Effectiveness of Tuscan primary schools: what role for school factors?

Enrico Conti (Irpet)

Silvia Duranti (Irpet)

Maria Luisa Maitino (Irpet)

Carla Rampichini (Università degli Studi di Firenze)

Nicola Sciclone (Irpet)

1. Introduction

The evaluation of schools' performance has not a long tradition in Italy, where only in 2007 a national committee (*Istituto nazionale per la valutazione del sistema educativo di istruzione e di formazione* – hereafter, Invalsi) has been established with the specific purpose of assessing the competences acquired by pupils, and thus evaluating the role of schools. As a consequence, a national standardized test has been developed to assess students in reading and mathematics at different stages of their educational career; the new evaluation system, gradually introduced in a subset of grades, is now administered on a census basis to the 2nd and 5th classes of primary school (Grades 2 and 5), the 1st and 3rd classes of middle school (Grades 6 and 8) and the 2nd class of high school (Grades 10).

The availability of these data enables researchers to fill the gap between Italy and other developed countries (especially US), where the analysis of students' and schools' results has a long tradition. Indeed, information about students' achievements have a high political relevance, which in Italy is emphasised by the poor performance of students with respect to their counterparts around the world, as stressed by international standardized tests' scores, like OECD-PISA[1] (OECD, 2010). Therefore, Invalsi data may represent for policy makers an important tool to establish if the Italian school system is achieving its objectives, in terms of both efficiency and equity, providing a starting point for designing educational reforms and rethinking the allocation of public budgets. This would be particularly important in a moment when public opinion calls for more accountability of schools' performance, not only to understand the efficiency of public spending, but also to have more information on the basis of which to choose the best school for their children.

From the point of view of economists, data on standardized achievement tests' scores make it possible to assess schools' performance in a variety of ways.

One possible approach is that of efficiency: given the amount of resources used by each school (inputs) and its pupils' achievement level (output), it is possible to evaluate the level of efficiency in transforming inputs into outputs and possibly to find its main determinants. This approach is particularly apt to understand how different schools employ public resources to reach a certain level of average pupils' competences; therefore, it allows policy makers to identify the cases in which a rationalization of resources is possible, also providing information on how to achieve it. In the literature about school performance several studies analyse the theme of efficiency (see MEF, 2012 for an international review), but only a few consider the Italian case (Barbetta and Turati, 2003; Quintano et al., 2010; Di Giacomo and Pennisi, 2011; Agasisti, 2011; Agasisti et al. 2012; Agasisti and Sibiano, 2013; MEF, 2012).

A second approach is that of effectiveness, which aims to identify the degree of achievement of schools' targets, that is the attainment of an adequate level of knowledge. In the case of education, effectiveness is measured in relative terms, that is comparing institutions offering the same service after having adjusted for factors out of their control (Grilli and Rampichini, 2009). The kind of adjustment required for assessing effectiveness depends on the type of effectiveness one wants to estimate; if the aim is assessing the "production process" in order to evaluate the ability of the institution to exploit the available resources, then the school's performance needs to be adjusted for the features of its students, of the school itself and of the context in which it operates. The analysis of school effectiveness can be done using both a static and a longitudinal perspective. In the first case, the aim is disentangling the role of individual, school and territorial characteristics in determining student's performance, with the main goal of identifying the relevance of "school factors" (see Section 2 for a review). Other studies try to analyse school effectiveness in a longitudinal perspective, with a panel

[1] Other international tests on students' achievements are the Program for international Student Assessment (PISA), the Trends in Maths and Science Study (TIMSS) and the Progress in International Reading Literacy Study (PIRLS). See Montanaro (2008) for a description.

or pseudo-panel approach, which allows to identify the "value added" provided by the school net of students' starting level of competences; value-added models, which tracks individual test scores over time, are very popular in England (FitzGibbon, 1997) and in the US (Sanders, Saxton and Horn, 1997), while in Italy only a few studies adopt such a perspective (Cipollone et al. 2010; De Simone, 2011; De Simone and Gavosto, 2013), because of the lack of a longitudinal design in the available data.

In this paper the focus is on the effectiveness of the Tuscan primary education system, restricting the attention to public schools. Our aim is to describe the main determinants of Grade 5 pupils' achievement, disentangling the role of individual factors (demographic, social, economic and cultural) from that of school factors, the latter divided into two categories: externally-determined factors (average characteristics of pupils who are attending the school, e.g. socio-economic status, nationality) and school-manageable factors (quality and quantity of school resources) (Agasisti and Vittadini, 2012). To this end, we adopt a multilevel regression model, which properly takes into account the hierarchical structure of data (pupils nested into schools) by partitioning the residual variance into pupils and school components. Using this kind of methodology it is possible to obtain results that can be used for several purposes (Grilli and Rampichini, 2009). The first, it is the study of the relationship between the outcome and the explanatory variables, which is a common aim of all statistical models; findings should be interpreted in terms of associations without giving them any causal interpretation, as the availability of a single cross-section of data prevents any attempt of identifying causality, which requires more sophisticated techniques and data (Murnane and Willet 2011). A second purpose is predicting the outcome for a given student in a given school, in order to understand how a different context can change a student's potential performance. A third purpose concerns ranking schools according to their effectiveness, which can be derived from school-level residuals. This represents a useful tool to identify areas with anomalous performances.

Our paper is innovative in this field of research in two aspects. First, the paper benefits from the construction of a novel dataset, which gathers data provided by Invalsi (containing test scores, individual characteristics and some information on schools) and the data available on the website of the Italian Ministry of Education (MIUR) (containing information on school resources). The availability of school, class and pupils variables allows an original analysis in respect to the existing literature on the effectiveness of Italian schools. Indeed, we have direct measures of school resources, while existing works in this field of research are mainly based on indirect measures of resources, created on the basis of information provided by school principals (Castellano et al., 2009; Benadusi et al., 2010; Agasisti and Vittadini, 2012; Agasisti, 2013); moreover, the availability of class level characteristics allows us to estimate peer effects much more precisely than through school composition variables, frequently used in the existing literature. A second innovative aspects concerns the focus of the paper on primary schools located into a single region, namely Tuscany; this avoids the high variance across regions highlighted in several Italian studies on the issue (Checchi, 2004; Bratti et al. 2007; Montanaro, 2008; Castellano et al. 2009; Benadusi et al. 2010; Agasisti and Vittadini, 2012) and allows to focus on the variability among smaller geographical units, such as Zonal Conferences (*Conferenze Zonali*[2]). To date, little research has focused on territorial disparities other than the north-south divide; this is the case of the research conducted by Bratti et al. (2007), Benadusi et al. (2010), Gerard Ferrer-Esteban (2011) and Agasisti (2011) where, however, the benchmark for each school is at most the Province.

2. A literature review

This section briefly reviews the main determinants of students' performance, as highlighted in the literature in this field, distinguishing between the effects of four distinct subjects: family, classmates, school and community.

