The future has early roots.

Learning outcomes and school's effectiveness in Tuscany's primary education system

Enrico Conti (Irpet)

Silvia Duranti (Irpet)

Maria Luisa Maitino (Irpet)

Carla Rampichini (Università degli Studi di Firenze)

Nicola Sciclone (Irpet)

Workshop

Youth and their future: Work, Education and Health

October 17th -18th, 2013 - University of Salerno

1. Introduction

The educational system has a key role in influencing the future of young generations: indeed, it is school that provides the necessary skills and competences to successfully enter the labour market.

The increasing difficulties encountered by Italian young people in the transition to work (Irpet, 2013) have stimulated studies on the role that the educational system can play in facilitating or hampering the entrance to the labour market. Such studies usually focus on upper secondary school or university (Ungaro and Verzicco, 2005; Di Patrizio et al., 2010; Aina and Partore, 2012; Caroleo and Pastore, 2012), mainly looking at the role that the type of education received (general or vocational) can play in the transition to work. But young people's destiny is decided well before secondary school: the choice of the type of secondary education (and of university) itself strongly depends on past attainment's levels, which, in turn, are largely explained by one's family's characteristics (Checchi and Flabbi, 2007; Mocetti, 2007; Giuliano, 2008).

This is why we think that it is from the earliest years of education that the schooling system can significantly influence pupils' future prospects. Primary school should indeed be able to guarantee all pupils, irrespective of their family background, at least a standard level of learning outcomes, thus providing them the necessary tools to be successful both in upper school grades and, eventually, in the labour market. Acknowledging the key role played by primary school in influencing pupils' future prospects, in this paper we intend to analyse the effectiveness of public primary schools in Tuscany.

In this paper, 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. Our analysis of school effectiveness is done by 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". Such an analysis is possible thanks to the availability of data on scores of Invalsi reading and mathematics tests, now yearly 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)[1].

Our aim is to describe the main determinants of Grade 5 pupils' outcomes, 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, but also quality and quantity of school resources, which in Italy are mainly managed by the Ministry of Education) and school-manageable factors. 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 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, created on the basis of information

[1] 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.

provided by school head teachers (Castellano et al., 2009; Benadusi et al., 2010; Agasisti and Vittadini, 2012; Agasisti, 2013). A second innovative aspect 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 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

[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).

[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.

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.

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' 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 pupils' outcomes (Goldstein, 2003; de Leeuw and Meijer, 2008).

Our estimation procedure follows 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 spatial dummies in order to check for spatial variability among test scores.

We estimate the 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 spatial 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. $u_j \sim \text{IID-N}(0,\tau^2)$, $e_{ij} \sim \text{IID-N}(0,\sigma^2)$, and $\text{cov}(u_j, e_{ij})=0$.

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 of both private and public schools. Selecting public ones, we have information on the math test for 868 schools and on the reading test for 871 schools, which represent the 92% of Tuscan public schools. 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. For each pupil, the dataset contains the math and reading standardized score and individual characteristics: gender, age, nationality 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, number of disabled and of repeating pupils.

The original data set has been matched at the school level with the administrative data available on the website of the Italian Ministry of Education, concerning financial, instructional and human resources employed by all public schools.

[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 54.8% 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.

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 the municipal level. As Invalsi does not communicate information at school-level, but only at the 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 math-database comprises 23,406 pupils nested in 1,343 classes and 844 schools and the reading-database includes 23,665 pupils nested in 1,369 classes and 848 schools[5].

4.2 Relevant variables

Pupil covariates considered in model (1) are:

- 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)[7];
- Foreign, a dummy variable taking value 1 if the pupil is foreign and 0 otherwise;
- Late, a dummy variable taking value 1 if the pupil has a delay in the schooling career and 0 otherwise;
- LateXforeign, an interaction term, which takes value 1 if the pupil is both foreign and late;
- Full time, a dummy variable taking value 1 if the pupil attends a full-time class (40 hours) and 0 otherwise.

