statements about the future. Each item is scored using a six-point Likert scale ranging from 1 = "Definitely False" to 6 = "Definitely True" and asks the students to "Think about the future…" and to focus on work or educational attainment, in general on the possibility to realize her/his own wishes.

To account for the cognitive abilities of the students, we used the students' performance at INVALSI tests in Italian Language and Mathematics at the end of the lower secondary school. The score at INVALSI test is measured using the Weighted Likelihood Estimation (WLE) estimated by the Rasch model (for more information a technical report is available on the official website: http://invalsi-areaprove.cineca.it/).

ISBN 9788835123330

3.3. Statistical analyses

An IRT (*Item Response Theory*) method called *Graded Response Model* (Samejima, 1969; Koch, 1983) for ordered polychotomous scale was applied to estimate measures of the five latent variables (demotivation, self-efficacy, parent's support, relationships with peers and future expectations) from Likert-type scale data. All the analyses are performed using a unidimensional approach.

A descriptive analysis to compare the characteristics of students at risk and not at risk of dropping out of school has been performed. For each group has been computed the students' proportion with respect to four different categorical variables: ESCS, gender, immigrant status and the late-enrolled indicator. The Chi Square test is used to study the association between the indicator of students at risk of dropping out and each categorical variable. For the five continuous factors, estimated using the *Graded Response Model*, the performance at INVALSI tests the Kruskall-Wallis test has been applied to compare the distributions of the two groups. The score in Italian Language and Mathematics have been rescaled to have a mean of 0 and a standard deviation of 1.

Focusing on the student classified in the high-risk group, we used a latent profile analysis (LPA) to examine the profile of these students in relation to seven measurements: demotivation, self-efficacy, parent's support, relationships with peers, future expectations, the score in Italian Language and the score in Math. LPA is a latent variable mixture model, widely used and known in literature with a variety of names, as finite mixture modelling (McLachlan and Peel, 2000) and latent class cluster analysis (Vermunt and Magidson, 2002). The aim of LPA is to analyse heterogeneity in a population by identifying a number of existing but unobserved subpopulations of individuals, which are referred to as latent classes. Unlike the traditional cluster analysis, the LPA is a model-based method and derives clusters using a probabilistic model able to describe the data distribution. A detailed introduction to this technique can be found in Magidson and Vermunt (2002, 2004), Muthén (2001, 2004), Muthén and Muthén (2000) and McLachlan and Peel (2000). In this work, LPA was conducted on the five measurements using *mclust* package (Scrucca *et al.*, 2016) of R software. To determine the optimal number of latent classes fitting the data, we examined the solutions containing no more than 3 clusters for different parameterizations of the covariance matrix (Fraley *et al.*, 2012) and we compared all the models based on the fit statistics using the Bayesian Information Criterion (BIC; Raftery, 1995) to evaluate the model fit. We used a multinomial regression model to evaluate the effect of the external covariates (ESCS, gender, immigrant status and the late-enrolled indicator) on the estimated class membership.

ISBN 9788835123330

4. Results

4.1. Descriptive analysis

The data reveal a significant difference in the distribution of the ESCS ($p < 2.2E-16$), the immigrant status ($p < 2.2E-16$), the gender ($p < 2.2E-16$) and the late-enrolled indicator ($p < 2.2E-16$) between the two groups of students (High-risk and Low-risk groups). Table 1 shows that 43.50% of the students classified as at risk for school dropout are characterized by low ESCS whereas only 10.50% of these have high ESCS. Focusing on the distribution of the ESCS in the group of students enrolled in grade 10 in 2019-2020, it is evident that the population is equally distributed in the four ESCS classes. The 20.00% of the students at high risk of dropping out are immigrant of first or second generation. The 57.10% of the students at high risk of dropping out are male, whereas the percentage of male in the group of students at low risk of dropping out is close to 50% (47.90%). Only 3% of the students that are attending the second year of the upper secondary school are late-enrolled, whereas this percentage increases to 20.70% for the high-risk group.

Tab. 1 – Results of Chi-Square test on status risk factors: demographic and socio-economic characteristics and late-enrolled indicator

	High-risk group		*Low-risk group*		*P-value*
	N = 66,604		*N = 352,257*		
ESCS					< 2.2E-16
Low	28,964	43.50%	76,566	21.70%	
Medium-Low	17,491	26.30%	86,914	24.70%	
Medium-High	13,165	19.80%	98,447	27.90%	
High	6,984	10.50%	90,330	25.60%	
Immigrant status					< 2.2E-16
Native	53,287	80.00%	32,7986	93.10%	
1st gen. immigrant	5,199	7.81%	7,451	2.12%	
2nd gen. immigrant	8,118	12.20%	16,820	4.77%	
Gender					< 2.2E-16
Female	28,573	42.90%	183,371	52.10%	
Male	38,031	57.10%	168,886	47.90%	
Late-enrolled					< 2.2E-16
No	52,811	79.30%	341,672	97.00%	
Yes	13,793	20.70%	10,585	3.00%	

ISBN 9788835123330

Table 2 shows the median with the 2.5th and the 97.5th percentile for the alterable risk factors included in the analysis: demotivation, future expectations, parent's support, relationships with peers and self-efficacy. There is a statically significant difference between the two groups for the distributions of demotivation, future expectations, perception of parents' support and self-efficacy. The high-risk group is characterized by high demotivation, low future expectations, low support and interest of the parents in the student school life and self-efficacy near the population mean since all these variables for construction are normal distributed with mean 0 and variance equal to 1.