The family influences student's performance through a series of channels, such as the quantity and quality of time parents devote to the education of children, which can be proxied by the number of children, working habits, age etc. Moreover, also the economic status exerts a direct influence on students' performance, as wealthier parents can invest more on education of children (through extra-activities, private lessons etc.). Therefore, empirical analyses on the determinants of students' performance usually include among explanatory variables the family background, described

[2] Zonal Conferences for Education are territorial units covering several municipalities, which are aimed at the coordination and planning of the Tuscan education system at the local level. Established by LR. 5/2005, they have as a primary goal the planning of primary and lower secondary education at a local level, by coordinating the action of the municipalities belonging to them. For a description of characteristics and functions of Tuscan Zonal Conferences, see Irpet (2012).

through characteristics of the family structure (number of children, marital status of parents)[3] but above all through proxies of the family's social, cultural and economic capital. In particular, the main variables used concern direct measures of family background, such as income, education level and occupation of parents, or proxies, such as the number of books or bathrooms at home. Research in this field agrees on the important role of family background in influencing students' performance, even if the magnitude of the effect differences across countries (Wößmann 2004). Also the nationality of the students and its family can have an impact on school results, exerting an effect independent from the socio-economic status. Indeed, children of immigrants may have a gap in educational attainment due not only to lower family endowments but also to problems in integrating with classmates or to language difficulties, which tend to hamper their school performance (Schnepf, 2007).

School plays an important role in determining students' performance. However, literature on the field has focused on the amount of school resources (class size, student/teacher ratio etc.) without finding any robust evidence of its effects on students' performance (Hanushek, 1997). Although from a theoretical point of view one could expect a negative correlation between class size and school performance (smaller classes help to improve school climate and increase students' attention), empirical research has generally found a weak relationship between the two (Ehrenberg et al., 2001; Piketty and Valdenaire, 2006; Minzyuk and Russo, 2012; West and Woβmann, 2002), which appears to be a bit stronger in the early years of schooling (Finn, 1998). One reason of the lack of clear evidence could be the endogeneity of class size, often influenced by a non casual sorting of students into different classes; indeed, according to compensatory policies, weaker students tend to be allocated in smaller classes, in order to ensure them greater support from teachers and a better climate (Minzyuk and Russo, 2012; Bouzer and Rouse, 2001). As far as human resources are concerned, literature reveals that is more their quality than quantity to make the difference in students' performance (Hanushek, 1997; Woβmann, 2003). However, the analysis of the role of teachers' quality on students' performance may be hampered by a non casual allocation of human resources to schools: teachers know which are the best schools and gradually attempt to move there when their level of seniority allows them (Barbieri et al., 2007). Also organizational factors may have a role in influencing students' performance; however no clear evidence is found with regard neither to school autonomy (Woβmann, 2003; Jürges and Schneider, 2004), nor to proprietary structure (Fuchs and Woβmann, 2003; Vandenberghe & Robin, 2004).

Classmates have a direct influence on student's performance through multiple channels, which in literature are usually summarised as "peer effect". On one side, peers provide the model to be followed and influence considerably the scale of values while, on the other, they affect student's behaviour through competition effects. In both cases, a higher quality of the class should determine a better student's performance, even if literature does not provide a clear evidence in this sense (Zimmer and Toma, 2000; Hanushek et al. 2003; Hanushek, 2003. One reason for this could be the use of school level variables to identify the peer group effect (motivated by the lack of class level data in most databases on students' performance), even if the entire population of a school cannot be considered a good proxy of the peer group (Bratti et al. 2007).

Finally, also the context in which the student lives contributes to its school performance, as social cohesion, average cultural level and shared values are factors that can affect the student 's aspirations and motivation for studying. This effect, usually called "neighbourhood effect" (Bratti et al. 2007), is usually proxied by a series of statistics concerning the average income, educational level, unemployment rate etc of the local community.

When looking at the literature about Italy, it should be noted that only recently there has been a growing attention towards the determinants of students' achievement, thanks to the availability of international (such as OECD-PISA) and Invalsi test scores. The Italian line of research on school effectiveness, to which our paper is meant to contribute, is mainly based on multilevel modelling and on higher secondary school, for which both Invalsi and PISA data are available.

Using OECD-PISA data, Checchi (2004) highlights the existence of regional disparities in 15-years-old students' performance in Italy, even after controlling for the type of secondary school attended and for individual background. The analysis shows that the main factors affecting students' achievements are related to the socio-economic status; however, average parental education and socio-economic status measured at school level appear to be much stronger predictors than individual variables, thus indirectly confirming that environmental and peer factors may be important determinants of student performance.

---

[3] See Bratty et al. (2007) for a review of the literature on the role exerted on pupils' performance by family's features other than cultural-socio-economic factors.

OECD-PISA data are also employed by Bratti et al. (2007) to explain the determinants of 15-years-old students' achievement with several individual and school characteristics. Their results confirm the relevance of the socio-economic status, of the macro-area and of the type of secondary school in determining students' achievements; another significant result is that private schools perform worse than public ones.

Other contributions in this field employ multilevel modeling to take explicitly into account the hierarchical nature of data, thus providing a more robust analysis than previous ones on the determinants of students' performance. It is the case of Castellano et al. (2009), who confirm the role of socio-economic factors in explaining interindividual differences in test scores, finding little evidence of the effect exerted by school resources. A similar analysis carried out by Benadusi et al. (2010) points to the higher relevance of the average school socio-economic level compared to the individual one in explaining students' achievements, providing some evidence of the existence of forms of socio-cultural segregation among schools.

Another contribution to the literature is provided by Agasisti (2011), which also relays upon OECD-PISA data. The results are pretty similar to those obtained in the previously cited works, but the analysis includes a measure of competition among the set of covariates, to investigate whether competition actually fosters schools' performance (hypothesis partially confirmed by the empirical results). The theme of school competition is explored also using Invalsi data on lower secondary school, obtaining similar results on the effects on pupils' performance (Agasisti, 2013).

Invalsi data are also used by Agasisti and Vittadini (2012), who carry out a multilevel analysis to decompose the overall variance of students' achievement scores into three components: within-schools variance, between-schools variance and between-Regions variance. The findings confirm that variance at the regional level is statistically significant (it accounts for the 4.6% of the total variance) and due to socio-economic structural differences among Regions, as measured by GDP per capita.

To our knowledge, the only analysis on primary schools' effectiveness is that conducted by Grilli and Sani (2011) on Invalsi data. As in many of the above-cited studies, the methodological approach is a multilevel model. The authors considered heteroscedastic variance components, thus allowing the pupil-level variance to change with gender and the school-level variance to change with the geographical area. The estimates of the regression coefficients are in line with empirical analyses on different school grades: lower test scores are found for foreigners and pupils with a low economic, social and cultural background. Peer and contextual effects, very relevant in empirical analyses on secondary school, are confirmed to influence pupils' performance in primary school too. However, the analysis cannot take into account school-level variables other than compositional ones (i.e., averages of the same variables inserted at the individual level); indeed, the data used do not include information on school characteristics and resources, such as the number of pupils per teacher or the availability of school facilities, thus hindering the analysis of the school effect.

3. Methodological approach

In this paper, we use a multilevel approach to analyze schools' and pupils' performances, taking into account the hierarchical structure of the data: pupils nested within schools[4]. Multilevel models) represent a good solution for studying the relationship between outputs and contextual and organizational variables in complex hierarchical structures, considering both individual and aggregate levels of analysis. The use of multilevel modeling prevents some common errors in the interpretation of individual data nested within larger units, such as interpreting at individual level some variables obtained by aggregating data at higher level and interpreting group' effects by using individual-level data. Indeed, the multilevel regression analysis estimates a regression equation which takes into account the correlation of the responses of the pupils of the same school, thus obtaining more correct estimates of the role played by different factors in determining pupil's attainment (Goldstein, 2003; de Leeuw and Meijer, 2008).

