Giornale degli Economisti e Annali di Economia
Volume 66 - N. 3 (Novembre 2007) pp. 299-333

GEOGRAPHICAL DIFFERENCES IN ITALIAN STUDENTS' MATHEMATICAL COMPETENCIES: EVIDENCE FROM PISA 2003[1]

MASSIMILIANO BRATTI,[2] DANIELE CHECCHI[3] AND ANTONIO FILIPPIN[4]

Received: October 2007; accepted: September 2007

In this paper we investigate the existence and the size of geographical differences in Italian students' mathematical competencies. We analyze a novel data set that combines the 2003 wave of the OECD *Programme for International Student Assessment* (PISA) with information about local economic conditions and school-level administrative data. We find there is significant positive correlation, across provinces, between mathematical literacy and school buildings maintenance and local employment probabilities. About 75% of the North-South differential in mathematical literacy is accounted for by resource differences, while geographical differences in school production functions account for the remaining fraction.

JEL Classification: J21, J24, H52.
Keywords: education, PISA, students, territorial differences.

1. INTRODUCTION

Italy typically scores poorly in international surveys on students' competencies such as the *Programme for International Student Assessment*

[1] This research project has been promoted by the *Fondazione per la Scuola della Compagnia di San Paolo*. We thank the *Ministero della Pubblica Istruzione* (MPI) for providing us some of the data used in this paper. We thank participants at the workshops held at the *Fondazione Collegio Carlo Alberto* (Turin, May 2006 and May 2007), the *University of Milan* (Milan, June 2006), the European University Institute (Florence, June 2007), the Bank of Italy (Rome, July 2007), and the conferences 'Economics of Education: Major Contributions and Future Directions' (Dijon, June 2006), 'V Brucchi Luchino conference' (Padua, November 2006) and the 'XIX International Conference of *Osservatorio Giordano dell'Amore* on Educational Systems and Human Capital' (Milan, June 2007) for useful suggestions. We thank in particular Giuseppe Bertola, Lawrence Kahn, Carmine Porello, Alfonso Rosolia, Paolo Sestito and an anonymous referee. The authors only are responsible for the content of this paper.

[2] Corresponding author: DEAS, Università degli Studi di Milano, via Conservatorio 7, I-20122, Milan, Italy and IZA (Bonn). Telephone: 02 503 21545. Fax: 02 503 21505 E-mail: massimiliano.bratti@unimi.it

[3] DEAS, Università degli Studi di Milano and IZA (Bonn). E-mail: daniele.checchi@unimi.it

[4] DEAS, Università degli Studi di Milano and IZA (Bonn). E-mail: antonio.filippin@unimi.it

(PISA), the *Trends in International Mathematics and Science Study* (TIMSS) and the *Progress in International Reading Literacy Study* (PIRLS).[5] As in other Mediterranean countries, this could be related to the relatively recent improvements in the level of education of the population. At the end of WWII Italy was a developing country, with more than half of the labour force employed in agriculture and a similar fraction of illiterate population. Sixty years later, the same country sits in the group of developed countries, but still lags behind in terms of educational attainment. This is still visible in the average educational attainment of the population: in 2003 the fraction of secondary school graduates was 22% in the population cohort aged 55-64 and 57% in the younger cohort aged 25-34; the corresponding figures for the OECD area were 49% and 64%.[6] For this reason, one would be tempted to attribute the poor performance of 15-year-old Italian students to the lack of an adequate cultural environment: if these students live in culturally deprived families, they do not receive sufficient support and incentives to achieve good results at school.

However, matters are probably more complex. A simple OLS regression of the level of PISA test scores in mathematical literacy on macro-area dummies (North-East, North-West, Centre, South and Islands) yields statistically significant differences across Italy. Indeed, the difference between South and Islands and the reference category (North-East) amounts to about –83 points, almost one standard deviation, while students in Centre Italy have on average a score which is 39 points lower than those in the North-East. North-West does not perform differently from the reference category [see column (1) in Table 1]. The geographical divide in mathematical competencies only partially reflects in grades achieved in the same subject, which are also lower in the South [column (2) in Table 1]. This is possibly due to different grading policies in different areas of the country: the (unconditional) correlation between mathematical competencies (average of the five plausible values) and marks obtained in the last transcript in mathematics is 0.42 in the North-West, 0.36 in the North-East, 0.35 in the Centre and 0.36 in the

[5] See for example OECD (2004) where Italy is among the countries with the highest improvements in numeracy performance between 2000 and 2003, but still remaining in the lowest quartile of countries. This is mainly due to a larger fraction of students in the two lowest levels of competencies. For a review of the performance of Italian students in international surveys see Montanaro (2007) and Cipollone - Sestito (2007).

[6] A similar dynamic is observed for the fraction of college graduates: they were 6% in the oldest cohort and 12% in the youngest (OECD 2003), while in the OECD area they were 10% and 18%, respectively.

[7] The interested reader is referred to Cipollone - Sestito (2007) for a thorough exploration of this hypothesis.

South of Italy.[7] Students seem partially aware of this situation when we look at self-perceptions: students from Centre and South schools are less self-confident and express greater anxiety in mathematics.[8] Even so, they exhibit a higher level of self-concept, which necessarily entails a lower correlation with both competencies and marks.

We do not find evidence of different aspirations, neither in terms of final educational attainment [column (6) of Table 1] nor in terms of occupational prestige associated to the desired occupations, expressed by 15-year-old students [column (7) of Table 1]. Finally, students in the South have higher opinions about schooling: the final four columns of Table 1 report the level of agreement with general statements about the school experience.[9] We note that, despite their lower level of competencies, which is partially signalled by the grading policies of their teachers and partially reflected by self-perceptions, students in Southern schools have higher opinions about the effectiveness of schools with respect to future working life. Differently from their peers in Northern schools, they maintain that schools are helpful, give them more confidence and provide skills for work.

We therefore face a situation where students in similar learning environments have different perceptions and significantly different outcomes. Such large territorial differences that characterize student performance are particularly striking given the highly centralised nature of the Italian educational system. School teachers are hired through a national competition, they are employed by the Ministry of Public Education (MPI, hereafter) and receive an identical pay, which evolves according to seniority only. Eighty per

---

[8] The PISA 2003 index of mathematics self-efficacy [MATHEFF, in column (3) of Table 1] is derived from students' responses to the eight items measuring the students' confidence with mathematical tasks (like using a train timetable or understanding a graph in a newspaper); positive values indicate higher levels of self-efficacy. The index of mathematics anxiety [ANXMAT, in column (4) of Table 1] is concerned with feelings of helplessness and emotional stress when dealing with mathematics, is derived from students' responses to questions like "I get very nervous doing mathematics problems" or "I worry that I will get poor marks in mathematics"; positive values indicate higher levels of anxiety. Finally the index of mathematics self-concept [MATSC, in column (5) of Table 1] is derived from students' responses to items like "I am just not good at mathematics" or "I have always believed that mathematics is one of my best subjects"; positive values on this index indicate a positive self-concept in mathematics.

[9] "School has done little to prepare me for adult life when I leave school" [variable ST24Q01 in column (8) of Table 1]. "School has been a waste of time" [variable ST24Q02 in column (9) of Table 1]. "School has helped give me confidence to make decisions" [variable ST24Q03 in column (10) of Table 1]. "School has taught me things which could be useful in a job" [variable ST24Q04 in column (11) of Table 1]. Variables have been recoded such that positive values indicate agreement with the statement.

TABLE 1 – *Territorial differences in students' mathematical literacy, aspirations (education and expected occupations), self-perception and opinion about school (PISA 2003) in Italy*

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)
	school performance		self perceptions			aspirations		opinions about schooling			
VARIABLE	mathemat. test score	past mark in mathem.	math. self-efficacy	anxiety in math.	math. self-concept	educational attainment	occupational prestige	school did little for adult life	school is a waste of time	school gave confidence make decisions	school taught useful for job
Descriptive statistics											
mean dep. variable	468.057	5.930	-0.098	0.283	0.006	4.011	58.993	-0.201	-0.857	0.315	0.794
st. deviation dep. var	93.663	1.499	0.822	0.848	1.010	1.119	17.542	0.733	0.655	0.762	0.741
min dep.var.	111.185	1.000	-3.890	-2.478	-2.122	1 (prim)	16.000	-1.5 disagr	-1.5 disagr	-1.5 disagr	-1.5 disagr
max dep.var.	841.439	10.000	2.531	2.697	2.416	5 (univ)	90.000	1.5 agree	1.5 agree	1.5 agree	1.5 agree
Estimation technique	OLS	OLS	OLS	OLS	OLS	OLS	OLS	OR.PROBIT	OR.PROBIT	OR.PROBIT	OR.PROBIT
Area											
North-West	0.492	-0.05	-0.069	0.031	-0.051	0.1	1.418	0.025	0.019	0.019	-0.081
	[0.05]	[0.58]	[0.99]	[0.87]	[0.96]	[0.85]	[0.91]	[0.50]	[0.28]	[0.41]	[2.60]***
Centre	-39.075	-0.251	-0.129	0.087	-0.033	0.143	0.007	-0.005	-0.07	0.086	0.084
	[3.95]***	[2.08]**	[1.97]*	[1.71]*	[0.45]	[0.89]	[0.00]	[0.08]	[0.90]	[1.33]	[1.80]*
South and Islands	-82.797	-0.303	-0.123	0.168	0.151	0.201	3.055	-0.061	-0.201	0.316	0.169
	[8.47]***	[2.55]**	[1.84]*	[3.91]***	[2.66]***	[1.63]	[1.66]	[0.98]	[2.30]**	[4.84]***	[4.06]***
Observations	11,565	11,318	11,488	11,474	11,475	11,456	9,364	11,418	11,414	11,412	11,472
R^2	0.16	0.01	0.00	0.01	0.01	0.00	0.01	0.00	0.00	0.01	0.00
Log likelihood	-67,936.66	-20,596.92	-14,036.04	-14,345.17	-16,349.48	-17,522.16	-40,082.91	-12,401.23	-10,328.94	-12,319.92	-11,555.42