Tab. 2 – Results of Kruskal-Wallis test on the alterable risk factors

	High risk group		*Low risk group*		*P-value*
	N = 66,604		*N = 352,257*		
Demotivation	0.33	[-1.72;2.32]	-0.10	[-1.72;1.67]	< 2.2E-16
Future Expectations	-0.10	[-1.63;1.80]	-0.04	[-1.63;1.80]	< 0.001
Parents	-0.17	[-1.49;1.48]	-0.03	[-1.38;1.48]	< 0.001
Peers	-0.03	[-1.74;1.85]	-0.05	[-1.74;1.85]	0.690
Self-Efficacy	0.03	[-1.53;1.65]	-0,07	[-1.53;2.06]	< 0.001

Focusing on the distribution of the alterable factors in Figure 1 it is clear the central role played by the demotivation in the characterization of the two groups. It is interesting to notice that the distribution of the self-efficacy of the two groups is quite similar, and the median of self-efficacy is slightly higher ($p < 0.001$) for the students at risk of dropping out.

Tab. 3 – Results of Kruskal-Wallis test on the score at INVALSI test of Italian Language and Mathematics

	High-Risk Group	*Low-Risk Group*	*P-value*
	N = 66,604	*N = 352,257*	
Italian Language	-0.67 [-2.58; 1.05]	0.12 [-1.71; 1.99]	< 2.2E-16
Mathematics	-0.68 [-2.10; 1.09]	0.08 [-1.64; 2.16]	< 2.2E-16

Table 3 reports the median with the 2.5th and the 97.5th percentile and the results of the Kruskal-Wallis tests to compare the distribution of the score at INVALSI tests in Italian Language and Mathematics for the high-risk and low-risk groups. A statistically significant difference between the groups is observed in both subjects. The students in the high-risk group have lower performance than those in the low-risk group. The boxplots in Figure 2 represents the distribution of score at INVALSI test for each group in cor-

ISBN 9788835123330

respondence of each subject. The 75% of the students at high-risk achieve a score lower than the median of the score of the low-risk students in both subjects.

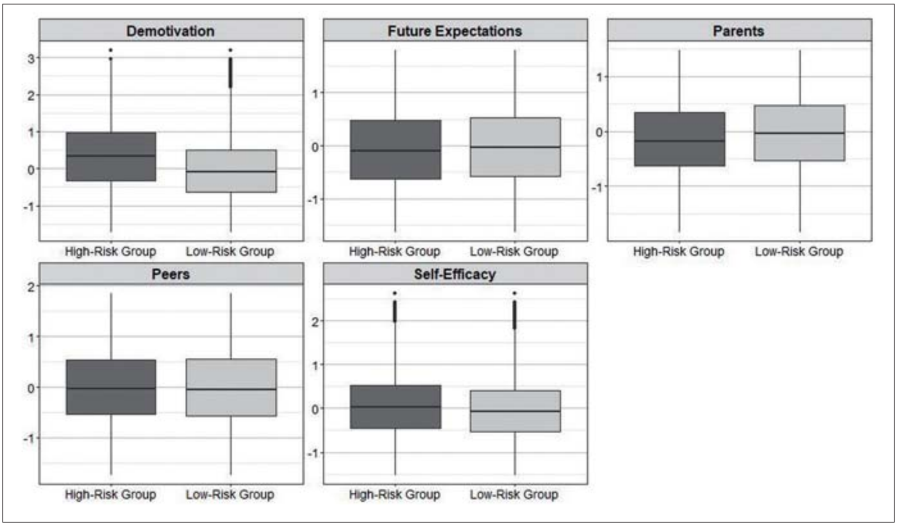


Fig. 1 – Boxplot of the distribution of the five alterable factors by two groups: students at high-risk and students at low-risk of dropping out school

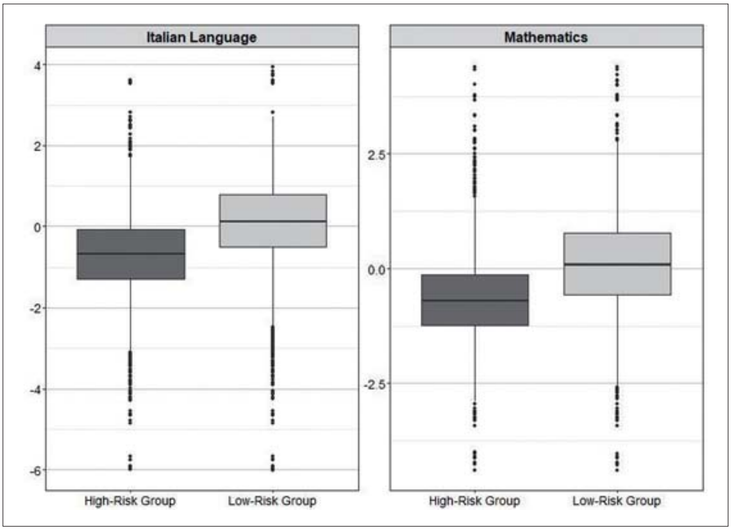


Fig. 2 – Boxplot of the distribution of the score at INVALSI test in Italian Language and Mathematics for the high risk group and the low risk group