---

[4] Considered the richness of our database, which contains data at the individual, class, school and territory, we have also attempted to exploit its potential by estimating a three level model.
A first attempt concerned the insertion of the class level, which showed that in the empty model the percentage of variance explained by the class is almost equal to the one explained by the school. However, when the second level covariates are added the variance between classes remains unchanged, indicating that the variability of learning between classes depends on factors other than the class composition, such as the quality of the teacher. Despite the relevance of the result, the model suffered from the low number of classes per school (1.6), due to the fact that the 33% of schools has only one class, and was thus set aside.
A second attempt concerned the insertion of the territorial level, represented by Zonal Conferences. A brief description of this model's results are presented in section 5.1.

In the two-level linear random intercept model, the role of hierarchy can be summarized by the Intraclass Correlation Coefficient (ICC), which is defined as the ratio of the level 2 variance on the total variance. We follow this methodological approach, considering four steps: in the first step, we estimate an "empty" model, to decompose the total variance into the student-level (within) variance and school-level (between) variance, while in the second and third steps we add explanatory variables respectively at student and school level. At the last step, we insert territorial dummies in order to check for territorial variability among test scores.

We estimate the random slope model (1) for both the math and reading scores:

$$Y_{ij}=\alpha+\beta' X_{ij}+\gamma' W_j+u_j+e_{ij}, \quad (1)$$

where Y_{ij} is the outcome of pupil *i* in class *j (*reading or math), $i=1,2,\ldots,n_i$, $j=1,2\ldots,J$, α is the intercept, X_{ij} is the vector of level 1 (pupil) covariates, including the territorial dummies, W_j is the vector of level 2 (school) covariates, β and γ are the corresponding vectors of fixed parameters, u_j is the level 2 residual and e_{ij} is the level 1 residual. The model assumes independent and identical normal distributed residuals, i.e. u_j~IID-N($0,\tau^2$), e_{ij} ~IID-N($0,\sigma^2$), and cov(u_j, e_{ij})=0.

4. Data description

4.1 The database

In this paper we exploit a database on Tuscan public schools obtained merging the Invalsi dataset with other sources of information at the school level. In particular, we use information concerning primary schools.

The Invalsi dataset contains individual data on the 2010/2011 test carried out on fifth grade pupils. The test includes 61 multiple-choice items for reading and 47 for mathematics; test scores have been standardized into a range [0;100], representing the percentage of right answers to the items. For each pupil, the dataset contains the math and reading standardized score and individual characteristics: gender, age, nationality, whether the student attended to nursery or kindergarten, the province where he/she lives and some variables on family characteristics.

The Invalsi dataset also contains information at the class level. They concern size and composition of the class: total number of pupils, number of foreign pupils (distinguishing between first and second generation), number of disabled and of repeating pupils.

The original data set has been matched at school level with the administrative data available on the website of the Italian Ministry of Education, concerning financial, instructional and human resources employed by public schools. For primary education, financial and human resources are available at the level of school institutions and have been attributed to single schools on the basis of the number of pupils. Information on the quality of school buildings, coming from the Tuscan Register of school buildings, were also merged to the main dataset.

The resulting database was also merged to variables concerning the geographical, social and economic context in which the school operates, collected at municipal level. As Invalsi does not communicate information at school-level, but only at Province level, Invalsi itself merged for us our Municipality-Based-Dataset with the main Invalsi dataset, removing school identifiers and then returning the output to us in the form of anonymous data.

The final database contains data for 25,720 pupils nested in 871 public schools.

In order to stabilize the sample from a step to another of the multilevel analysis, this database has been "cleaned" by removing all records with at least a covariate missing. Therefore, our final database is made up of 22,005 pupils with information on all covariates and on at least one of the two test scores. In particular, the math-database comprises 21,300 pupils nested in 1,230 classes and 766 schools and the reading-database includes 21,525 pupils nested in 1,243 classes and 772 schools[5].

4.2 Relevant variables

Pupil covariates considered in model (1) are:

[5] The number of classes and schools differs in the math and in the reading database because some classes and schools have missing values in one of the two test scores.

- Male, a dummy taking value 1 if the pupil is male and 0 if female;
- ESCS, a proxy variable of socio-economical status, built by Invalsi trough a principal component analysis of three indicators: employment status of pupil's parents, the level of education of pupil's parents and the possession of a range of specific goods[6]. The ESCS variable has been standardized with mean equal to 0 and standard deviation equal to 1 (Campodifiori et al., 2010). Among Tuscan pupils, the ESCS value ranges between -3.14 and 2.61, with an average of 0.11, slightly higher than the Italian average of zero;
- Foreign, a dummy variable taking value 1 if the pupil is foreign and 0 otherwise;
- Repeating, a dummy variable taking value 1 if the pupil has a delay in the scholar path and 0 otherwise;

To control for peer effects, we include in the model some composition variables, which are identified at the class level instead of the school level, the latter usually revealing little explicative power (Bratti et al. 2007). Class variables are:

- Class size, a categorical variable taking value 1 if the number of pupils per class is less than 10, 2 if it is between 10 and 25, and 3 if it is higher than 25;
- Disabled, number of disabled pupils per class;
- Repeating, number of pupils held in the same grade for an extra year rather than being promoted to a higher grade, per class.

The school-level covariates we insert in model (1) are both composition variables, obtained as the average of pupil level covariates, and variables measuring the resources employed in the education process.

- School average of pupil ESCS;
- School building's status, a continuous variable which ranges between 1 (if the building needs radical interventions in fixtures or structure) and 6 (if the building's status is optimal)[7];
- Fixed-term teachers, a three category ordinal variable taking on the following values: 1 if the incidence of fixed-term teachers over the total number of teachers in the school is lower than the 25th percentile (i.e. 9.5%); 2 if the incidence is between the 25th and the 75th percentile (i.e., between 9.5% and 20.8%) and 3 if it is higher than the 75th percentile;
- % of teachers over 55, a continuous variable;
- Financial resources per pupil, which include all resources not aimed at the payment of the contractual wages.

The data set contains other variables measuring school resources, in particular the number of computers per student and the student/teacher ratio. When added to model (1), these covariates are not statistically significant, so they are not considered in the final model.

In order to control for territorial differences within Tuscany, we consider two different sets of dummies (one at a time).

- Dummies for Zonal Conferences, 35 territorial units aimed at the coordination and planning of the Tuscan education system at the local level,
- Remote areas, a dummy taking value 1 if the municipality where the school is located is considered a "remote area", on the basis of the presence or lack of some basic services (hospital, school, train station) in the nearby, and 0 otherwise.

Tables 1 and 2 report descriptive statistics on dependent variables and covariates. In particular, they show that both dependent variables range between 0 and 100, with the math score presenting a slightly lower average value and a higher variability than the Italian one.

[6] More specifically, these "goods" concerns: a quiet place to study, a personal desk for homework, encyclopedias, internet connection, burglar alarm, a room exclusively devoted to the student, more than one bathroom, more than one car in the family, more than one hundred books.