Note: The first column reports t-statistics computed using 80 balanced repeated replications and 5 plausible values. The reference group is North-East. Weighed estimates. Errors are clustered by provinces.

cent of the teaching curricula are set by the central government, while the remaining is left to the autonomous design of each school. Two thirds of total financial resources are centrally set, while the remaining fraction is left upon local authorities (which are in charge of providing buildings and basic services such as transport, food and sport infrastructures). It is therefore rather surprising to observe this huge variation across regions, and this begs the question of what determines students' competencies.

In this paper we address the 'puzzle' of territorial variations of student performance[10] resorting to different sources of data. Information on the students' families was provided by the students themselves immediately after sitting the PISA test. The school head provided information about the teachers' behaviours and attitudes. In both cases information is potentially plagued by reporting errors (especially in the case of students reporting information on their parents) and by subjective perceptions (especially in the case of school managers). Thus, we have complemented original data with data from other sources (administrative data from the archives of MPI, data on geographical distribution of social phenomena such as immigration, unemployment, illegal activity, suicides from Census data or from *ad hoc* surveys) in order to capture additional determinants of student performance that may be related to resource and social capital locally available to students.

Our view is that student competencies as measured in PISA are likely to be significantly affected by the surrounding socio-economic environment. Indeed, as we will stress in Section 3, PISA tests are not intended to measure curricular competencies, but rather specific forms of literacy and numeracy which are formed through interactions with external factors. In this regard OECD (2004, p. 23) states: "the acquisition of literacy is a lifelong process taking place not just at school or through formal learning, but also through interactions with peers, colleagues and wider communities".

This paper has descriptive aims, since it estimates statistical associations without giving them any causal interpretation. The results are nonetheless interesting for several reasons. Firstly, to the best of our knowledge, this is the first paper to explore a territorial dimension going beyond the simple inclusion of macro-area dummies in the analysis of PISA data in Italy. Secondly, although correlation does not necessarily imply causation, the latter cannot be excluded *a priori*. Therefore, we believe to have been able to identify territorial variables whose importance should be further investigated in order to estimate possible causal effects.

[10] When it is not stated otherwise, we refer to "student performance" meaning PISA test scores in mathematical literacy.

After reviewing the scant literature existing on Italian data (Section 2), we describe our data-set (Section 3) and our empirical strategy (Section 4). Our main results are presented in Sections 5 and 6, while Section 7 reports some concluding remarks.

2. The literature on the argument

Montanaro (2007) shows that the North-South divide characterizes the scores of all the international surveys conducted so far. Such gap is larger in maths than in reading and widens along the school career.

Using the initial PISA survey conducted in 2000, Checchi (2004) puts forward the existence of regional disparities in student performance in Italy, even after controlling for the type of secondary school attended.[11] He investigates the determinants of secondary school track and finds that parental education and socio-economic status are the main drivers of track choice. The same exercise has been repeated on the 2003 survey, where additional information on pre-existing ability of students was collected, without finding significantly novel results. Contrary to the German case, students in Italy are streamed in different tracks according more to their background than according to their ability (Checchi - Flabbi, 2006). Checchi (2004) also analyses the role of family background and school level peer effects in affecting students' performances in PISA test scores. His main result is that average parental education and socio-economic prestige measured at school level are much stronger predictors than individual variables, thus indirectly confirming that environmental factors may be important determinants of student performance. Indeed, even when controlling for individual background and school types, the significance of regional dummies (in the order of 50 test score points) indicates that additional territorial variables unrelated to family factors are likely to be correlated with student competencies.

[11] It should be reminded that the Italian upper secondary school system can be described as tripartite, with an academic oriented generalist education provided by high schools (5 years, called *licei*, with further division in humanities, sciences, languages, pedagogy), a technically oriented education provided by technical schools (5 years, called *istituti tecnici*, with further differentiations according to the type of job), and a vocational training offered by local schools organized at regional level (3 years, called *istituti di formazione professionale*). After a debated reform in 1969, students from any track are entitled to enrol in Colleges and Universities, conditional on having successfully completed 5 years of upper secondary schooling (even students from vocational schools can enrol if they attend two integrative years). However, each of these tracks still predicts very different outcomes in terms of additional education acquired and labour market performance. More than 88% of students who graduate from *licei* enrol in a University as opposed to 17.8% of the students coming from the vocational track.

Using the PISA 2000 survey, Tramonte (2004) applies multilevel statistical modelling decomposing the total variance of student literacy tests into a within-school component (explaining 45% of the variance), a between-school within-region component (explaining 47% of the variance) and a residual 8% between-region component. However, the 5 macro regions she considers[12] (North-East, North-West, Centre, South, South and Isles) are very aggregated and heterogeneous. Using a multilevel model controlling for individual characteristics and the average characteristics of students in the same school, the author explains 19% of within-school variance, 88% of between-school variance and 58% of between-region variance. Hence, although the model is quite successful in explaining between-school variance, a consistent part of both individual and regional differences remains unexplained.

We think that these residual differences may relate to the local availability of social capital or other local resources and we aim to test this hypothesis using PISA 2003 data.

3. Data

PISA data have by now become very popular not only among researchers in several disciplines, such as economists, educationalists, political scientists and sociologists, but they have also drawn the attention of the media and the general public. For this reason we report here only their main characteristics.[13] The purpose of PISA is to gather highly standardised data that can be used to compare student competencies in various domains both within and between countries.

We use in this paper the second wave of PISA, which refers to data collected in Spring of 2003 and whose main focus is on measuring performance in mathematics. As emphasised in OECD (2004), PISA considers students' competencies in some areas "not in isolation but in relation to students' ability to reflect on their knowledge and experience and to apply them to real world issues" (p. 24). Therefore, unlike other large-scale surveys such as TIMSS or PIRLS, PISA does not focus on curricular competencies but on knowledge and skills that can be used in every day life, helping the individual to fully realise his/her potential in the 'knowledge society'.

PISA data gather a wealth of information on both students' and schools' characteristics. The latter are collected through a questionnaire filled in by

[12] Since only this level of information was available in PISA 2000.

[13] A detailed description of the general characteristics of the survey can be found in OECD (2004), while for more technical details the interested reader is referred to OECD (2005a).

the head of each school that entered PISA national samples. For our purpose of analysing geographical differences, the main drawback of the PISA survey in Italy is that its sample only contains very aggregated information of school geographical location (11 areas), which does not enable an analysis of the role of territorial factors.[14] However, thanks to a research effort of the Ministry of Public Education, and the Italian agency for the assessment of the educational system (INVALSI), the original data set has been matched at school level with administrative data and at province level with data from other statistical sources.[15]

The procedure to build the data set worked as follows. Several variables were collected at the level of *Provincia* (province) from the 2001 Population Census, the 2002 Italian Labour Force Survey, and many other surveys run by the Italian National Statistical Institute (ISTAT), including cultural and judiciary statistics. *Province* (comparable to counties in the Anglo-Saxon context) are the intermediate level in which the Italian territory is organized by the Italian Constitutional Law, the upper level being *Regioni* and the lower *Comuni*. Currently there are 110 *province* in Italy, for a population of 59 millions. This province-based data set was sent to MPI and merged with the Italian PISA data set, by province.[16] MPI also merged the PISA data set with information on students and schools collected through the *Sistema Integrato Segreterie Scolastiche Italiane* (SISSI) information system. The latter mainly includes information on students and limited information on teachers (teachers with permanent contracts, teachers with temporary contracts, type of teacher's qualification, teachers' age and gender). After removing the school identifiers, MPI finally returned the merged set of anonymous data to us.

It must be noted that the explicit stratification variables used in the Italian sample design (see OECD 2005a, p. 54) were aggregate geographical regions (11), school types (4) and school size (2), while only one implicit stratification variable was used (public/private). Therefore, the sample is likely not to be representative of the Italian population at province level, the territorial unit at which we measure local variables. However, we think this is a

[14] The official data set released by the OECD does not even allow an analysis by macro-region, e.g. North, Center and South, since some Central and Southern regions are aggregated in a common macro-area.