ISBN 9788835123330

4.2. Latent profile analysis

A more detailed characterization of the high-risk group has been obtained using the LPA. Figure 3 presents the values of the Bayesian Information Criterion for 1- to 3-class solutions in correspondence of several parametrizations of the covariance matrix. A 3-class solution with a VVV (ellipsoidal, varying volume, shape and orientation) parametrization yielded the best fit for the data, resulting in the highest BIC.

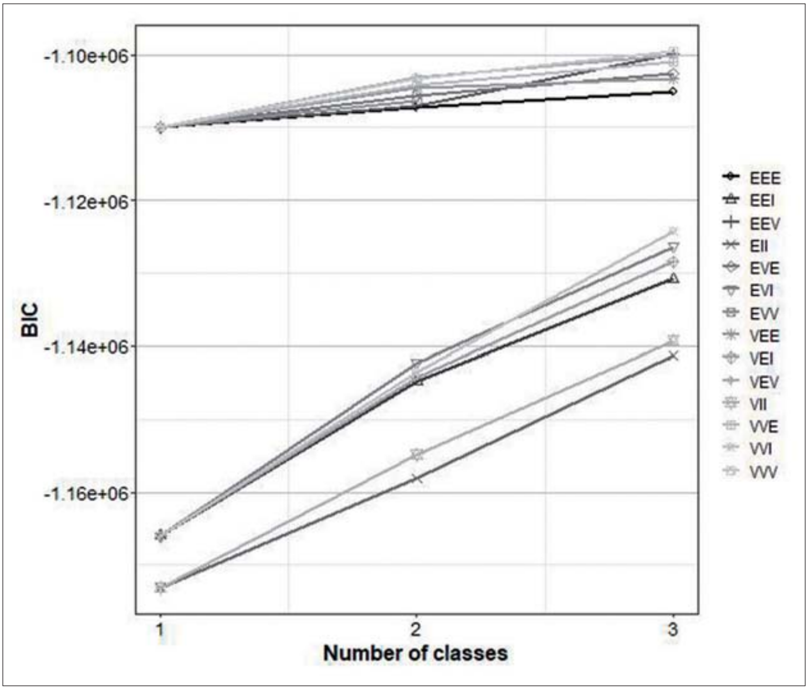


Fig. 3 – Bayesian information criterion (BIC) for various model-based clustering solutions

Figure 4a shows the class-conditional mean profiles of the three classes, whereas the estimated mean probability of being in each class by the ESCS, the gender, the immigrant status and the late-enrolled indicator is represented in Figure 4b, Figure 4c, Figure 4d and Figure 4e, respectively. Class 1 represents about 53% of the students with high-risk of dropping out of school. This cluster is characterized by a high mean on demotivation (z = 0.46), a very low mean on the expectations for the future (z = -0.84), low self-efficacy (z = -0.39) and a moderately low mean on the school performance (Italian Language: -0.51;

ISBN 9788835123330

Mathematics: -0.47). The lowest mean value of self-efficacy is observed in correspondence of the students allocated in the Class 1. From the analysis of the relation between latent class membership and the predictors, it is clear that the Class 1 membership in a growing function of the ESCS and of the gender. The first generation immigrants have on average a lower probability to be allocated in the Class 1 than the second generation immigrants and the native. The mean probability of Class 1 is a decreasing function of the late enrolled indicator.