[7] For an analytical description of the variable's construction method, see Irpet (2012).

Table 1. Descriptive statistics on continuous variables

Variable	Level	Mean	Std. Dev.	Min	Max	Obs
Math Score	individual	69.68	16.61	0	100	21,300
Reading Score	individual	73.99	14.08	0	100	
ESCS	individual	0.11	0.96	-3.14	2.61	22,005
% teachers over 55	School	28.16	9.21	0.00	55.20	772
Repeating per class	class	1.00	1.11618	0	8	1,270
Disabled per class	class	0.9	1.132461	0	8	1,270
Number of pupils per class	class	19.28	4.978719	1	28	1,270
Average school ESCS	School	0.08	0.39	-1.52	1.47	772
School building's status	School	5.20	0.64	2.14	6.00	772
% fixed-term teachers	School	15.90	8.71	0.00	45.71	772
Financial resources	School	409.28	340.14	132.62	4386.00	772

Table 2. Descriptive statistics on categorical and dummy variables

Variable	Level	Proportion %	Obs
Male	individual	50,73	22,005
Foreign	individual	13.12	22,005
Repeating	individual	3.81	22,005
Full time pupils	individual	42.29	22,005
Small class	class	5.83	1,270
Medium class	class	90.31	1,270
Large class	class	3.86	1,270

5. Results

5.1 The multilevel model's results

Results of the multilevel analysis, obtained via the XTMIXED command of STATA (Stata, 2011) have been reported in Tables 3 and 4.

The total variability is higher in math scores than in reading scores. Empty model's results show that most part of the variance is at student-level, even though between-school variance is significantly different from zero, suggesting the existence of some degree of segmentation among schools. Indeed, 22.3% of the variance in math scores and 19.7% of that in reading scores is explained by between schools variance.

Pupil covariates (see column of B of Tables 3 and 4) reduce unexplained pupil-level variance and also unexplained between school variance, highlighting the importance of compositional effects. The insertion of school variables (see column of C of Tables 3 and 4) reduces between variance in both the math and the reading model, although the unexplained variability remains high, stressing the opportunity to look for the role of territorial factors in determining differences between schools. However, even after the insertion of territorial dummies (see columns of D and E of Tables 3 and 4), both in the form of Zonal Conferences and of remote areas, the model does not explain much between school variance in test scores. In the end, the model explains only a minor part of the total variance (10.4% for math and 13% for reading). At the pupil level the main unobserved factors can be reasonably identified with the pupil's ability of learning and inclination to study. School level residuals can be interpreted as school's effectiveness, adjusted for the available explanatory variables. In turn, school effectiveness is strictly linked to the ability and vocation of teachers and to management quality, factors which are difficult to measure.

Table 3. Two level linear model on math score

Variable	MODEL				
	A	B	C	D	E
	Empty	pupill variables	+ school variables	+ Zonal Conferences dummies	+ remote areas dummies
Constant	***69.3****	***68.1****	***63.1****	***65.7****	***62.9****
Male		2.17***	2.17***	2.17***	2.17***
Escs		3.93***	3.87***	3.88***	3.87***
Foreigner		-2.93***	-2.9***	-2.90***	-2.90***
Repeating		-3.48***	-3.26***	-3.24***	-3.26***
Full time		0.28	0.09	0.09	0.10
Disabled pupils per class			0.51***	0.50***	0.52***
Repeating pupils per class			-0.45***	-0.42***	-0.45***
Class size: 10-25 pupils			2.53*	1.89	2.61*
Class size: more than 25 pupils			4.19***	3.47**	4.26***
Average school escs			2.34***	2.13**	2.35***
School building status			0.73	0.54	0.74
% of fixed-term teachers: medium			-0.38	-1.43	-0.37
% of fixed-term teachers: high			-2.39***	-4.13***	-2.41***
% of teachers older than 55			-0.03	-0.05	-0.03
Financial resources per pupil			0.00	0.00	0.00
Territorial dummies	no	no	no	yes	yes
Random effects					
Between variance	64.72	59.88	57.07	52.00	57.05
Within variance	219.98	203.23	203.03	203.06	203.03
Total variance	284.70	263.12	260.10	255.06	260.08
% between over total (ICC)	22.7%	22.8%	21.9%	20.4%	21.9%
% change in within variance		-7.6%	-0.1%	0.0%	0.0%
% change in between variance	-	-7.5%	-4.7%	-8.9%	0.0%
LR test vs. linear regression: chibar2(01)	3262.96	3280.39	3037.18	2723.97	3036.87

Table 4. Two level linear model on reading score

Variable	MODEL				
	A	B	C	D	E
	Empty	pupill variables	+ school variables	+ Zonal Conferences dummies	+ remote areas dummies
Constant	***74.1****	***75.3****	***76.0****	***76.8****	***76.3****
Male		-0.33**	-0.33*	-0.33**	-0.33*
Escs		3.38***	3.35***	3.35***	3.35***
Foreigner		-5.09***	-5.08***	-5.07***	-5.07***
Repeating		-5.46***	-5.38***	-5.38***	-5.38***
Full time		-1.27***	-1.39***	-1.30***	-1.40***
Disabled pupils per class			0.65***	0.63***	0.64***
Repeating pupils per class			-0.19*	-0.18	-0.19*
Class size: 10-25 pupils			0.62	0.38	0.42
Class size: more than 25 pupils			0.90	0.60	0.69
Average school ESCS			0.95	0.98	0.89
School building status			-0.23	-0.46	-0.25
% of fixed-term teachers: medium			-0.86	-1.14	-0.86
% of fixed-term teachers: high			-2.95***	-3.55***	-2.91***
% of teachers older than 55			0.02	0.00	0.02
Financial resources per pupil			0.00	0.00	0.00
Territorial dummies	no	no	no	yes	yes
Random effects					
Between variance	40.04	35.62	33.58	31.11	33.53
Within variance	163.40	146.08	145.94	145.87	145.93
Total variance	203.44	181.70	179.52	176.98	179.47
% between over total (ICC)	19.7%	19.6%	18.7%	17.6%	18.7%
% change in within variance	-	-10.6%	-0.1%	0.0%	0.0%
% change in between variance	-	-11.0%	-5.7%	-7.4%	-0.1%
LR test vs. linear regression: chibar2(01)	2713.34	2678.07	2433.71	2267.61	2428.15

Looking at regression coefficients of the two-levels model without dummies (column of C of Tables 3 and 4), the role of cultural and socio-economic background of pupils in explaining educational attainments is highlighted; ESCS gets a positive and highly significant coefficient in both the reading and the math model. Pupils' demographic features appear to be very relevant. On average, male pupils perform significantly better in math than their females counterparts (+2.17), while they show a slightly worse performance in reading (-0.33). Foreign pupils have lower performances, particularly in reading (-5.1), as expected, but also in math (-2.5). Also the fact of having repeated one or more years negatively impacts on test scores (-3.26 and -5.38 for math and reading respectively). With regard to the type of education received, full-time courses have a negative influence only on reading attainment (-1,27), whereas they are not significant for math's score.