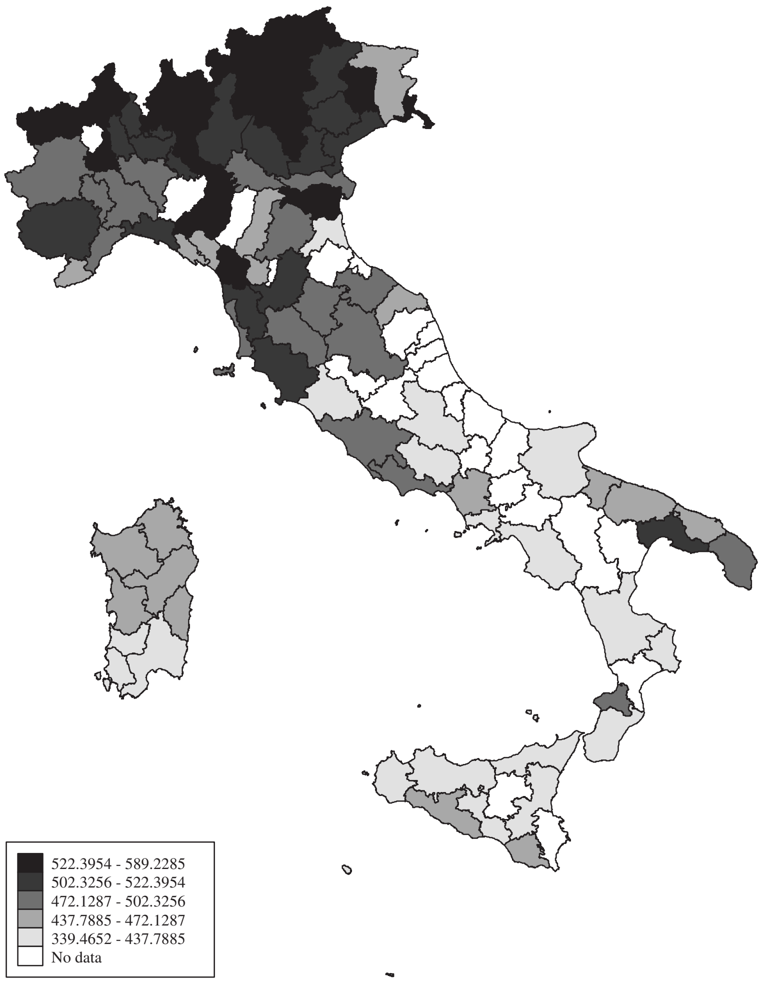
[15] We are very grateful to Aura Micali (formerly director of MPI), Prof. Bruno Losito (Università degli Studi Roma Tre) and Prof. Giacomo Elias (Invalsi) for supporting our research effort.

[16] 7 out of 110 *province* have been created after 2003. Therefore we assign them to the provinces to which they belonged when PISA data were collected. In particular: Barletta-Andria-Trani to Bari; Carbonia-Iglesias and Medio Campidano to Cagliari; Fermo to Ascoli Piceno; Monza-Brianza to Milan; Ogliastra to Nuoro; Olbia-Tempio to Sassari.

minor problem in our case, since our goal is not to assess differences in student performance by province, but only to assess which local variables are correlated with student performance, and whether once we control for individual, school and local variables the North-South divide reduces or disappears.

With this cautionary note in mind, Figure 1 reports the quintiles of the average students' mathematical literacy by province. The map shows only

FIGURE 1 - *Quintiles of the distribution of raw test scores – mathematical literacy in PISA 2003 by province*



Note: PISA scores have an average of 500 and a standard deviation of 100 in the cross-country OECD data set. In the PISA 2003 Italian sample the following provinces are not represented: Ascoli Piceno, Avellino, Benevento, Biella, Campobasso, Catanzaro, Chieti, Enna, Forlì, Isernia, Macerata, Matera, Pescara, Piacenza, Potenza, Prato, Reggio Emilia, Rieti, Rimini, Siracusa, Teramo, Terni.

unconditional differences (i.e. unadjusted for contextual factors such as family background characteristics or school type) of the student population sampled in each province. It suggests that students in Northern and Centre Italy perform significantly better than those in the South (as we already saw in Table 1). Given the lack of representativeness of the Italian sample by province, the averages should not be strictly interpreted as the performances of the student populations in each province, and the map is only meant to give a broad picture of territorial differences in students' mathematical literacy. [17] After controlling for contextual factors, such as individual, school and local variables, we will able to assess how this overall picture changes.

4. Econometric strategy

The analytical tool we use in our empirical analysis is the estimation of educational production functions (EPFs), aimed at assessing the size and the determinants of territorial differences in student performance. In order to achieve such a goal we follow a 2-step strategy.

1. Step 1: we adopt an 'incremental' strategy and estimate alternative specifications where we progressively add new controls. In particular, we start from a specification including only individual variables, then we add information at school level, and finally variables related to the local socio-economic environment. The aim of this step is to identify a set of local variables correlated to student performance. Models estimated at this stage do not include macro-area dummies;
2. Step 2: we estimate the same models as in the previous step, but also including macro-area dummies. This enables us to check whether the area dummies are likely to capture other unobserved effects that are common within macro-areas. If this does not happen, we are able to account for the North-South divide using only observable information.[18]

It is important to note that our paper has a descriptive character. Indeed, our strategy is to investigate the statistical significance of the correlations

[17] It is important to note that some of the richest provinces of Southern and Centre Italy are not represented in PISA 2003 and this may contribute to overestimate the gap between Northern Italy and the rest of the country. However, this possible bias is counterbalanced by the fact that the worse students are more likely to drop out, and therefore not showing up in PISA, in the South (see Cipollone - Sestito 2007).

[18] By including in the models macro-area dummies (which capture the effect of other unobservables common to provinces located in the same macro-geographic area), in this second stage we are implicitly exploiting only the within-area variation in the provincial variables.

between some individual, school and local variables and student performance, without giving a causal interpretation to the estimated coefficients. Our analysis is nonetheless important to isolate such factors whose statistical correlation with student performance may hide 'causal effects' and on which future research should focus.

We assume a linear EPF. In this case, the relation between student performance in mathematics (y_{ijp}) and the explanatory variables including vectors of individual ($\mathbf{x}_i$), school ($\mathbf{s}_j$) and local characteristics ($\mathbf{q}_p$), can be written as:

$$y_{ijp} = \mathbf{x}'_i\boldsymbol{\beta} + \mathbf{s}'_j\boldsymbol{\gamma} + \mathbf{q}'_p\boldsymbol{\eta} + \varepsilon_{ijp} \qquad (1)$$

where i =1, …, N, j =1, …, J and p =1, …, P are the subscripts for individuals, schools and provinces, respectively. ε_{ijp} is a random term capturing unobserved factors. $\boldsymbol{\beta}$, $\boldsymbol{\gamma}$ and $\boldsymbol{\eta}$ are the vectors of coefficients on individual, school and province variables, respectively. Initial specifications exclude $\mathbf{s}_j$ and/or $\mathbf{q}_p$.

In order to take into account the complex survey design of PISA (two-stage stratified sample) when estimating model (1) it is necessary to use the balanced repeated replications (BRRs) weights provided in the data set (see OECD, 2005b, pp. 31-52). Moreover, as PISA does not provide a point estimate for student performance but estimates a distribution of scores from which five values are drawn for each individual (*plausible values*, PVs), it is necessary to correct the standard errors of the estimates for the fact that PISA scores are imputed to individuals (OECD, 2005b, pp. 71-80). For this reason all the estimates in this paper use the 'unbiased shortcut' described in OECD (2005b, p. 109) to obtain unbiased standard errors.