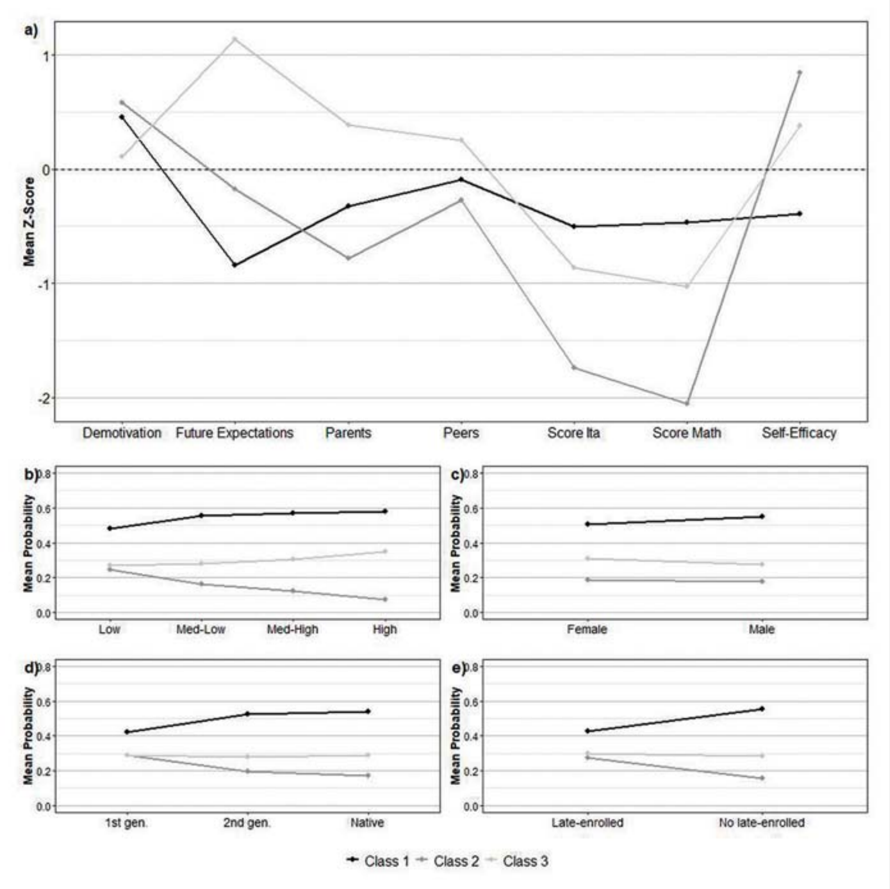


Fig. 4 – Results of latent profile analysis for 3 classes: a) estimated means of the seven observed variables with each classes; b) predicted probability of latent class membership at varying levels of the ESCS; c) predicted probability of latent class membership at varying levels of the gender indicator; d) predicted probability of latent class membership at varying levels of the immigrant indicator; e) predicted probability of latent class membership at varying levels of the late-enrolled indicator

ISBN 9788835123330

Class 2 is composed by about 18% of the students. This group is characterized mainly by four variables: demotivation, self-efficacy and performances at INVALSI tests. The students classified in this cluster show on average a strong demotivation ($z = 0.59$) and a strong self-efficacy ($z = 0.84$). This group have the lowest mean value in correspondence of both measures of the cognitive skills (Italian Language: $z = -1.74$; Mathematics: $z = -2.06$). The probability to belong to this class is a decreasing function of the ESCS and shows higher values in correspondence of late-enrolled students and immigrant of first-generation, respectively.

The last cluster, Class 3, represents about the 29% of the students with high-risk of dropping out of school. These students show a particular profile since they have a very high mean on the future expectations ($z = 1.14$), a moderately high mean on parent's supports ($z = 0.39$) and on self-efficacy ($z = 0.38$) and a low mean on Italian Language performance ($z = -0.86$) and Maths performance ($z = -1.03$). The Class 3 membership is an increasing function of the ESCS; the probability to be classified in this cluster seems to be constant with respect to the varying of the categories of the gender, the immigrant status and the late-enrolled indicator.

5. Discussion

In this exploratory study we explored deeply the Goal 4 of the 2030 Agenda by investigating the characteristics that accurately identify students more likely to drop out in the transition between the last year of the lower secondary school and the upper secondary school. We compared the characteristics of students at risk and not at risk of dropping out school and we described the risk group using Latent Profile Analysis.

The descriptive analysis of the characteristics of the two groups (high-risk group and low-risk group) confirms the impact of dropout risk factors related to socio-demographic characteristics as gender, immigrant status and the socio-economic background. There is also a significant difference in the distribution of the late-enrolled indicator.

By observing the distribution of the alterable factors it is clear the crucial role of the demotivation in the characterization of the two groups. This result agrees with an increasing research activity, which investigates risk factors related to characteristics at school level and organization at school level and with theories on the school engagement. In all cases where the school makes a strong commitment to meet specific needs of students in the family, community, classmates and school areas, the dropout rate decreases (Wang

ISBN 9788835123330

and Eccles, 2012a). The construct of engagement is useful for capturing the gradual process by which students disconnect from school (Finn, 1989). In this work, the concept of demotivation is closely linked to how the student perceives being in school. The student at high-risk of dropping out of school perceives school as a waste of time, an effort or a boredom, a place where they do not do interesting things and they would like to stop going to school as soon as possible.

On the contrary, it is interesting to notice that the distribution of the self-efficacy of the two groups is quite similar, and the median of self-efficacy is slightly higher for the students in the high-risk group. If decades of research suggest that school effort decreases rapidly during secondary school (Wang and Eccles 2012b, 2012c), general self-efficacy remains moderately stable from 9th through the 12th grade (Grabowski and Mortimer, 2001). General self-efficacy is conceptualized as "broad and stable" (Luszczynska *et al.*, 2005, p. 81) assessments of individuals' overall beliefs about their abilities "to perform across a variety of situations" (Judge *et al.*, 1998, p. 170). In this work, we refer to a general concept of self-efficacy because we investigate the level of the student believes in his/her ability to think, to communicate his/her own opinion, the ability to learn new things and to have good ideas. However, it seems unlikely that general self-efficacy is totally unrelated to academic experience (Brown *et al.*, 2019).