Variables inserted to catch peer group effects reveal high statistical significance especially in the math model. Disabled classmates appear to exert a positive effect on pupils' scores; such an unexpected result can be explained by the role played by the support teacher in classes with disabled pupils. On the contrary, repeating classmates appear to have a negative, although limited, impact on pupils' attainments in reading and math[8]. The categorical variable on class size provides very interesting results in the math model, showing that being in a medium or big class significantly advantages pupils. Such a result can be partly explained by the fact that most small classes (under 10 pupils) are grouped together with other grades' classes into a single multi-grade class, a context which is certainly not an optimal learning environment.

With respect to school level variables, the model estimation confirms the important influence that school's average cultural-socio-economic level exerts on pupils' performance in math (+2,34), while no significant effect is found for reading. In addition, the analysis shows that school's resources do not play a significant role in determining test scores: indeed, the school building's status, the age of teachers and financial resources do not get a statistically significant coefficient. On the contrary, the variable about fixed-term teachers appears to exert a statistical significant effect on both reading and math test scores, even if only when the percentage of fixed term teachers is higher than the 75th percentile. This finding seems to show that the quality of school's human resources matters, even though this quality lies more in the continuity of the relationship with pupils than in teachers' personal features.

Columns D and E of Tables 3 and 4 show estimation results of the models which include territorial dummies. Coefficients of dummies are not listed in tables, because they show very little (in the case of some Zonal Conferences) or none (in the case of remote areas) statistical significance. However, the aggregate significance of the Zonal Conferences dummies is confirmed by the likelihood-ratio test, while it is rejected for the remote areas dummies. Therefore, our results seem to indicate that it is not the location of the school in an urban / rural context to make the difference in pupils' educational attainment but rather the placement in a given administrative area.

These results appear to be confirmed also by our three-level estimation of pupils' test scores[9], where the third level is represented by Zonal Conferences; indeed, the between areas variance is very limited and not explained by territorial covariates (economic specialisation, average income), which do not exert any statistical significant effect on pupils' test scores.

Given the limited variance explained by the inclusion of Zonal Conferences dummies, in what follows we refer to results obtained with the model without territorial dummies (column C of Tables 3 and 4).

5.2 Expected test scores for different profiles

After having estimated the multilevel models, we computed the expected math and reading scores for some hypothetical pupil and context profiles, in order to better understand the importance of each type of determinant. In this exercise we consider a pupil with level 1 residuals equal to zero and a school with level 2 residuals equal to zero, i.e. average unobserved pupil and school characteristics.

---

[8] Among class composition variables, we also inserted the percentage of immigrant pupils: this covariate revealed a very little explicative power and its insertion reduced the impact of repeating pupils. In our interpretation, this is due to the fact that in most cases (70%) repeating pupils are foreigners, probably of recent immigration, held in the same grade for an extra year because of language difficulties ecc. We then conclude that the presence of foreigners without delay in a class does not significantly influence the performance of pupils.

[9] Results are not shown in this paper and are available from authors upon request.

We considered the following three pupil profiles:

- Lucky pupil: male in the case of math and female in the case of reading, Italian, non repeating and with a high ESCS (equal to the 90th percentile);
- Unlucky pupil: male in the case of reading and female in the case of math, foreign, repeating and with a low ESCS (equal to the 10th percentile);
- Medium pupil: Italian, non repeating and with a medium ESCS (equal to the 50th percentile);

And the following context profiles:

- Good school: medium sized class, no repeating students in class, high average school ESCS (equal to the 90th percentile), low percentage of fixed-term teachers;
- Bad school: small sized class, two repeating students in class, low average school ESCS (equal to the 10th percentile), high percentage of fixed-term teachers;
- Medium school: medium sized class, one repeating student in class, medium average school ESCS (equal to the 50th percentile), medium percentage of fixed-term teachers.

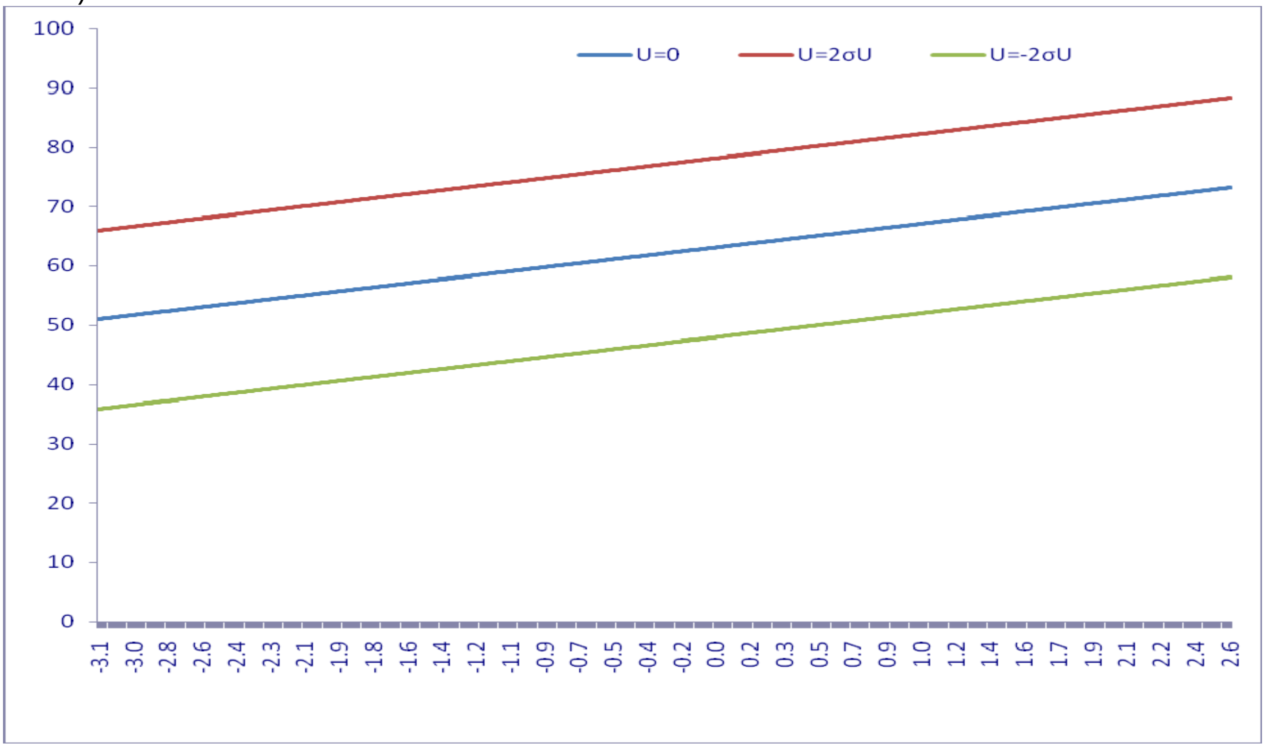
Table 5 reports the nine profiles obtained by crossing pupils profiles with context profiles. In this way, it is possible to compare the expected score of a certain type of pupil when attending a good, a medium or a bad school. For example, a lucky pupil's math score can range between 70.43 and 78.67, depending on the school's observable characteristics. The same happens with the reading score, which for an unlucky pupil can range between 56.75 and 61.99 depending on the type of school.