As usual, in order to obtain unbiased estimates of $\boldsymbol{\beta}$, $\boldsymbol{\gamma}$ and $\boldsymbol{\eta}$ the error term ε_{ijp} must be uncorrelated with the explanatory variables included in the right-hand side (RHS, hereafter) of equation (1), i.e. we must not have omitted any relevant variable that is correlated with those included in the regression. It is important to note that since PISA is a cross-sectional data set, which does not provide information on past inputs, we are only able to estimate a 'contemporaneous specification' in the language of Todd - Wolpin (2003). In such a specification the identification of causal effects is difficult and requires some 'heroic' assumptions.[19] Therefore, the coefficients we es-

[19] These assumptions are (Todd - Wolpin, 2003, p. F16):
"(i) Only contemporaneous inputs matter to the production of current achievement.
Or: (ii) Inputs are unchanging over time, so that current input measures capture the entire history of inputs.
In addition to (i) or (ii):
(iii) Contemporaneous inputs are unrelated to (unobserved) endowed mental capacity".

timate must be interpreted as 'robust correlations', in the sense that they are net of the contribution of other factors potentially affecting mathematical competencies.[20]

However, one important variable is missing in PISA: student's innate ability. Since student ability may be correlated with other individual explanatory variables included in the model, such as parents' education, these are likely to partly capture the effect of ability and therefore to lead to biased estimates (*ability bias*). We think that this might be less of a problem for the analysis in this paper where we mainly focus on the relation between territorial variables and student performance. Indeed, given that we run an analysis of performance at the individual level while local variables are measured at province level, we think that the estimated coefficients are generally less prone to an *ability bias* or to other possible forms of *endogeneity bias*. Indeed, it is sufficient to assume an analogous distribution of innate ability within each province and to consider that single individuals or single families can hardly affect the local environment (such as employment rates) to get unbiased coefficients. A possible reason why *endogeneity* may still be a problem for territorial variables is migration, i.e. individuals may choose their preferred local environment. For instance, wealthier and more educated parents may choose to reside in provinces characterised by 'good' schools and better socio-economic environments. By controlling for several parents' characteristics (such as wealth, socio-economic status and education), we partially address this form of *endogeneity*.[21] However, as we already said, a full assessment of the 'causal effects' of local variables would require the use of other methods (such as instrumental variables methods or an experimental setting), which cannot be applied with the information available in our data set.

5. Results

In this section we report the results of the two stages of the empirical analysis, i.e. the estimation of linear EPFs excluding macro-area dummies and then including them.

[20] Our main objective is to be reasonably sure that the correlation we estimate reflects a true correlation with the territorial variable we have included in the regressions and does not reflect the correlation with other unobserved territorial variables.

[21] That is, migration decisions are mainly driven by the observables included in our econometric models.

5.1 The role of individual, school and local factors

All the results of the first stage are reported in Table 2. The different columns correspond to different models obtained by progressively adding control variables.[22] Sometimes when including local variables, we preferred to rely upon a preliminary factor analysis in order to address the presence of high correlation between these variables, which often raises mult-collinearity problems.[23] In what follows we describe our main findings.

Individual factors. For the choice of the individual variables $\mathbf{x}_i$ included in the initial model, we rely on the findings of the previous literature, in particular of the contributions using PISA,[24] but at the same time we mainly include factors that are presumably exogenous and for which a problem of reverse causality can be reasonably excluded.[25] We include several controls for students' economic and cultural family background, as well as the type of learning strategies in model (1). The main findings are as follows. Female students have a lower performance in maths, the difference with respect to male students amounting to –18 points in the model including only individual variables; the same difference changes only slightly in the models including further controls. The gender gap in mathematical performance turns out to be very robust and is a well-established result in the PISA literature.[26] In line with previous findings, we observe a positive correlation of student performance with household's economic capital, proxied by an index of home possessions (HOMEPOS) and by parents' highest occupational status (measured by the International Socio-Economic Index of Occupational Status – *ISEI* - see Ganzeboom - Treiman, 1996). Also cultural capital (proxied by the number of books at home, the availability of computer facilities at home and parental education) is positively associated with student per-

[22] Table A1 in the Appendix reports sample summary statistics. Note that we excluded 15-year-old students who were still in lower secondary schools, due to repeated failures (74 observations). In our econometric analysis we included individuals in grades 8, 9 and 10 without controlling for grade, since it is likely to be endogenous with respect to the level of mathematical competencies. In the final column of Table 2, we also estimated our preferred specification only on students attending the 9th grade, without finding significant differences.

[23] The results of this factor analysis are available upon request from the authors.

[24] For a survey of the empirical evidence on the theoretical determinants of student performance see chapter 1 of Bratti *et al.* (2007), while for a survey of the evidence from large international surveys (TIMSS, PISA, PIRLS, IALS) with a particular focus on Italy see chapter 2 of the same volume.

[25] Indeed, there are some factors such as mathematical anxiety and mathematics self-efficacy (ANXMAT and MATHEFF, respectively, in PISA 2003) that are highly correlated with performance in mathematics, but for which reverse causality is very likely.

[26] An opposite gender gap is found in reading literacy test scores.

TABLE 2 – *Models of students' mathematical literacy excluding macro-area dummies (PISA 2003)*

Variable	Model (1) individual	(2) + school	(3) + infrastructures	(4) + expenditures	(5) + labour mkt	(6) + cultural	(7) + demographic	(8) + social capital	(9) (7) + social cap. indx	(10) (9) only on 9th grade students
Small town (< 15,000)	11.116	2.947	-10.433	-7.191	-5.115	-6.335	-3.945	-4.863	-3.905	-5.979
	[1.32]	[0.34]	[1.40]	[0.93]	[0.76]	[0.97]	[0.61]	[0.76]	[0.61]	[0.87]
City (> 100,000)	2.436	-3.119	-13.966	-9.404	-3.564	-3.483	-2.387	-2.682	-2.197	-3.035
	[0.30]	[0.45]	[3.03]***	[2.31]**	[0.85]	[0.79]	[0.55]	[0.60]	[0.50]	[0.64]
Female	-18.254	-23.292	-22.234	-22.323	-22.541	-22.675	-22.172	-22.129	-22.045	-24.034
	[4.20]***	[7.53]***	[8.12]***	[8.09]***	[8.32]***	[8.48]***	[8.34]***	[8.30]***	[8.27]***	[8.02]***
Age	15.035	8.100	9.135	8.792	8.372	8.394	8.988	8.882	9.064	4.139
	[3.10]***	[1.79]*	[2.10]**	[1.99]*	[1.91]*	[1.91]*	[2.03]**	[2.00]**	[2.04]**	[0.90]
Single parent	-11.021	-6.636	-6.129	-6.180	-5.798	-5.853	-5.598	-5.620	-5.537	-2.368
	[3.60]***	[2.15]**	[2.10]**	[2.22]**	[2.14]**	[2.16]**	[2.07]**	[2.07]**	[2.04]**	[0.75]
Highest parental occupational status	0.786	0.295	0.311	0.335	0.335	0.334	0.335	0.335	0.336	0.216
	[7.56]***	[3.94]***	[4.16]***	[4.63]***	[4.62]***	[4.59]***	[4.62]***	[4.61]***	[4.63]***	[2.86]***
Highest parental education in years of schooling	17.340	10.095	8.051	7.995	7.499	7.427	7.134	7.111	7.188	7.535
	[8.64]***	[5.77]***	[4.78]***	[4.49]***	[4.07]***	[4.08]***	[3.97]***	[3.95]***	[3.99]***	[3.48]***
Highest parental education in years of schooling squared	-0.733	-0.499	-0.405	-0.407	-0.385	-0.382	-0.369	-0.368	-0.371	-0.387
	[8.33]***	[6.64]***	[5.66]***	[5.36]***	[4.93]***	[4.93]***	[4.82]***	[4.79]***	[4.83]***	[4.15]***
Computer facilities at home (*comphome*)	12.051	9.564	7.478	7.800	7.890	7.625	7.384	7.401	7.335	7.045
	[6.77]***	[5.67]***	[4.70]***	[5.16]***	[5.38]***	[5.31]***	[5.27]***	[5.26]***	[5.26]***	[4.35]***
No. books at home	0.056	0.026	0.027	0.027	0.027	0.027	0.027	0.027	0.027	0.024
	[7.39]***	[3.77]***	[4.03]***	[4.08]***	[4.02]***	[4.00]***	[4.03]***	[4.04]***	[4.04]***	[3.73]***