To shed light on the role of self-efficacy in the group of students at high-risk and, in general, to test the hypothesis that this group is composed by different latent subpopulations characterized by different profile in terms of self-efficacy, demotivation, future expectations, relationship with parents and with peers and school performance, we used a Latent Profile Analysis. Latent models provide a powerful tool to classify individual based on a set of manifest characteristics, identifying latent subgroups of students, and results useful when the "average" learning pattern does not adequately describe the phenomenon (Hichendorff *et al.*, 2018). The comparison of the models with a different number of latent classes corroborate our hypothesis that the population of students at high-risk of dropping out is not homogenous and is composed by three different subpopulations representing different students' profiles. The relationship between the class membership and the status variables has been studied using a multinomial regression model and helps to characterize and better understand the identified student's profile.

The lowest average self-efficacy value is observed in students assigned to Class 1, where it is classified the largest group of students at risk of dropping out of school (about 53%). On the other hand, in the other two groups of students at risk, the self-efficacy value is above the average value. The Class

ISBN 9788835123330

1 is also characterized by a high mean on demotivation, a very low mean on the expectations for the future, a moderately low mean on the school performance and the belonging to this class is an increasing function of the ESCS and gender. This group, which could be called "socially disadvantaged students", has been widely studied in the literature; they usually come from low-income, low-educated families or single parents and usually have a lack of expectations of staying at school, unsatisfactory relationships with teachers and classmates, expectations of failure, poor motivation and low commitment. However, not all students with socio-economic disadvantages are equally vulnerable to circumstances. Much of the research examines students' ability to thrive despite negative effects circumstances investigating the key role played by the character's strengths, such as trust their academic skills, assertiveness, ability to work hard, high internal standards motivation to achieve and ambitious aspirations for their future (Martin and Marsh, 2009). For this group at risk, a possible starting point to reduce the risk of dropout could be the self-efficacy. Positive general self-efficacy is believed to contribute to individuals' overall levels of optimism and adjustment and their capacity to deal with stressful situations and to produce successful results (Luszczynska *et al.*, 2005).

A strong role of the school, as previously said, could bring many benefits especially to students classified in the Class 2 that we name "students with low confidence in the school". These students despite having high self-efficacy, perceive going to school as futile and consequently they have low performances at INVALSI tests. In fact, students who belong to this class are more likely to be late-enrolled students or immigrant of first-generation. Considering the characteristics of this group, the low performance of the INVALSI tests could be due to a poor commitment in carrying out the test. To clarify this aspect, it may be useful to measure the student's effort during the test using the item response time.

The last cluster identified, Class 3, which we call "low performer", is characterized by a low performance at INVALSI tests and by scores above average in all alterable variables. The probability to be classified in this cluster seems to be constant with respect the varying of the categories of the gender, the immigrant status and the late-enrolled indicator. Probably more than the other groups, this group is the one that benefits the most from standard dropout prevention programs, based only on performance. One aspect that could clarify the characteristics of this group and that we will analyze in the future concerns the students attending the IeFP. In this work, in the group of students at risk of dropout there are also those enrolled in these courses that could fit in this profile. It is possible that these students are addressed to these

ISBN 9788835123330

courses because they have low performances and have high expectations for the future because they hope to enter the work world soon. Future studies will include the information about IeFP.

In summary, the present study describes the characteristics of students at high and low risk of dropout in Italy using the INVALSI data. In particular, this work extends the existing literature analyzing the students at high-risk of dropping out, using a latent model approach, demonstrating the presence of different high-risk students' profiles in terms of self-efficacy, demotivation, future expectations, relationship with parents and with peers and school performance. The findings of this work could be useful for targeting efficiently students destined to dropout prevention programs.

References

Ballarino G., Checchi D., Fiorio C., Leonardi M. (2010), "Le disuguaglianze nell'accesso all'istruzione in Italia", *Quaderni Rassegna Sindacale-Lavori*, 11, 1, pp. 177-231.

Ballarino G., Schadee H. (2006), "Espansione dell'istruzione e disuguaglianza delle opportunità formative nell'Italia contemporanea", *Polis*, 20, 2, pp. 207-228.

Borgna C., Struffolino E. (2017), "Pushed or pulled? Girls and boys facing early school leaving risk in Italy", *Social Science Research*, 61, 298-313, retrieved on January, 21, 2020, from https://doi.org/10.1016/j.ssresearch.2016.06.02.

Brown T.M., Galindo C., Quarles B., Cook A.L.J. (2019), "Self-Efficacy, Dropout Status, and the Role of In-School Experiences Among Urban, Young Adult School-Leavers and Non-leavers", *Urban Review*, 51, pp. 816-844, retrieved on January, 21, 2020, from https://doi.org/10.1007/s11256-019-00508-3.