Table 5. Expected scores for different profiles of pupil and school

MATH				
		SCHOOL		
		Bad	Medium	Good
PUPIL	Unlucky	52.04	57.79	60.28
	Average	64.15	69.90	72.39
	Lucky	70.43	76.18	78.67
READING				
		SCHOOL		
		Bad	Medium	Good
PUPIL	Average	71.57	74.60	76.81
	Lucky	76.24	79.27	81.48
	Unlucky	56.75	59.78	61.99

However, test scores can differ significantly also in relation to unobservable factors and thus to school's effectiveness. To show how important the latter factor can be, Figure 6 reports regression lines of math scores with respect to ESCS for an average pupil in an average school for different values of level 2 residuals.

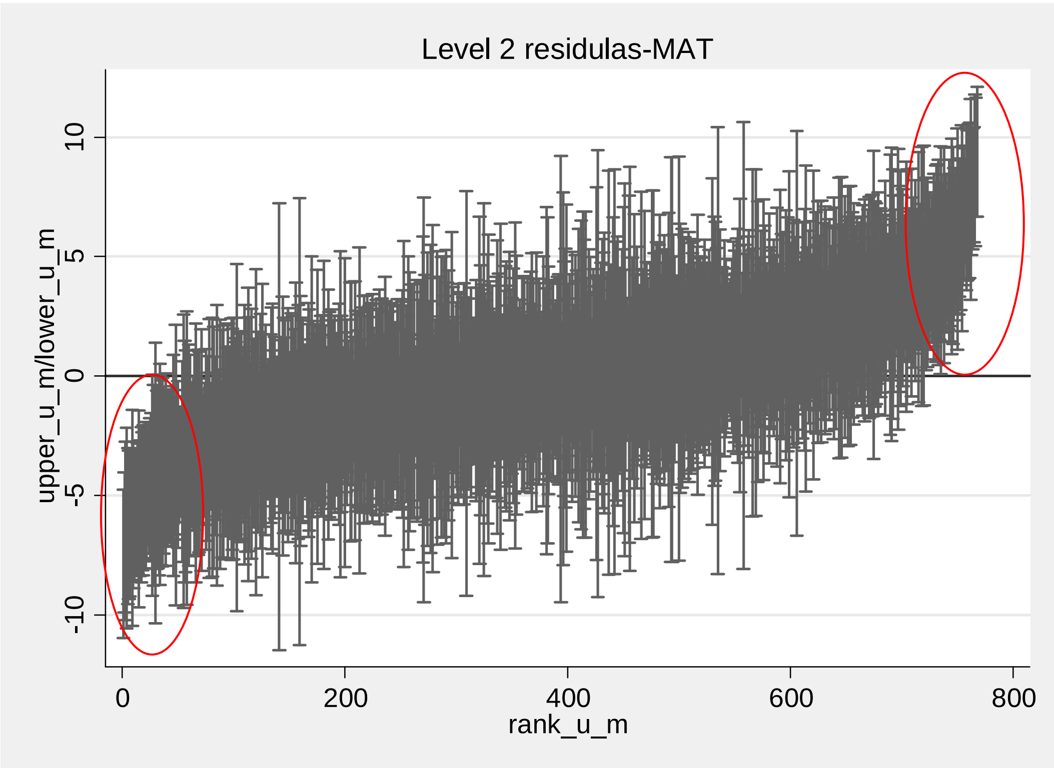
Figure. 6 Expected math scores for different pupil ESCS and school effectiveness (average pupil, average school)



5.3 School rankings

As already said, level 2 residuals can be interpreted as school effectiveness, conditional on observed pupil and context covariates. We ranked Tuscan schools according to their effectiveness, estimated separately for math and reading. Figure 7 shows the EB predictions of level 2 residuals from the math model alongside with their comparative confidence intervals: only few schools, in the upper right and lower left part of the graph, have a predicted residual significantly different from zero. These schools should be better investigated and monitored to correct bad practices (schools with a residual in the lower left part of the graph) and to discover determinants of good practice (schools with a residual in the upper right part of the graph).

Figure. 7 Ranking of Tuscan primary schools by level 2 residuals of the math model



Afterwards, from the two rankings we indentified the best (more effective) and worst (least effective) schools in reading and in math. Then, we selected the 45 Tuscan best schools, those resulting effective from both the math model and the reading model; similarly, we selected the 22 Tuscan worst schools, those resulting ineffective both from the math model and from the reading model.

The identification and mapping of such schools in the Tuscan region is hindered by the anonymity imposed by the privacy policy of Invalsi. Nonetheless, the distribution by Zonal Conference (Table 8) shows that the best and worst schools of Tuscany are not homogeneously distributed among geographical areas. This is a very important result for the evaluation of the regional education system, which can be used by policy makers to identify areas deserving particular attention. Obviously, this instrument could be strengthened by allowing researchers and policy makers to identify the effectiveness of single schools.