Index of home	14.702	3.704	3.099	2.376	2.207	2.458	2.481	2.456	2.486	2.549
possessions (*homepos*)	[6.61]***	[1.97]*	[1.61]	[1.39]	[1.31]	[1.47]	[1.50]	[1.48]	[1.50]	[1.54]
Memorisation strategies	-9.122	-8.129	-8.158	-8.342	-8.286	-8.165	-8.053	-8.043	-8.048	-9.138
(*memor*)	[5.01]***	[4.63]***	[4.89]***	[5.28]***	[5.36]***	[5.31]***	[5.21]***	[5.20]***	[5.22]***	[4.99]***
Elaboration strategies	0.905	3.638	5.229	5.234	5.747	5.611	5.505	5.559	5.507	6.102
(*elab*)	[0.63]	[2.45]**	[3.84]***	[3.86]***	[4.37]***	[4.23]***	[4.12]***	[4.14]***	[4.09]***	[4.10]***
Competitive learning	0.142	1.887	4.410	4.561	5.020	5.088	5.133	5.149	5.147	5.278
(*complrn*)	[0.08]	[1.18]	[2.90]***	[2.82]***	[3.23]***	[3.27]***	[3.34]***	[3.34]***	[3.34]***	[3.44]***
Co-operative learning	-6.285	-5.069	-4.297	-4.395	-4.232	-4.345	-4.318	-4.358	-4.286	-4.495
(*cooplrn*)	[4.92]***	[4.92]***	[4.06]***	[4.04]***	[3.81]***	[3.89]***	[3.85]***	[3.89]***	[3.81]***	[3.77]***
High school oriented towards humanities (*liceo classico*)		-27.064	-3.831	-0.547	4.825	5.116	4.510	5.952	5.982	6.266
		[2.14]**	[0.36]	[0.05]	[0.46]	[0.49]	[0.45]	[0.59]	[0.60]	[0.58]
High school oriented towards sciences (*liceo scientifico*)		2.939	27.627	30.942	39.864	40.324	39.936	39.589	40.511	40.272
		[0.28]	[2.81]***	[3.24]***	[4.21]***	[4.31]***	[4.33]***	[4.28]***	[4.38]***	[4.00]***
Technical school		25.469	25.104	25.190	20.419	20.068	19.081	18.716	19.488	22.188
(*istituto tecnico*)		[3.40]***	[3.80]***	[3.47]***	[3.07]***	[3.01]***	[2.90]***	[2.85]***	[2.95]***	[3.20]***
Vocational school		-6.467	-11.308	-14.158	-23.130	-22.925	-25.395	-25.283	-25.563	-24.947
(*istituti professionali*)		[0.69]	[1.49]	[1.63]	[2.82]***	[2.81]***	[3.27]***	[3.22]***	[3.26]***	[2.99]***
Private school		-37.848	-19.503	-19.645	-20.148	-21.148	-22.926	-22.813	-22.606	-23.374
		[3.26]***	[1.50]	[1.47]	[1.48]	[1.58]	[1.71]*	[1.70]*	[1.68]*	[1.57]
Proportion of computers		17.177	12.709	14.221	10.724	11.805	13.701	13.107	13.584	14.519
connected to internet		[2.15]**	[1.95]*	[2.21]**	[1.73]*	[1.90]*	[2.14]**	[2.03]**	[2.12]**	[2.22]**
Teacher/student ratio		0.498	0.337	0.328	0.565	0.505	0.489	0.499	0.469	0.444
		[1.50]	[0.99]	[0.92]	[1.23]	[1.13]	[1.05]	[1.06]	[1.00]	[0.95]
Highest parental education in yrs of schooling - school average		9.548	4.337	3.422	0.849	0.584	-0.447	-0.234	-0.297	-0.777
		[2.26]**	[1.54]	[1.17]	[0.32]	[0.21]	[0.16]	[0.08]	[0.10]	[0.27]
Highest parental occupational status – school average		0.619	0.927	1.130	0.830	0.845	1.146	1.163	1.122	1.256
		[0.69]	[1.40]	[1.78]*	[1.29]	[1.33]	[1.66]	[1.72]*	[1.62]	[1.74]*

(segue)

TABLE 2 – *(continua)*

Variable	(1) individual	(2) + school	(3) + infrastructures	(4) + expenditures	(5) + labour mkt	(6) + cultural	(7) + demographic	(8) + social capital	(9) (7) + social cap. indx	(10) (9) only on 9th grade students
No. books at home -		0.211	0.098	0.075	0.063	0.055	0.031	0.022	0.028	0.029
school average		[3.74]***	[2.21]**	[1.61]	[1.54]	[1.35]	[0.70]	[0.48]	[0.61]	[0.61]
Student-teacher relations		-5.700	-4.557	-4.676	-4.613	-4.708	-4.693	-4.682	-4.683	-3.989
at school (*sturel*)		[4.29]***	[3.89]***	[4.21]***	[4.35]***	[4.44]***	[4.49]***	[4.50]***	[4.46]***	[3.42]***
Disciplinary climate		6.825	7.428	7.585	7.921	7.870	7.796	7.820	7.812	6.740
in maths lessons (*disclim*)		[4.49]***	[5.83]***	[6.17]***	[6.83]***	[6.92]***	[6.69]***	[6.77]***	[6.68]***	[5.57]***
Fraction of students held		-71.115	-91.555	-86.850	-87.696	-91.657	-90.269	-93.439	-89.489	-70.197
back		[2.87]***	[3.97]***	[3.62]***	[4.17]***	[4.25]***	[4.19]***	[4.21]***	[4.16]***	[2.84]***
Temporary-permanent		1.185	-0.021	0.410	-0.082	-0.163	-0.373	-0.281	-0.256	-0.114
teacher ratio × 100		[3.98]***	[0.08]	[1.40]	[0.33]	[0.63]	[1.41]	[1.02]	[0.85]	[0.38]
% buildings unfit to school			-0.937	-0.891	-0.195	-0.180	-0.346	-0.446	-0.374	-0.526
and improperly adapted										
to schools			[3.48]***	[3.20]***	[0.79]	[0.72]	[1.30]	[1.68]*	[1.38]	[1.76]*
% school buildings			-0.877	-0.762	-0.482	-0.373	-0.277	-0.096	-0.211	-0.059
rented for schooling			[5.00]***	[3.96]***	[2.60]**	[2.03]**	[1.50]	[0.43]	[1.04]	[0.28]
Maintenance of buildings			-11.458	-9.499	-2.717	-3.936	-3.228	-2.270	-3.256	-3.462
– 1st factor			[5.74]***	[6.13]***	[1.42]	[2.03]**	[1.84]*	[1.34]	[1.90]*	[1.81]*
Log govrn. exp. on capital				16.748	11.603	11.268	10.696	10.561	11.006	10.251
account per student				[2.98]***	[2.21]**	[2.08]**	[2.00]**	[1.97]*	[2.02]**	[1.73]*
Log govern. exp. on				-31.993	-32.905	-25.810	-37.332	-40.261	-38.492	-40.590
intermediate inputs per student				[2.00]**	[2.96]***	[2.30]**	[3.33]***	[3.73]***	[3.40]***	[3.52]***

Log govern.exp. on teachers				73.339						
per student				[1.73]*						
Class size (no. students)				3.092						
				[0.98]						
Employment probability					1.943	1.444	1.389	1.491	1.196	1.332
(100-unempl.rate, at province level)					[4.92]***	[2.92]***	[2.48]**	[2.46]**	[2.00]**	[2.03]**
% irregular work					-0.650	-0.481	-0.632	-0.498	-0.651	-0.671
					[2.49]**	[1.83]*	[2.41]**	[1.82]*	[2.48]**	[2.54]**
N. crimes with unknown					-0.008	-0.010	-0.005	-0.003	-0.004	-0.004
author per 100,000 inhab.					[3.02]***	[3.27]***	[1.77]*	[1.03]	[1.26]	[1.03]
% without formal education						-3.192	-5.240	-4.630	-5.432	-6.024
(on pop > 6 years)						[1.91]*	[2.74]***	[2.26]**	[2.80]***	[2.99]***
% rented houses							-1.377	-1.214	-1.321	-1.547
							[2.92]***	[2.56]**	[2.81]***	[3.07]***
% foreign-born in resident							-4.096	-3.398	-2.976	-2.479
population (2001)							[1.99]*	[1.51]	[1.33]	[1.03]
social capital – 1st factor									1.391	1.760
									[0.85]	[0.96]
No. attempted suicides per								0.654		
per 100,000 inhab.								[1.29]		
No. volunteers in non-profit								1.190		
organisations per 100 inhab								[1.19]		
Observations	10,894	8,468	8,468	8,410	8,410	8,410	8,410	8,410	8,410	7,045
R^2	0.20	0.40	0.45	0.46	0.48	0.48	0.48	0.48	0.48	0.48
Log likelihood	-63,590.08	-48,200.71	-47,796.54	-47,430.51	-47,295.95	-47,284.43	-47,260.50	-47,254.96	-47,259.28	-39,459.34

Note. Absolute value of t-statistics in brackets. t-statistics are computed using 80 balanced repeated replications and 5 plausible values. The reference individual is male, lives in a town (15,000-100,000), comes from an intact family and is enrolled in *other schools* in the public education sector (see Section 5.1). The number of observations may differ across columns due to missing data. Errors are clustered by province. * significant at 10%; ** significant at 5%; *** significant at 1%.

formance, although the last variable exhibits a non-linear relation. The use of elaboration learning strategies and competitive learning have a positive and significant correlation with performance, while memorisation learning strategies and cooperative learning exhibit negative signs. Some of the estimated coefficients change in size across specifications but they are generally statistically significant (except for the index of home possessions).

School factors.[27] From model (2) we introduce school factors. The type of upper secondary school is highly statistically significant in most specifications. The reference group are *other schools* (including language, art and pedagogical schools, namely *Liceo Linguistico, Istituto/Scuola Magistrale, Istituto d'Arte, Liceo Artistico*). Thanks to MPI we obtained an indicator of the type of school program that is more detailed than that commonly available through PISA.[28] The type of school attended comes out to be a very important factor since even the most academic track (*liceo*) is characterised by large within-differences: while students in scientific high schools generally perform better than the reference group, students in humanities high schools are not statistically different from the reference case. Students attending technical schools perform better than those in *other schools*, while the reverse occurs for students in vocational tracks. The advantage of scientific high schools does not necessarily reflect a causal effect, since it might capture self-selection of mathematically oriented students in this type of schools. Students in privately managed schools have worse performance than students in public schools, and this can be explained in terms of their prevalent feature of remedial schools in Italy (see Brunello - Checchi, 2004). Better student-teacher relations are associated with a lower performance, while a tougher disciplinary climate correlates positively with performance, another well know result from the PISA literature. The percentage of computers connected to Internet is positively associated with performance. Parents' education and the number of books at home, averaged at school level to proxy for the quality of peer group, turn out to be significant only in mod-

[27] In our analysis, we exclude many school variables provided by PISA, while preferring the inclusion of administrative information provided by MPI, since we hold administrative data more reliable than opinions expressed by the local schools' heads. Regarding PISA 2000, for instance, Checchi (2004) observes that teacher heads in Southern Italy complained about the lack of personnel more than those in Northern regions, who conversely complained more about the lack of equipments. Needless to say, the territorial distribution of these resources pointed just to the opposite direction.