To control for the learning environment in which the pupil is inserted, we include in the model a variable on the size of the class, which is 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. We choose not to insert composition variables at the class level because the composition of classes might fall under the control of the head teacher, thus being part of the school's effectiveness. As our primary aim is not the identification of peer effects but instead the comparison of schools' effectiveness, we include composition variables at the school level, in order to control for the catchment area of the school, on which the head teacher has no influence.

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;
- Percentage of 5^{th} grade pupils who have a delay in the schooling career;
- Size of the municipality where the school is located, expressed as the inverse of the municipality's population;
- 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) [8];

[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.

[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] 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.

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

- 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;
- Percentage of teachers over 55.

In order to control for differences within Tuscany, we insert in the model dummies for Zonal Conferences, which are 35 units aimed at the coordination and planning of the Tuscan education system at the local level.

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.

Table 1. Descriptive statistics on continuous variables

	Level	U	Mean	Std. Dev.	Min	Max
Math Score	individual	n.	69.63	16.60	0.00	100.00
Reading Score	individual	n.	73.93	14.10	0.00	100.00
ESCS	individual	n.	0.11	0.97	-3.14	2.61
% late students in school's 5th Grade classes	School	%	5.87	6.83	0.00	50.00
Average value of ESCS per school	School	n.	0.08	0.40	-1.52	1.47
Conservation status of the scholar building	School	n.	5.20	0.63	2.14	6.00
% of Teachers with fixed-end contract per school	School	%	15.78	8.62	0.00	45.71
% teachers per school older than 55 years	School	%	28.16	9.21	0.00	55.20

Table 2. Descriptive statistics on categorical and dummy variables

	Level	Proportion %	Obs
Male	individual	50.94	24,199
Foreign	individual	13.2	24,199
Late	individual	3.8	24,199
Full time	individual	42.21	24,199
Small class	class	5.73	1,397
Medium class	class	90.55	1,397
Large class	class	3.72	1,397

5. Results

5.1 The multilevel model's results

Results of the multilevel analysis, obtained via the XTMIXED command of STATA (Stata, 2011) are shown 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, 22.4% of the variance in math scores and 19% of that in reading scores is explained by between schools variance.

Pupil covariates (see column 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 spatial factors in determining differences between schools. However, even after the insertion of spatial dummies (see column D of Tables 3 and 4) in the form of Zonal Conferences, 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 (10.2% for math and 12.8% for reading), leaving the most part of it in the residual, which thus collects all unobserved factors at the individual and school level. At the pupil level the main unobserved factors can be reasonably identified with the pupil's ability of learning and inclination to study, while 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

	A	B	C	D
	Empty	Invididual variables	+ school variables	+ Zonal Conferences dummies
Constant	*69.3****	*68.2****	*66.1****	*68****
Male		2.14***	2.14***	2.14***
Escs		3.95***	3.91***	3.91***
Foreign		-3.62***	-3.60***	-3.60***
Late		-8.01***	-7.92***	-7.87***
LateXforeign		6.24***	6.20***	6.14***
Full time		0.40	0.34	0.33
Class size: less than 10 pupils			-2.9**	-1.61
Class size: more than 25 pupils			1.47**	1.39**
% of late students in 5th grade classes			-0.09**	-0.07
Average school escs			1.87**	1.56*
School building status			0.857*	0.61
% of fixed-term teachers: medium			-0.743	-1.57*
% of fixed-term teachers: high			-2.60***	-3.95***
% of teachers older than 55			-0.05	-0.05
Inverse of municipality's population			3705**	3329*
Territorial dummies	no	no	no	yes
Between variance	63.61	58.97	56.31	52.20
Within variance	220.34	202.84	202.80	202.84
Total variance	283.95	261.81	259.11	255.04
% between over total	22.4%	22.5%	21.7%	20.5%
% change in between variance	-	-7.3%	-4.5%	-7.3%
% change in within variance	-	-7.9%	0.0%	0.0%
LR test vs. linear regression: chibar2(01)	3545.94	3588.28	3368.97	3240.89
Prob >= chibar2	0.00	0.00	0.00	0.00