Checchi D., Flabbi L. (2006), "Mobilità intergenerazionale e decisioni scolastiche in Italia", in G. Ballarino, D. Checchi (a cura di), *Sistema scolastico e disuguaglianza sociale. Scelte individuali e vincoli strutturali*, il Mulino, Bologna.

Esch P., Bocquet V., Pull C., Couffignal S., Lehnert T., Graas M., Laurence F.H., Ansseau M. (2014), "The downward spiral of mental disorders and educational attainment: A systematic review on early school leaving", *BMC Psychiatry*, 14, 237, retrieved on January, 21, 2020, from https://doi.org/10.1186/s12888-014-0237-4.

Fine M., Rosenberg P. (1983), "Dropping out of high school: The ideology of school and work", *Journal of Education*, 165, pp. 257-272, retrieved on January, 21, 2020, from https://doi.org/10.1177/002205748316500304.

Finn J. D. (1989), "Withdrawing from school", *Review of Educational Research*, 59, 117-142, retrieved on January, 21, 2020, from http://dx.doi.org/10.3102/00346543059002117.

Fortin L., Marcotte D., Diallo T., Potvin P., Royer E. (2013), "A multidimensional model of school dropout from an 8-year longitudinal study in a general high

ISBN 9788835123330

school population", *European Journal of Psychology of Education*, 28, pp. 563-583, retrieved on January, 21, 2020, from https://doi.org/10.1007/s10212-012-0129-2.

Fraley C., Raftery A.E., Murphy T.B., Scrucca L. (2012), *mclust version 4 for R: normal mixture modeling for model-based clustering, classification, and density estimation*, University of Washington – Departement of Statistics, Technical report, 597.

Grabowski L.J.S., Call K.T., Mortimer J.T. (2001), "Global and economic self-efficacy in the educational attainment process", *Social Psychology Quarterly*, 64, 2, pp. 164-179.

Hadjar A., Lupatsch J. (2010), "The Lower Educational Success of Boys", *Köln Z Soziol*, 62, 4, pp. 599-622.

Hascher T., Hagenauer G. (2010), "Alienation from School", *International Journal of Educational Research*, 49, 6, pp. 220-232, retrieved on January, 21, 2020, from: https://doi.org/49. 220-232. 10.1016/j.ijer.2011.03.002.

Hawkins R.L., Jaccard J., Needle E. (2013), "Nonacademic factors associated with dropping out of high school: Adolescent problem behaviors", *Journal of the Society for Social Work and Research*, 4, 2, pp. 58-75.

Hickendorff M., Edelsbrunner P.A., McMullen J., Schneider M., Trezise K. (2018), "Informative tools for characterizing individual differences in learning: Latent class, latent profile, and latent transition analysis", *Learning and Individual Differences*, 66, pp. 4-15.

ISTAT (2019), *Rapporto SDGs 2019: informazioni statistiche per l'Agenda 2030 in Italia*, Roma.

Judge T.A., Erez A., Bono J.E. (1998), "The power of being positive: The relation between positive self-concept and job performance", *Human Performance*, 11, 2-3, pp. 167-187.

Koch W. R. (1983), "Likert scaling using the graded response latent trait model", *Applied Psychological Measurement*, 7, 1, pp. 15-32, retrieved on January, 21, 2020, from https://doi.org/10.1177/014662168300700104.

Lamb S., Markussen E. (2011), "School dropout and completion: an international perspective", in S. Lamb, E. Markussen, R. Teese, J. Polesel, N. Sandberg (*ed.*), *School dropout and completion*, Springer, Dordrecht.

Lessard A., Butler-Kisber L., Fortin L., Marcotte D. (2014), "Analyzing the discourse of dropouts and resilient students", *The Journal of Educational Research*, 107, 2, pp. 103-110, retrieved on January, 21, 2020, from https://doi.org/10.1080/00220671.2012.753857.

Luszczynska A., Gutiérrez-Doña B., Schwarzer R. (2005), "General self-efficacy in various domains of human functioning: Evidence from five countries", *International journal of Psychology*, 40, 2, pp. 80-89.

McLachlan G., Peel D. (2000), *Finite mixture models*, John Wiley & Sons, New York.

Malmberg-Heimone I., Sletten M.A., Tøge A.G., Gyüre K., Borg. E. (2018), "Research protocol: Systematic follow-up in order to reduce dropout in upper sec-

ISBN 9788835123330

ondary schools. A cluster-randomised evaluation of the IKO model", *International Journal of Educational Research*, 89, pp. 153-160, retrieved on January, 21, 2020, from https://doi.org/10.1016/j.ijer.2017.11.001.

Magidson J., Vermunt J. K. (2002), "Latent class models for clustering: a comparison with K-means", *Canadian Journal of Marketing Research*, 20, 1, pp. 36-43.

Magidson J., Vermunt J.K. (2004), "Latent class models", in D. Kaplan (*ed.*), *Handbook of quantitative methodology for the social sciences*, Sage Publications, Newbury Park (CA).