[28] For 7 schools (6 in the province of Bolzano and 1 in the province of Aosta, corresponding to 199 students) data on school type are not available from MPI. According to the information available in PISA (variable PROGN) they should consist of two high schools (*licei*) and three vocational schools (*scuole professionali*). However, given the uncertainty on their specific school types, we preferred to stick to MPI data and dropped these schools.

el (2) while they loose significance once we include territorial variables. To be noted the positive correlation with mathematical literacy of the temporary-permanent teacher ratio (measured at school level). This correlation is unexpected, since we believe that a tenured teacher provides continuity in teaching; in addition the teacher team has greater opportunities to coordinate their efforts in raising student learning. However, when we introduce further territorial controls, the point estimates become negative, but the coefficients do not resume to statistical significance.

Local school infrastructures. From model (3) we start including some characteristics of school infrastructures.[29] We remind the reader that provincial administrative units are responsible for maintenance of upper secondary school buildings. We use three variables relating to the state of schooling infrastructures. The first two, i.e. the share of buildings that are unfit for schooling and the share of rented buildings, can be considered as proxies for an insufficient endowment of infrastructures. The third variable (maintenance state) corresponds to the first factor extracted from six building's attributes,[30] and can be taken as a proxy of the poor state of maintenance of school buildings. All three variables are negatively correlated with school performance, although their statistical significance decreases when we include other local variables. However, the correlation between poor maintenance and student performance survives (at 10%) in model (9), our preferred specification, suggesting the possible existence of causal effects running from the state of school infrastructures towards performance that should be further investigated.

Local educational expenditures. From model (4) we introduce the (log) expenditure per student at province level provided by MPI articulated into expenditures for teachers, expenditures for intermediate consumptions and other personnel, and expenditures on capital account.

We observe a highly significant positive correlation between the (log) educational expenditure on capital account[31] and student performance. In contrast, our estimates show a negative and significant correlation between per-

[29] When we use these data, which are provided by MPI, we loose 2,500 observations relating to the provinces of Trento, Bolzano and Aosta corresponding to 69 schools.

[30] Roofing, flooring, water system, sewerage plant, electric system, heating-plant.

[31] These data refer to 1998 and 1999. The expenditure on capital account (mostly equipment) is expressed as percentage of total expenditures defined as: total budgetary expenditures (1998) augmented with the salaries of managing, teaching and ATA (administrative, technical and auxiliary) personnel (1999) divided by the number of pupils. These expenditures include only those borne by MPI, while those under the responsibility of other territorial administrative entities, such as provinces (e.g. expenditures on buildings for upper secondary schools) are excluded.

formance and the (log) expenditure in intermediate consumptions and other personnel.[32] Both correlations are quite robust across specifications. A higher expenditure on capital account can be a proxy for a higher quantity and/or quality of school infrastructures, while a higher expenditure on intermediate goods or other personnel may partly reflect inefficiencies. The most surprising result, although in line with the international literature, is the absence of any positive correlation with per-student expenditures for teachers (reported in model (4) and then removed due to insignificance).[33] This variable is likely to be affected by the student-teacher ratio and by the (average) teacher seniority (which in the Italian context corresponds to higher salaries), which may exhibit opposite correlations with students' performance.[34] Another possible explanation is that in the Italian context the student-teacher ratio varies across schools also due to the different number of support teachers (*insegnanti di sostegno*). Alternatively, if we consider average class size we do not find any significant correlation.[35] For this reason, and in order to reduce the dimension of the models, both (log)expenditure on teachers (per student) and class size were excluded from the following specifications.

Local labour market. From model (5) we add some variables describing local labour market conditions. The employment probability (defined as the complement to 100 of the unemployment rate) is highly correlated with individual student performance. An increase by one percentage point in the employment probability is associated with a more than one-point increase in the PISA score. On the contrary, an increase in the incidence of irregular work at province level is associated to a decrease in the PISA score. A similar negative correlation is found for the extension of the illegal sector (proxied by the number of crimes with unknown author per 100,000 inhabitants), which however disappears in the following models. It is clear that these variables may capture broader local socio-economic effects, which go beyond

[32] Which is obtained residually from total expenditures after substracting expenditures on capital account and for teachers.

[33] The expenditures for teachers are provided by MPI and are given by the teachers' salaries plus the *indennità integrativa*. The province of Udine did not provide these data.

[34] It must be noted that the same finding emerged when controlling for student-teacher ratios provided by school head teachers. When we control for teacher seniority at school level, we also did not find statistically significant coefficients (not reported here).

[35] A possible theoretical explanation for the absence of any effect of class size is offered in Lazear (2001), who emphasises the role of the class composition (smaller classes could be remedial for low ability students - thus implying negative correlation with performance – or 'magnet' for brilliant students – yielding a positive correlation). For two different points of view on the very rich literature on the effect of school resources see Hanushek (1997) and Greenwald - Hedges - Laine (1996).

what they directly measure.[36] However, since we control for family background and school factors, this correlation can be suggestive of a potential 'causal' effect. Individuals living in areas with a well functioning local labour market experience higher employment probabilities; by anticipating higher expected returns to education they are induced to invest more in education. In contrast, the larger the extension of the irregular sector, which mainly employs unskilled labour, and of the illegal economy (especially when criminals are less likely to be caught), the smaller the incentive for individuals to invest in human capital and the higher the incentive to devote time to alternative activities.[37]

Local cultural factors. In model (6) we make an attempt to introduce some proxies for the cultural environment surrounding schools. However, this attempt is unsuccessful. We tried to include variables related to the presence of libraries, university proximity, consumption of books, cinemas and TV programmes[38] which all turned out to be uncorrelated with student performance. We also tried to include the second factor extracted from this group of variables that, given the factor loading, mostly reflects the con-

[36] We did not include in this specification the (log) GDP per capita at province level given its high correlation with the employment rate.

[37] The variable we include is different from the employment rate (number of workers divided by the working age population) since it does not consider non participation. However, we consider what we have defined "employment probability", i.e. the likelihood of finding a job for individuals who are looking for a job, as a better proxy of the correlation we want to capture. The expected wage (w) for an individual conditional on her level of education (e) is $E(w|e) = \Pr(Empl|e)E(w|Empl, e)$, where *Empl* is a dichotomous variable denoting employment, $E(.)$ stands for mathematical expectation, and $\Pr(.)$ for probability. It is then clear that education may have a twofold positive effect, on the probability of employment (employment return) and on the wage conditional on employment (conditional wage return), that the unconditional (to employment) wage return can be written as $\frac{\partial E(w|e)}{\partial e} = \frac{\partial \Pr(Empl|e)}{\partial e} \cdot E(w|Empl,e) + \Pr(Empl|e) \cdot \frac{\partial E(w|Empl,e)}{\partial e}$. Therefore, the contribution of the conditional wage return is higher the higher the employment probability. However, in principle one could also observe a correlation of the variable 'employment probability' of the opposite sign if, for instance, the wage distribution is compressed and education has a low employment return (e.g., there are high employment opportunities irrespective of education). In the latter case individuals may exert a low effort in education since it has low economic returns. We have also explicitly included a proxy for the level of participation in the irregular sector in the econometric models.

[38] Proxied by the number of subscriptions to the national television broadcasting company (RAI) per 1,000 inhabitants. Since this subscription is compulsory by law for all individuals possessing a TV set, given that only one subscription is necessary per residential unit, this variable has lower values in the presence of larger families. We are aware that this variable may also be a proxy of tax evasion.

sumption of TV programmes, and we did not find any statistically significant correlation, and therefore all these variables do not appear in our estimated model. Vice versa, the fraction of population without formal schooling is negatively correlated with student performance. This evidence is a bit surprising, especially in the light of the fact that our model already controls for an individual's parental education and for the average parental education at school level.[39]

Local demographic and residential factors. From model (7) we add some information on demographic and residential characteristics of provinces. The first is the share of rented houses and the second one the fraction of foreign residents. The most robust correlation emerges for the first variable, which may approximate municipality size, complementing the categorical variables already included in the individual data section (reclassifying the categorical variable provided by PISA).[40] Indeed, the share of rented houses is 36% for Naples, 26% for Turin, 24% for Palermo, Milan and Rome, while the Italian average is 19%. If we accept this interpretation, then our findings suggest that students in very large cities have worse performances.