Table 4. Two level linear model on reading score

	A	B	C	D
	Empty	Invididual variables	+ school variables	+ Zonal Conferences dummies
Constant	*74****	*75.2****	*77.7****	*77.8****
Male		-0.37**	-0.36**	-0.37**
Escs		3.36***	3.34***	3.34***
Foreign		-5.58***	-5.57***	-5.56***
Late		-7.66***	-7.55***	-7.55***
LateXforeign		3.19***	3.15***	3.14***
Full time		-0.90***	-0.87***	-0.78***
Class size: less than 10 pupils			-1.22	-0.91
Class size: more than 25 pupils			-0.02	-0.07
% of late students in 5th grade classes			-0.11***	-0.090**
Average school escs			0.49	0.45
School building status			-0.16	-0.37
% of fixed-term teachers: medium			-.948*	-1.25*
% of fixed-term teachers: high			-2.88***	-3.34***
% of teachers older than 55			0.00	-0.00
Inverse of municipality's population			854	1574
Territorial dummies	no	no	no	yes
Between variance	38.68	34.43	32.47	30.31
Within variance	164.79	147.19	147.22	147.18
Total variance	203.48	181.62	179.69	177.49
% between over total	19.0%	19.0%	18.1%	17.1%
% change in between variance	-	-11.0%	-5.7%	-6.6%
% change in within variance	-	-10.7%	0.0%	0.0%
LR test vs. linear regression: chibar2(01)	2876.81	2860.5	2648.94	2477.78
Prob >= chibar2	0.00	0.00	0.00	0.00

The categorical variable on class size provides very interesting results in the math model, showing that being in a small-sized class significantly disadvantages 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 the catchment area of the school exerts on pupils' performance, especially in math. Indeed, both the average school ESCS and the percentage of late students in 5^{th} grade have a statistically significant effect on test scores, which is positive in the former case and negative in the latter[9]. An interesting result is the one on the size of the municipality where the school is located, which seems to indicate that small-sized municipalities favour pupils' learning outcomes, probably thanks to a better institutional environment and to closer relationships between school and families.

With regard to school's resources, the analysis shows that they play a certain role in explaining pupils' test performance; however it appears to be more their quality than quantity to make the difference. Indeed, the number of computers per student, financial resources and the student/teacher ratio, inserted in an early version of the model, do not exert a statistically significant effect on pupils' outcomes. On the contrary, two variables used as proxies of resources' quality appear to be statistically significant: the school buildings' maintenance level exerts a positive and statistically significant impact on students' achievements (only in the math model) and the percentage of fixed-term teachers a negative one, even if only when it is higher than the 75th percentile. Instead, no effect is found for the age of teachers: students' outcomes do not appear to be influenced by a higher percentage of mature teacher. Together with the result on temporary teachers, this finding shows 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.

Column D of Tables 3 and 4 shows estimation results of the model with spatial dummies. Coefficients of dummies are not listed in tables, because they show very little statistical significance. However, the aggregate significance of the Zonal Conferences dummies is confirmed by the likelihood-ratio test, indicating that the placement in a given administrative area can make the difference in pupils' educational outcomes.

This result appears to be confirmed also by our three-level estimation of pupils' test scores[10], where the third level is represented by Zonal Conferences; indeed, the between areas variance is very limited and not explained by spatial 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 spatial 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 residual equal to zero and a school with level 2 residual equal to zero, i.e. average unobserved pupil and school characteristics.