Table 8. Best and worst primary schools by Zonal Conference

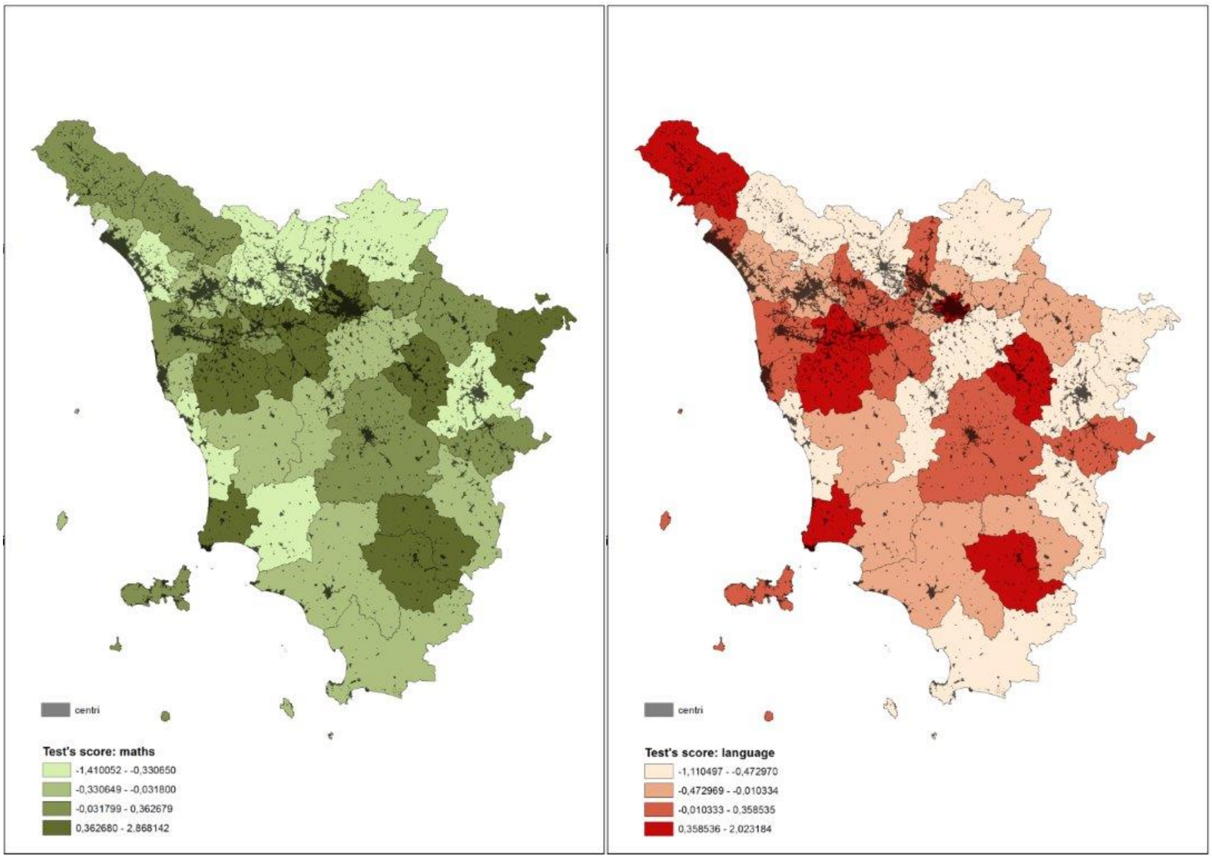
		BEST		WORST	
Zonal Conference	N. of schools	N.	%	N.	%
Alta Val d'Elsa	13	0	0%	0	0%
Amiata - Val d'Orcia	7	0	0%	0	0%
Amiata Grossetana	7	0	0%	0	0%
Apuane	45	3	7%	2	4%
Aretina	18	1	6%	0	0%
Bassa Val di Cecina	14	0	0%	2	14%
Casentino	10	0	0%	0	0%
Colline dell'Albegna	15	1	7%	0	0%
Colline Metallifere	17	0	0%	1	6%
Elba	6	0	0%	0	0%
Empolese	35	4	11%	0	0%
Fiorentina Nord-Ovest	28	3	11%	0	0%
Fiorentina Sud-Est	21	0	0%	0	0%
Firenze	39	4	10%	0	0%
Grossetana	23	1	4%	1	4%
Livornese	23	5	22%	2	9%
Lunigiana	20	1	5%	0	0%
Mugello	14	0	0%	0	0%
Piana di Lucca	29	1	3%	2	7%
Pisana	46	3	7%	2	4%
Pistoiese	47	0	0%	4	9%
Pratese	43	3	7%	2	5%
Senese	21	1	5%	1	5%
Val d'Era	39	3	8%	0	0%
Val di Cecina	12	0	0%	0	0%
Val di Chiana Aretina	19	0	0%	1	5%
Val di Chiana Senese	9	0	0%	0	0%
Val di Cornia	10	2	20%	0	0%
Val di Nievole	28	4	14%	1	4%
Val Tiberina	5	1	20%	0	0%
Valdarno	21	2	10%	0	0%
Valdarno e Valdisieve	11	0	0%	1	9%
Valdarno Inferiore	15	1	7%	0	0%
Valle del Serchio	28	1	4%	0	0%
Versilia	34	0	0%	0	0%

Note: the number of schools refers to the number of schools in our database after the cleaning process and not to the real number of schools in each Zonal Conference

5.4 Primary schools' effectiveness by Zonal Conferences

In order to map primary schools' effectiveness in different areas of Tuscany, we calculated average values of level 2 residuals, which are shown in Figure 9 and 10. The maps highlight the lower effectiveness of schools located in the northern and eastern borders of the region and generally in the empty South, with a few exceptions. On the contrary, the more urbanized Zonal Conferences generally reveal higher levels of average schools' effectiveness.

Figures 9-10. Average level 2 residuals by Zonal Conference, from the math and the reading model



6. Conclusions

In this paper we analysed the determinants of Invalsi test scores for 22,005 pupils in 772 primary schools of Tuscany. Our analysis benefited from a novel database which gathers data on test scores, individual and school characteristics and contains information at the level of very small territorial units, such as Zonal Conferences.

We choose to use a multilevel approach, to properly account for the hierarchical structure of data, confirmed by our empty model, which shows that a significant part of test scores' variance is explained by between-schools variance.

Our model shows that individual variables tend to explain most of the variance in pupils' achievements: being foreign, repeating and with a low socio-economic and cultural family background drastically reduces both math and reading test score. When looking at the school-level variables, it turns out that the composition of the student body (inserted to catch the so called "peer effect") matters much more than school's resources; however also the quality of resources appears to exert a certain influence on the achievement of pupils, especially when it concerns the stability of the teaching body. Finally, territorial differences appears to matter little for pupils' test scores in Tuscan primary school; this is a positive result, which seems to shows the limited role of pupils' residential area in determining their educational attainments. However, we should consider the fact that some school level covariates already explain differences in the performance of pupils' attending schools in different areas of Tuscany. Thus, a pupil attending a school in a remote area is disadvantaged by the typical characteristics of a marginal school (high percentage of fixed-term teachers, due to self-selection processes, small class size) and not by the school's location itself.

The ranking of schools according to their residual, interpreted as their level of effectiveness once accounted for observed variables, has revealed a non homogenous distribution of effective and ineffective schools among areas. This is a very important result for the evaluation of the regional schooling system, which could be used by the policy maker to identify the areas deserving particular attention. Clearly, it would be desirable to inform policy makers of the single school's effectiveness, in order to allow targeted actions.

Given the availability of such a rich database, we want to conclude by specifying our intentions for further research. First, of all, our analysis on Tuscan primary schools could be deepened by estimating a bivariate multilevel model, in order to decompose the covariances between the two scores, thus estimating the unexplained correlation due to pupil and school levels respectively. Second, we intend to continue our analysis of Tuscan schools' effectiveness by considering secondary schools, where the higher average school size could allow the insertion of the class level and the estimation of a three multilevel model, in order to look for within school segmentation and for a residual to be interpreted as teacher's effectiveness.

Bibliografia

Agasisti, T. (2011), "The effect of competition on schools' performance: preliminary evidence from Italy through OECD-PISA data", *European Journal of Education*, Vol. 46, Issue . 4, pp. 549-565.

Agasisti, T. (2013), "Competition Among Italian Junior-Secondary Schools: A Variance-Decomposition Empirical Analysis", *Annals of Public and Cooperative Economics*, Vol. 84, issue 1, pp. 17-42.

Agasisti, T. and Sibiano, P. (2013), "Efficiency and heterogeneity of public spending in education among Italian regions", *Journal of Public Affairs*, Volume 13, Issue 1, pp 12–22.

Agasisti, T., Bonomi, F. and Sibiano, P. (2012), Adjusted efficiency measures for schools: a two-stage empirical analysis with Bootstrap DEA and Tobit regression, paper presented at the 37th AEFP Annual Conference, 15th-17th March, Boston, MA.

Barbetta G.P. and Turati G. (2003), "Efficiency of junior high schools and the role of proprietary structure", *Annals of Public and Cooperative Economics*, Vol. 74, Issue 4, pp. 529-551.

Barbieri, G., Cipollone, P. and Sestito, P. (2007), "Labour Market for Teachers: Demographic Characteristics and Allocative Mechanisms", *Giornale degli Economisti, GDE (Giornale degli Economisti e Annali di Economia)*, Bocconi University, Vol. 66, Vol. 3, pp. 335-373.

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

Boozer, M. A. and Rouse, C. E. (2001), "Intra school variation in class size: Patterns and implications", *Journal of Urban Economics*, Vol. 50, pp. 163-189.

Bratti, M., Checchi, D. and Filippin, A. (2007), *Da dove vengono le competenze degli studenti? I divari territoriali nell'Indagine OECD-PISA 2003*, Bologna: Il Mulino.