Martin A.J., Marsh H.W. (2009), "Academic resilience and academic buoyancy: Multidimensional and hierarchical conceptual framing of causes, correlates and cognate constructs", *Oxford Review of Education*, 35, 3, pp. 353-370.

MIUR (2019), *La dispersione scolastica nell'anno scolastico 2016/2017 e nel passaggio all'anno scolastico 2017/2018*, retrieved on January, 21, 2020, from https://www.miur.gov.it/documents/20182/2155736/La+dispersione+scolastica+nell%27a.s.2016-17+e+nel+passaggio+all%27a.s.2017-18.pdf/1e374ddd-29ac-11e2-dede-4710d6613062?version=1.0&t=1563371652741.

Mocetti S. (2012), "Educational choices and the selection process: before and after compulsory schooling", *Education Economics*, 20, 2: 189-209, retrieved on January, 21, 2020, from https://doi.org/10.1080/09645291003726434.

Muthén B.O. (2001), "Latent variable mixture modelling", in G.A. Marcoulides, R.E. Schumacker (*ed.*), *New developments and techniques in structural equation modelling*, Lawrence Erlbaum Associates, Hillsdale (NJ).

Muthén B.O. (2004), "Latent variable analysis: Growth mixture modeling and related techniques for longitudinal data", in D. Kaplan (*ed.*), *Handbook of quantitative methodology for the social sciences*, Sage, Newbury Park (CA).

Muthén B.O., Muthén L.K. (2000), "Integrating person-centered and variable-centered analyses: growth mixture modelling with latent trajectory classes", *Alcoholism: Clinical and Experimental Research*, 24, pp. 882-891.

Nouwen W., Clycq N. (2020), "Assessing the added value of the self-system model of motivational development in explaining school engagement among students at risk of early leaving from education and training", *European Journal of Psychology of Education*, 1-19, retrieved on January, 21, 2020, from https://doi.org/10.1007/s10212-020-00476-3.

Nylund K., Asparouhov T., Muthén B.O. (2007), "Deciding on the number of classes in latent class analysis and growth mixture modeling: a Monte Carlo simulation study", *Structural Equation Modeling*, 14, pp. 535-569.

OECD (2005), *PISA 2003 Technical Report*, OECD, Paris.

Raftery A. (1995), "Bayesian model selection in social research", *Sociological Methodology*, 25, pp. 111-163.

Renaud-Dubé A., Guay F., Talbot D., Taylor G., Koestner R. (2015), "The relations between implicit intelligence beliefs, autonomous academic motivation, and school persistence intentions: a mediation model", *Soc. Psychol. Educ.*, 18, pp. 255-272, retrieved on January, 21, 2020, from https://doi.org/10.1007/s11218-014-9288-0.

ISBN 9788835123330

Rumberger R.W., Lim S.A. (2008), “Why students drop out of school: A review of 25 years of research”, *California Dropout Research Project*, 15, pp. 1-3.

Sachs J., Schmidt-Traub G., Kroll C., Lafortune G., Fuller G. (2019), *Sustainable Development Report 2019*, Bertelsmann Stiftung and Sustainable Development Solutions Network (SDSN), New York.

Sagatun Å., Heyerdahl S., Wentzel-Larsen T., Lien L. (2014), “Mental health problems in the 10th grade and non-completion of upper-secondary school: The mediating role of grades in a population-based longitudinal study”, *BMC Public Health*, 14, 1, p. 16.

Samejima F. (1969), “Estimation of latent ability using a response pattern of graded scores”, *Psychometrika Monograph*, 17, pp. 1-97.

Scrucca L., Fop M., Murphy T.B., Raftery A.E. (2016), “mclust 5: clustering, classification and density estimation using Gaussian finite mixture models”, *The R Journal*, 8, 1, pp. 289-317, retrieved on January, 21, 2020, from https://doi.org/10.32614/RJ-2016-021.

Wang M.T., Eccles J.S. (2012a), “Social support matters: longitudinal effects of social support on three dimensions of school engagement from middle to high school”, *Child Development*, 83, pp. 877-895, retrieved on January, 21, 2020, from https://doi.org/10.1111/j.1467-8624.2012.01745.x.

Wang M.T., Eccles J.S. (2012b), “Adolescent behavioural, emotional, and cognitive engagement trajectories in school and their differential relations to educational success”, *Journal of Research on Adolescence*, 22, pp. 31-39, retrieved on January, 21, 2020, from https://doi.org /10.1111/j.1532-7795.2011.00753.x.

Wang M.T., Eccles J.S. (2012c), “Social support matters: longitudinal effects of social support on three dimensions of school engagement from middle to high school”, *Child Development*, 83, pp. 877-895, retrieved on January, 21, 2020, from https://doi.org /10.1111/j.1467-8624.2012.01745.x.

ISBN 9788835123330

The authors

Cecilia Bagnarol, graduated in Statistics, Economics and Business at the Alma Mater Studiorum of Bologna. She currently works at Statistical Service of INVALSI where she performs support activities for statistical analysis on large data bases of national surveys on learning.