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, low percentage of late students in the school (equal to the 10th percentile), high average school ESCS (equal to the 90th percentile), low percentage of fixed-term teachers (lower than the 25th percentile);
- Bad school: small sized class, high percentage of late students in the school (equal to the 90th percentile), low average school ESCS (equal to the 10th percentile), high percentage of fixed-term teachers (higher than the 75th percentile);

[9] Among school composition variables, we also inserted the percentage of immigrant pupils in 5th grade classes: this covariate revealed a very little explicative power and its insertion reduced the impact of late 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 the school does not significantly influence the performance of pupils.
[10] Results are not shown in this paper and are available from authors upon request.

- Medium school: medium sized class, medium percentage of late students in the school (equal to the 50th percentile), medium average school ESCS (equal to the 50th percentile), medium percentage of fixed-term teachers (between the 25th and the 75th percentile).

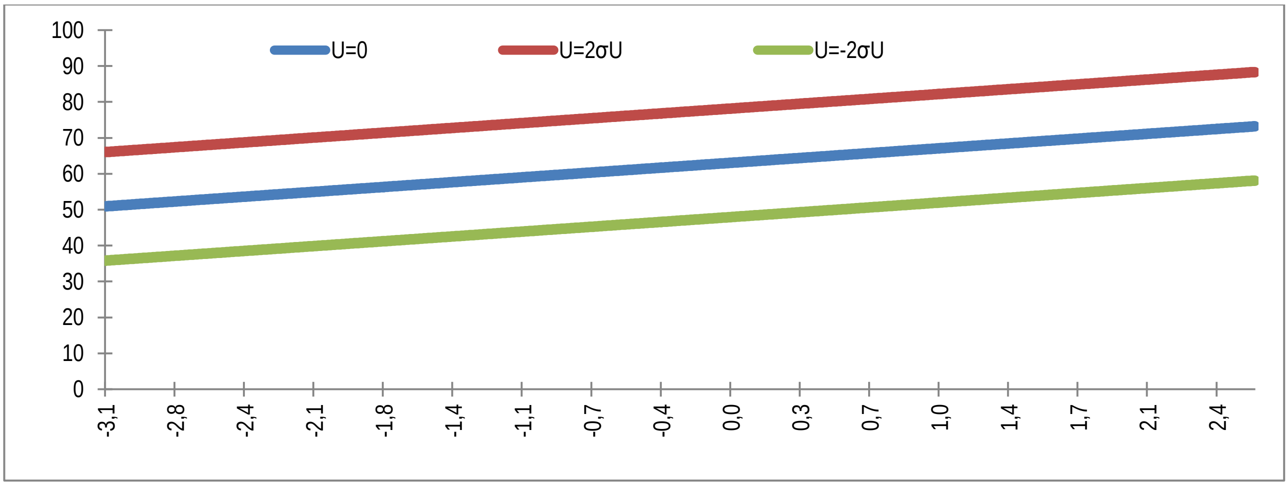
Table 5 reports the nine profiles obtained by crossing pupils profiles with context profiles. In this way, it is possible to compare the expected score of a certain type of pupil when attending a good, a medium or a bad school. For example, a lucky pupil's math score can range between 70.04 and 78.61, depending on the school's observable characteristics. The same happens with the reading score, which for an unlucky pupil can range between 56.39 and 62.71 depending on the type of school.

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

MATH				
		SCHOOL		
		Bad	Average	Good
PUPIL	Unlucky	52.50	58.33	61.00
	Average	63.81	69.64	72.30
	Lucky	70.04	75.94	78.61
READING				
		SCHOOL		
		Bad	Average	Good
PUPIL	Unlucky	56.39	60.32	62.71
	Average	70.72	74.65	77.04
	Lucky	75.39	79.33	81.72

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.