Local social capital. In the final models we add some proxies for social capital. Once again we experimented with several variables, among which only two seemed worth including in our regressions: the incidence of attempted suicides and the participation to non-profit activities. Both variables indirectly describe the quantity and quality of social interactions observed in the province territory. We expect lonely individuals being more inclined to attempt suicide, while the number of volunteers in non-profit organisations (per 100 inhabitants) captures the degree of solidarity within each province. Unfortunately both variables turn out being statistically insignificant [see model (8)]. We also tried other variables used in the literature as alternative proxies of social capital [such as participation to political elections, blood donations, etc. and the factor extracted using principal component analysis: see Micucci - Nuzzo, 2003) but all were not significant (see model (9)].[41]

---

[39] To have a rough idea of its impact, it is enough to say that by reducing this rate by one percentage point the average PISA student score would rise by 3-5 points depending on the particular estimated specification.

[40] In the current specification, municipalities are divided into small towns (less than 15,000 inhabitants), towns (15,000-100,000 inhabitants) and cities (more than 100,000 inhabitants).

[41] Including indicators of social capital in the educational performance function is also important because regional dummies may be proxying different levels of students' effort and motivation when sitting the PISA test. For instance, students from South Italy might be less responsive than those living in Northern Italy to the 'social returns' produced by

We also estimated the final model on the subsample of individuals attending the 9th grade only (i.e. who were regular in their previous educational career),[42] in order to control potential sample distortions introduced by different policies of grade retention adopted at school level. In this case, the coefficients are to be interpreted as the correlations with performance of individual, school and local factors conditional on the probability of not being held back at school in previous years. Model (10) shows that the correlations are both qualitatively and quantitatively very similar to model (9), the only notable exception is the negative correlation of being in a single parent family, which disappears when considering 9th graders only. This indirectly suggests that being in a non-intact family mostly raises the likelihood of being held back at school.

As to the performance of our econometric analysis in terms of explained variance, our most general model is able to account for about 50% of individual student performance. This value may not seem very high, however it is not lower than that one would have expected on the grounds of the empirical evidence that innate ability determines almost 50% of the variance in individuals' performances in IQ tests (see Plomin - Petrill, 1997).

5.2 Individual, school, local factors and the North-South divide

Table 3 reports the coefficients on the macro-area dummies obtained when re-estimating all the specifications in Table 2 (that excluded these dummies). The reference group are individuals living in North-Eastern Italy. Model (1) shows that individual variables are not sufficient to account for macro-regional differences. Indeed the difference between the North-East and the South (including also Islands) is almost 70 test points. Including school variables in model (2) improves the explanatory power of the model but does not help explain the North-South divide in students' mathematical peformance. Including information on school infrastructures in model (3) contributes to reducing the North-South gap by more than 20 points. The next large drop in the coefficient on the South dummy is produced by the

a good school performance in PISA. However, the fact that the proxies of social capital commonly used in the literature, such as blood donations and participation to elections, do not correlate with PISA performance and the good opinion of Southern students about the role of schools (see Section 1) are not consistent with this explanation. This does not exclude that students from Southern and Northern Italy may exert different levels of effort during the PISA tests for reasons uncorrelated with the proxies of social capital (i.e., they may have different incentives).

[42] In principle we cannot exclude that we are retaining in the sample students who enrolled primary school one year in advance, and later on had to repeat one year.

inclusion of the variables related to the local labour market. In this case the coefficient on the South dummy falls by 50%, becoming statistically insignificant at 5%. Inclusion of other control variables induces further loss of significance in the coefficient.[43] The coefficient on the Centre Italy dummy remains instead statistically significant also in the most general specification, altough the inclusion of territorial variables reduces its size by about 50%.

This can be illustrated by comparing figures 1 and 2. Figure 1 reported the distribution of raw test scores, providing a visual perception of the territorial divide between Northern and Southern regions. In Figure 2 we report the (quintiles of the) averages of the estimated residuals from model (9) in Table 2 at province level. These residuals show the distribution of the other half of variance that our model is unable to explain, which therefore can be attributed to individual unobserved heterogeneity (such as individual innate ability or motivation), to school unobserved characteristics (such as teachers' motivation) and in measurement errors related to the tests. It is evident that, once the contextual factors are controlled for, there are some areas of good performance both in the North and in the South of Italy, including several provinces of Southern Sicily and Sardinia.

[43] A possible concern with our strategy might be that the provincial variables which were progressively added might capture the role of geographical unobservables that are correlated with the macro-area dummies. If this were the case, including the variables we used might be just as good as including others which differ by macro-area, and the reduction of the significance of the coefficient on the macro dummies could be only a statistical artifact. As a robustness check we estimated two different specifications:

i) model (2) in Table 3 including maintenance of school buildings, expenditures on capital account and intermediate goods, the probability of employment and employment in the irregular sector, the percentage of illiterate population and the percentage of rented houses, which all turned out to be significant in model (9) of Table 2;

ii) specification (2) in Table 3 augmented by the percentage of unfit and rented school buildings, teacher expenditure, class size, the crime rate, the percentage of foreign population and the indicator of social capital, which were generally insignificant.

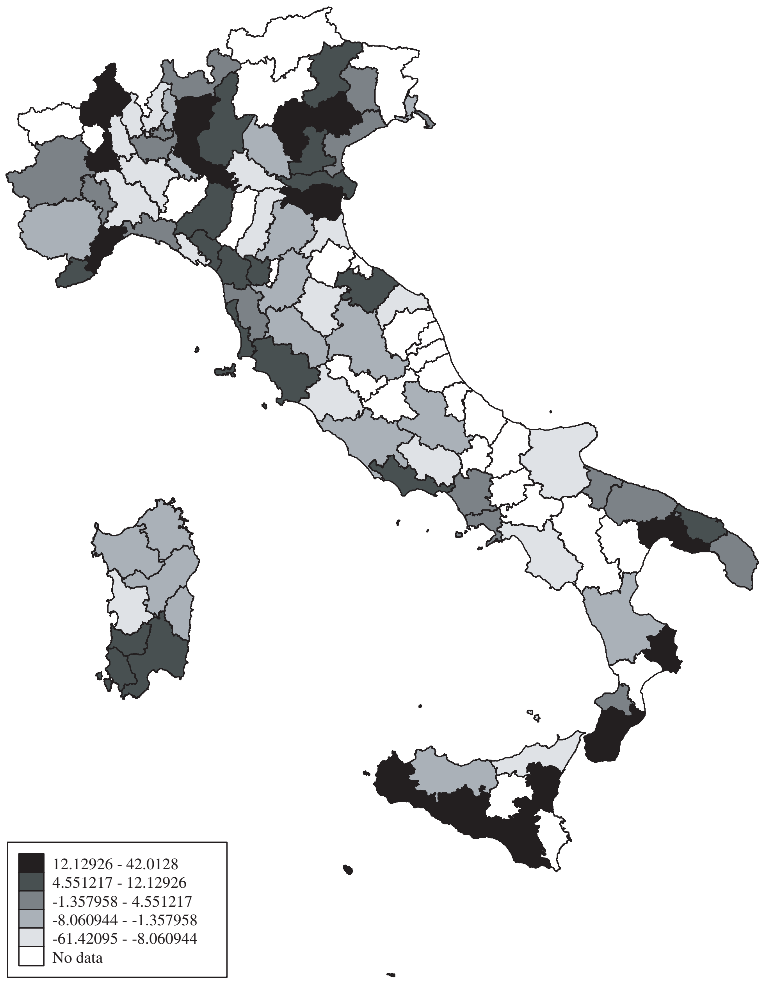
From the first specification we obtained the following coefficients and p-values (in parentheses): North-West –0.117 (0.98), Centre -20.457 (0.00), South and Islands -9.853 (0.50). The coefficients estimated from the second specification were instead: North-West 5.628 (0.37), Centre -30.577 (0.00), South and Islands -50.8 (0.00). This result, jointly with the fact that the coefficients on many provincial variables retained statistical significance while the area dummies generally lost significance and reduced in size, suggests that the significant provincial variables we included have a genuine correlation with student performance and do not simply reflect North-South differences in unobservables. Indeed, once we include the area dummies the coefficient on the provincial variables is estimated using only within macro-area variation in the same variables.