Campodifiori, E., Figura, E., Papini, M. and Ricci, R. (2010), An indicator for students' socio-economic background, Invalsi Working Paper n. 02/2010, Invalsi.

Castellano R., Quintano C. and Longobardi S. (2009), "L'influenza dei fattori socio-economici sulle competenze degli studenti italiani. Un'analisi multilevel dei dati PISA 2006", *Rivista di Economia e Statistica del territorio,* Vol. 2, pp. 109–149.

Castellano, R., Longobardi, S. and Quintano, C. (2010), The Efficiency of Italian Schools: a Stochastic Frontier Analysis, paper presented at the DYSES International Conference, Benevento, September 20-25.

Checchi, D. (2004), "Da dove vengono le competenze scolastiche? L'indagine PISA 2000 in Italia", *Stato e mercato*, Vol. 72, pp. 413-453.

Cipollone P., Montanaro P. and Sestito P. (2010), *Misure di valore aggiunto per le scuole superiori italiane: i problemi esistenti e alcune prime evidenze*, Temi di discussione Bank of Italy n. 754.

de Leeuw J. and Meijer E. (2008), *Handbook of Multilevel Analysis*, Springer, New York.

De Simone, G. (2011), Render unto primary the things which are primary's. Inherited and fresh learning divides in Italian lower secondary education, FGA Working Paper n. 44, Fondazione Giovanni Agnelli.

Di Giacomo, G. and Pennisi, A. (2011), Efficienza e autonomia scolastica nella suola primaria e secondaria di primo grado italiana, paper presented at the 33th Conference of Società italiana di economia pubblica, Pavia, 19-20 September.

Ehrenberg, R., *et al.* (2001), "Class Size and Student Achievement", *Psychological Science in the Public Interest*, Vol. 2, No. 1, pp. 1-30.

Ferrer-Esteban, G. (2011), Beyond the traditional territorial divide in the Italian Education System, FGA Working Paper n. 43, Fondazione Giovanni Agnelli.

Finn, J. (1998), "Class Size and Students at Risk: What is Known? What is Next?", US Department of Education, Office of Educational Research and Improvement, National Institute on the Education of At-Risk Students, Washington, DC.

Fitz-Gibbon, C. (1997), *The Value Added National Project Final Report: Feasibility Studies for a National System of Value-Added Indicators*. London: School Curriculum and Assessment Authority.

Fuchs, T. and Woessmann, L. (2007), "What Accounts for International Differences in Student Performance? A Re-examination using PISA Data", *Empirical Economics*, Vol. 32, Issues 2-3, pp. 433-464.

Gavosto, A. and De Simone, G. (2013), "Patterns of Value-Added Creation in the Transition from Primary to Lower Secondary Education in Italy", FGA Working Paper n. 47, Fondazione Giovanni Agnelli.

Goldstein, H. (2003), *Multilevel statistical models*, 3rd edn. Arnold, London.

Grilli, L. and Rampichini, C. (2009), "Multilevel models for the evaluation of educational institutions: a review", in Bini M., Monari P., Piccolo D., Salmaso L., *Statistical Methods for the Evaluation of Educational Services and Quality of Products*, pp. 61-80, Heidelberg: Physica-Verlag HD.

Grilli, L. and Sani, C. (2011), "Differential variability of test scores among schools: a multilevel analysis of the fifth-grade Invalsi test using heteroscedastic random effects", *Journal of Applied Quantitative Methods*, 6 (4), pp 88-99.

Hanushek, E. A. (1997), "Assessing the effect of school resources on student performance: An update", *Educational Evaluation and Policy Analysis*, Vol. 19, pp. 141-164.

Hanushek, E. A. (2003), "The failure of input-based schooling policies," *Economic Journal*, Vol. 113, pp. F64-F98.

Hanushek, E. A., Kain, J. F., Markman, J. M and Rivkin, S. G. (2003), "Does Peer Ability Affect Student Achievement?" *Journal of Applied Econometrics,* Vol. 18, Issue 5, pp. 527-544.

MEF (2012), *Analisi dell'efficienza delle scuole italiane rispetto agli apprendimenti degli studenti - differenze territoriali e possibili determinanti*, Ministero dell'Economia e delle Finanze, Dipartimento della Ragioneria generale dello Stato, Servizio Studi, Roma.

Minzyuk, L. and Russo, F. (2012), The causal effect of class size on pupils' performance: evidence from Italian primary schools, paper presented at the 34th Conference of Società italiana di economia pubblica, Pavia, 24-25 September.

Montanaro, P. (2008), I divari territoriali nella preparazione degli studenti italiani: evidenze dalle indagini nazionali e internazionali, Bank of Italy, Occasional Papers, n. 14.

Murnane R.J. and Willet J.B. (2011), *Methods Matter: Improving Causal Inference in Educational and Social Science Research*, Oxford University Press: New York.

OECD PISA (2009), *PISA 2009 Results: What Students Know and Can Do—Student Performance in Reading, Mathematics and Science*. OECD: Parigi.

Piketty, T. and M. Valdenaire (2006), *L'Impact de la taille des classes sur la réussite scolaire dans les écoles, collèges et lycées français : Estimations à partir du panel primaire 1997 et du panel secondaire 1995*, ministère de l'Éducation nationale, de l'Enseignement supérieur et de la Recherche, Direction de l'évaluation et de la prospective, Paris.

Raudenbush S. and Willms, J. (1995), "The estimation of school effects", *Journal of Educational and Behavioral Statistics,* Vol. 20, pp. 307-335.

Sanders, W., Saxton, A. and Horn, B. (1997), "The Tennessee Value-Added Assessment System: A Quantitative Outcomes-Based Approach to Educational Assessment", in J. M. (Ed.), *Grading Teachers, Grading Schools: Is Student Achievement a Valid Evaluational Measure?*. Thousand Oaks, CA: Corwin Press, Inc.

Schneider, K. (2004), "International differences in student achievement: an economic perspective", *German Economic Review*, Vol. 5, Issue 3, pp. 357-380.

Schnepf, S.V. (2007). 'Immigrants' educational disadvantage: an examination across ten countries and three surveys'*, Journal of Population Economics*, Vol. 20, Issue 3, pp. 527-545.

Stata (2011), *Stata Longitudinal-Data/Panel-Data Reference Manual*, Stata Press.

Vandenberghe, V. and Robin, S. (2004), "Evaluating the effectiveness of private education across countries: a comparison of methods", *Labour Economics*, Vol. 11, pp.487-506.

Wossman, L.(2003), "Schooling Resources, Educational Institutions and Student Performance: The International Evidence", *Oxford Bulletin of Economics and Statistics*, Vol. 65, Issue 2, pp.117-170.

Wößmann, L. (2004), "How Equal Are Educational Opportunities? Family Background and Student Achievement in Europe and the United States," IZA Discussion Papers 1284, Institute for the Study of Labor (IZA).

Wößmann, L. and West, M. R. (2002), "Class-Size Effects in School Systems Around the World: Evidence from Between-Grade Variation in TIMSS", Discussion Paper No. 485, April.

Zimmer, R. W. and Toma, E. F. (2000), "Peer effects in private and public schools across countries", *Journal of Policy Analysis and Management* , Vol. 19 (Winter), pp. 75-92.