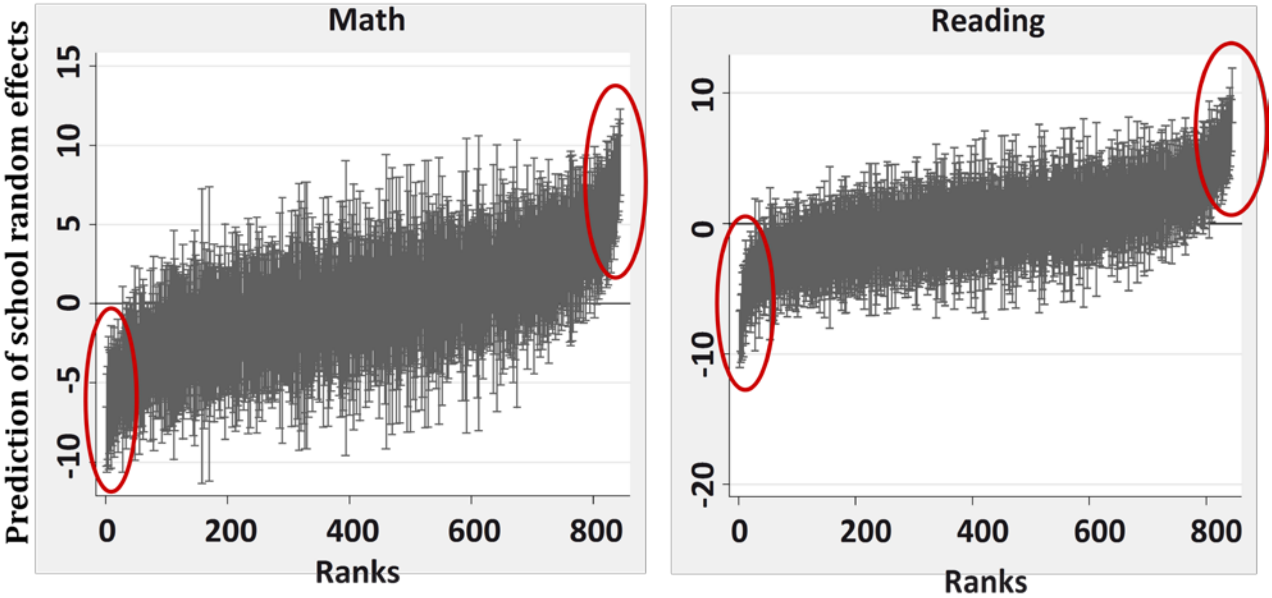
Figure. 6 Expected math score for different individual ESCS and school effectiveness



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 residual from the math and the reading 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 and the reading 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 46 Tuscan best schools, those resulting effective from both the math model and the reading model; similarly, we selected the 30 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.

In particular, some areas show a prevalence of good schools over bad ones (in green in Figure 9), while in some others bad schools outnumber good ones (in red in Figure 9). At the same time, many Zonal Conferences show a uniform degree of effectiveness of schools (in white in Figure 9); in some of them this is due to the presence of only medium-effective schools, in some others to the same number of good and bad schools.