Andrea Bendinelli got master degree in Statistics and works at INVALSI's statistical service. He carries out statistical analysis activities on large databases and conducts research support activities in the assessment of students learning.

Emiliano Campodifiori, graduated in Statistics and Economics at the University of Rome "La Sapienza". Currently he works in the Statistical Service of INVALSI, he performs statistical analysis of the National Assessment data.

Michele Cardone holds a Degree in Statistics for demographic and social sciences and a Master (I level) in Statistics for the management of information systems (Università di Roma "La Sapienza"). Working for INVALSI since 2004, member of the Statistical Service since 2010, mainly involved in the analysis of the school tests data and in the management of the annual data return to schools.

Silvia Donno, graduated in Demographic Sciences for Social and Health Policies at the University of Rome "La Sapienza". Currently she works in Statistical Service of INVALSI, she carries out activities to support the elaboration and statistical analysis of the data of the national surveys on learning.

ISBN 9788835123330

Patrizia Falzetti is Head of the INVALSI Statistical Service, which manages the acquisition, analysis and return of data concerning national and international surveys on learning to individual schools, stakeholders and the scientific community.

Paola Giangiacomo is a researcher at the National Institute for Educational and Educational Education Assessment (INVALSI), where she holds the position of National Data Manager for the surveys promoted by the OECD. Her main activities concern the revision and adaptation of survey instruments, the definition of sampling plans, the statistical analysis of quantitative and qualitative data, the drafting of technical and scientific reports, training activities for data analysis.

Patrizia Giannantoni, PhD in Statistics/Demography. She has worked on psychometric evaluation of developmental tests with CNR and on research projects on migration as research fellow at University of Naples. Since 2017 she has joined the statistical office of INVALSI, keeping her research interests on migrant integration support and educational inequalities.

Jana Kopecna, PhD in Demography. She has been involved in research projects on international migration and educational integration of migrants, at the University Sapienza of Rome. She is currently employed at the Statistical Office of INVALSI.

Francesca Leggi, graduated in Sociology, specializing in Economics, Labour and Organizations at the University of Rome “La Sapienza”. Currently, she works at the Statistics Office of INVALSI, focusing on the statistical analysis on large databases.

Giuseppina Le Rose, psychologist, psychotherapist and expert in psychological evaluation and counseling, currently works at INVALSI. She has performed numerous educational and vocational interventions and collaborated in the preparation of psycho-aptitudinal, cognitive and personality tests.

Michele Marsili, graduated in Statistics at Sapienza University of Rome. He worked in Business Intelligence consulting, providing software development solutions for analysis and support for company’s decision making in insurance and pharmaceutical industries. Since January 2018 he has been working in the Statistical Service of INVALSI.

ISBN 9788835123330

Veronica Pastori, PhD in Methodology of Social Sciences. Her main interests of research are social inequalities, migratory phenomena, evaluation of educational processes, data quality and construction of standardized questionnaire. Currently, she works at the Statistics Office of INVALSI.

Veronica Riccardi, PhD in Pedagogy, currently works at INVALSI. Her research interests mainly concern the field of studies of intercultural pedagogy, adult education and social pedagogy.

Maria Carmela Russo, PhD in Methodology of Social Sciences, works at the Statistics Office of INVALSI. Her research interests concern territorial inequalities and migration background in student's achievement. She worked on evaluation of tertiary education, indicators to measure violence against women in comparative perspective, and marriage dissolution.

Chiara Sacco, PhD in Statistical Methodology for Scientific Research, currently works at the INVALSI Research Institute as statistician. The main research interests are in the context of multivariate data analysis for high dimensional data with particular focus on dimension reduction strategies, model based clustering and latent variable models.

Valeria F. Tortora is a researcher at the National Institute for Educational and Educational Education Assessment (INVALSI), where she is National Manager for the International Association for the Evaluation of Educational Achievement (IEA). She is PhD in Comparative Education with a thesis on the use of OECD-PISA results by teachers to improve their teaching strategies. The most recent research concerns the study of social inequalities, the variables connected to educational performance of students.

ISBN 9788835123330

Agenda 2030 for Sustainable Development is a shared action plan for people, for the planet and for prosperity, adopted in September 2015 by the 193 United Nations (UN) Member States. The guidelines of this journey are summarized by 17 Sustainable Development Goals (SDGs) and the associated 169 Targets approved by the UN, with the shared aim to reach them by 2030. The volume, composed of 7 chapters, discusses 3 of the SDGs: to ensure inclusive and equitable quality education, to achieve gender equality and to reduce economic inequality within and across national borders. The INVALSI database provided a valuable resource to the authors to investigate the characteristics of the Italian school system. We wish the volume reading encourages the discussion about possible ameliorative interventions and it is a starting point to measure potential progress.

Patrizia Falzetti is Head of the INVALSI Statistical Service, which manages the acquisition, analysis and return of data concerning national and international surveys on learning to individual schools, stakeholders and the scientific community.