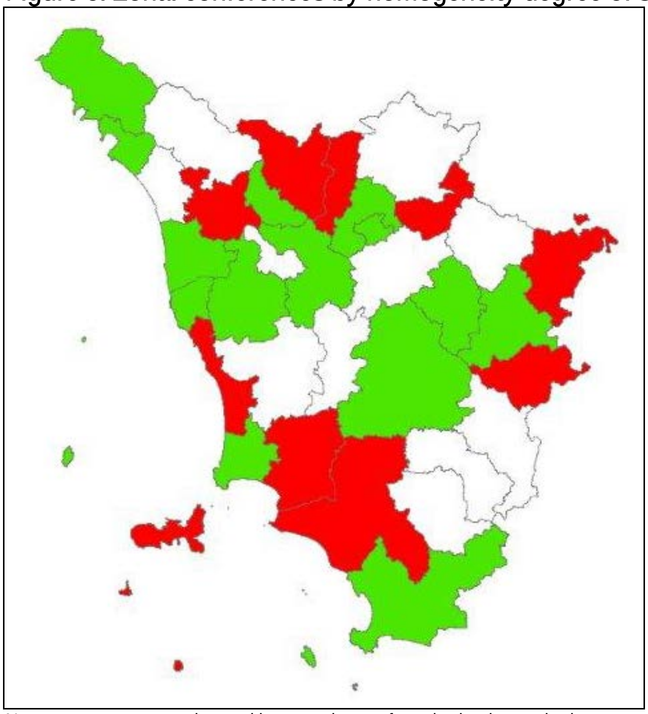
This is a very important result for the evaluation of the regional education system, which can be used by policy makers to identify areas deserving particular attention. Obviously, this instrument could be strengthened by allowing researchers and policy makers to identify the effectiveness of single 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	46	3	7%	2	4%
Aretina	35	1	3%	0	0%
Bassa Val di Cecina	14	0	0%	2	14%
Casentino	15	0	0%	0	0%
Colline dell'Albegna	16	1	6%	0	0%
Colline Metallifere	17	0	0%	1	6%
Elba	9	0	0%	1	11%
Empolese	37	5	14%	0	0%
Fiorentina Nord-Ovest	36	4	11%	1	3%
Fiorentina Sud-Est	25	0	0%	0	0%
Firenze	44	4	9%	2	5%
Grossetana	25	1	4%	2	8%
Livornese	23	5	22%	2	9%
Lunigiana	20	1	5%	0	0%
Mugello	14	0	0%	0	0%
Piana di Lucca	36	1	3%	2	6%
Pisana	46	3	7%	2	4%
Pistoiese	47	0	0%	4	9%
Pratese	44	2	5%	3	7%
Senese	22	2	9%	0	0%
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	12	0	0%	0	0%
Val di Cornia	12	1	8%	0	0%
Val di Nievole	28	4	14%	1	4%
Val Tiberina	9	0	0%	1	11%
Valdarno	23	2	9%	0	0%
Valdarno e Valdisieve	11	0	0%	1	9%
Valdarno Inferiore	15	1	7%	1	7%
Valle del Serchio	31	1	3%	1	3%
Versilia	39	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

Figure 8. Zonal conferences by homogeneity degree of schools' effectiveness



Notes: green areas are those with a prevalence of good schools over bad ones, red areas are those with a prevalence of bad schools over good ones and white areas are those with a uniform degree of effectiveness of schools.

6. Conclusions

In this paper we analysed the determinants of Invalsi test scores for 24,199 pupils in 848 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, spatial 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 Vittadini G. (2012), "Regional Economic Disparities as Determinants of Students' Achievement in Italy", *Research in Applied Economics*, Vol. 4, Issue 1, pp. 33-53.

Aina, C. and Pastore, F. (2012), "Delayed Graduation and Overeducation: A Test of the Human Capital Model versus the Screening Hypothesis," IZA Discussion Papers 6413, Institute for the Study of Labor (IZA).

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.

Caroleo F. E. and F. Pastore (2012), “Overeducation at a glance. The wage penalty of the educational mismatch in the Italian graduate labour market, looking at the AlmaLaurea data”, Paper presented at the XXVII AIEL Conference, Seconda Università di Napoli - 27-28 September.

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.

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

Checchi, D. and Flabbi, L. (2007), “Intergenerational mobility and schooling decisions in Germany and Italy: the impact of secondary school tracks”, IZA Discussion Paper n.2876

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

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.

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.

Giuliano, P. (2008), "Culture and the Family: An Application to Educational Choices in Italy," *Rivista di Politica Economica*, Vol. 98, Issue 4, pp. 3-38, July-Augu.

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.

IRPET (2013), *La condizione giovanile ai tempi della crisi.*

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.

Mocetti S. (2007), “Scelte post-obbligo e dispersione scolastica nella scuola secondaria”, paper presented at the XIX SIEP Conferences, Pavia, 13 - 14 September.

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

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.

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.