TABLE 3 – *Models of students' mathematical literacy including macro-area dummies (PISA 2003)*

	MODEL									
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Area	individual	+ school	+ infrastructures	+ expenditures	+ labour mkt	+ cultural	+ demographic	+ social capital	(7) + social cap. index	(9) only on 9th grade students
North-West	-0.308	-1.048	-0.233	2.383	4.636	3.801	2.835	2.603	3.048	4.793
	[0.04]	[0.22]	[0.04]	[0.40]	[0.67]	[0.57]	[0.45]	[0.42]	[0.51]	[0.73]
Center	-37.143	-32.419	-23.218	-21.257	-16.965	-16.486	-15.585	-14.845	-18.329	-17.257
	[5.04]***	[5.44]***	[4.00]***	[3.79]***	[2.66]***	[2.43]**	[2.46]**	[2.39]**	[2.69]***	[2.25]**
South and	-69.436	-68.277	-48.824	-44.721	-23.942	-22.464	-15.221	-13.188	-17.689	-16.455
Islands	[9.51]***	[11.63]***	[5.76]***	[4.78]***	[1.82]*	[1.61]	[1.05]	[0.90]	[1.22]	[0.99]
Observations	10,894	8,468	8,468	8,410	8,410	8,410	8,410	8,410	8,410	7,045
R-squared	0.30	0.46	0.47	0.47	0.48	0.48	0.48	0.48	0.48	0.48
Log likelihood	-62,871.91	-47,717.22	-47,662.18	-47,325.01	-47,263.01	-47,262.61	-47,244.29	-47,240.50	-47,238.99	-39,441.83

Note: This table reports the estimated coefficients on the macro-area dummies. Models (1)-(9) also include all explanatory variables listed in Models (1)-(9) in Table 2. The number of observations may differ across columns due to missing data. Errors are clustered by province. The reference individual is male, lives in North-Eastern Italy, in a town (15,000-100,000), comes from an intact family and he is enrolled in *other schools* in the public education sector (see Section 5.1).
Absolute value of t-statistics in brackets. t-statistics are computed using 80 balanced repeated replications and 5 plausible values. * significant at 10%; ** significant at 5%; *** significant at 1%.

FIGURE 2 - *Quintiles of the average estimated residuals (unexplained component of mathematical literacy) by province from model (9) in Table 2*



Note. Regression residuals are estimated as the difference between the mean of the five plausible values and the linear predictions obtained from model (9). In addition to the Italian provinces that were not sampled in PISA 2003 (see the note to Figure 1) residuals for other provinces are missing due to missing values in some of the explanatory variables included in model (9) (see Table 2).

6. ACCOUNTING FOR DIFFERENCES IN 'SCHOOL EFFECTIVENESS' ACROSS ITALY

In the previous section we have assumed the existence of a homogenous EPF across regions, while in this section we go one step further and partly relax this assumption. If we estimate equation (1) by macro-areas as

$$y_{ijk} = \alpha_k + \mathbf{x}'_i\boldsymbol{\beta}_k + \mathbf{s}'_j\boldsymbol{\gamma}_k + \mathbf{q}'_p\boldsymbol{\eta}_k + \varepsilon_{ijk}, \quad \mathrm{k} = N,C,S \qquad (2)$$

we can obtain the following decomposition

$$\bar{y}_N - \bar{y}_S = \underbrace{(\alpha_N - \alpha_S)}_{\substack{\textbf{unobservable}\\ \textbf{endowment}}} + \underbrace{(\bar{\mathbf{x}}'_N - \bar{\mathbf{x}}'_S)\boldsymbol{\beta}_N}_{\substack{\textbf{individual}\\ \textbf{endowment}}} + \underbrace{\bar{\mathbf{x}}'_S(\boldsymbol{\beta}_N - \boldsymbol{\beta}_S)}_{\substack{\textbf{individual}\\ \textbf{return}}} + \underbrace{(\bar{\mathbf{s}}'_N - \bar{\mathbf{s}}'_S)\boldsymbol{\gamma}_N}_{\substack{\textbf{school}\\ \textbf{endowment}}} + \underbrace{\bar{\mathbf{s}}'_S(\boldsymbol{\gamma}_N - \boldsymbol{\gamma}_S)}_{\substack{\textbf{school}\\ \textbf{effectiveness}}} + \underbrace{(\bar{\mathbf{q}}'_N - \bar{\mathbf{q}}'_S)\boldsymbol{\eta}_N}_{\substack{\textbf{local territory}\\ \textbf{endowment}\\ \textbf{effect}}} + \underbrace{\bar{\mathbf{q}}'_S(\boldsymbol{\eta}_N - \boldsymbol{\eta}_S)}_{\substack{\textbf{local}\\ \textbf{returns}}} \tag{3}$$

where a bar indicates sample averages. Equation (3) clearly shows that territorial differences depend on differences in endowments (at individual, school and territory levels) as well as on differences in the (implicit) returns to these endowments. We do not have good reasons to expect territorial differences in unobservables (like abilities, religious attitudes, political orientation, cultural differences, and so on), and therefore we impose $\alpha_N = \alpha_S$. We also do not find robust arguments to expect significant territorial differences in the returns of individual or local endowments, and therefore we also impose the identity of these implicit prices.[44] In addition, we are mostly interested in the effect of variables that in principle can be modified by the educational authorities (like school endowments and/or their impact on competencies formation). As a consequence, the territorial differences are decomposed according to the following equation

$$\bar{y}_N - \bar{y}_S = \underbrace{(\bar{\mathbf{x}}'_N - \bar{\mathbf{x}}'_S)\boldsymbol{\beta}_N}_{\substack{\textbf{individual}\\ \textbf{endowment}\\ \textbf{effect}}} + \underbrace{(\bar{\mathbf{s}}'_N - \bar{\mathbf{s}}'_S)\boldsymbol{\gamma}_N}_{\substack{\textbf{school}\\ \textbf{endowment}\\ \textbf{effect}}} + \underbrace{\bar{\mathbf{s}}'_S(\boldsymbol{\gamma}_N - \boldsymbol{\gamma}_S)}_{\substack{\textbf{school}\\ \textbf{effectiveness}}} + \underbrace{(\bar{\mathbf{q}}'_N - \bar{\mathbf{q}}'_S)\boldsymbol{\eta}_N}_{\substack{\textbf{local territory}\\ \textbf{endowment}\\ \textbf{effect}}} \tag{4}$$

Essentially, we have estimated model (9) of Table 2 interacting school variables with the macro-area dummies, after pooling North-East and North-West in a common category (North). This enables us to investigate differences in school effectiveness across geographic areas. The results are reported in Table 4.

[44] If we estimate directly equation (3) by interacting all individual and local variables with geographical dummies, we are unable to identify some of the components since many variables drop from the model due to perfect multicollinearity. If we instead estimate equation (3) by subsamples, we are puzzled by the intercept estimation where we obtain values that are inexplicably high and rather different among areas (+983 for the North, -1470 for the Centre and +135 for the South). As a consequence, the coefficients on the other variables become significantly different across areas, thus making the Oaxaca (1973) decomposition almost meaningless.

TABLE 4 – *Models of students' mathematical literacy [Model (9), Table 2] with interactions between school variables and macro-areas (North, Centre and South)*

Variable	EPFs			means **x**, **s**, **q**		
	North	Centre	South	North	Centre	South
Small town (< 15,000)		1.599		0.18	0.07	0.12
		[0.25]				
City (> 100,000)		-2.824		0.34	0.31	0.22
		[0.59]				
Female		-22.721		0.52	0.52	0.52
		[9.62]***				
Age		8.345		15.71	15.71	15.70
		[1.89]*				
Single parent		-5.489		0.15	0.17	0.14
		[2.07]**				
Highest parental occupational status		0.331		47.83	48.74	44.63
		[4.55]***				
Highest parental education		7.166		12.85	13.14	12.05
in years of schooling		[3.90]***				
Highest parental education		-0.370		177.10	185.08	159.69
in years of schooling squared		[4.77]***				
Computer facilities		7.694		0.04	-0.01	-0.30
at home (*comphome*)		[5.59]***				
No. books at home		0.027		187.51	198.89	129.78
		[4.06]***				
Index of home possessions		2.393		0.07	0.11	-0.22
(*homepos*)		[1.52]				
Memorisation strategies (*memor*)		-8.139		-0.04	-0.02	0.14
		[5.28]***				
Elaboration strategies (*elab*)		5.639		-0.11	-0.05	0.23
		[4.08]***				
Competitive learning (*complrn*)		5.321		-0.11	-0.06	0.35
		[3.59]***				
Co-operative learning (*cooplrn*)		-4.020		0.03	0.08	0.25
		[3.71]***				
High school oriented towards	25.302	-100.63	-82.28	0.08	0.13	0.04
humanities (*liceo classico*)	[1.87]*	[1.17]	[1.04]			
High school oriented towards	51.613	-72.973	-52.523	0.20	0.17	0.27
science (*liceo scientifico*)	[4.34]***	[0.86]	[0.70]			
Technical school	28.003	-94.654	-83.524	0.39	0.37	0.34
(*istituto tecnico*)	[3.32]***	[1.24]	[1.17]			
Vocational school	-34.412	-164.73	-110.97	0.20	0.20	0.23
(*istituti professionali*)	[3.62]***	[2.24]**	[1.62]			
Other schools	-	-122.34	-82.575	0.13	0.14	0.12
	-	[1.58]	[1.18]			
Private school	-0.850	-50.451	-38.303	0.05	0.03	0.03
	[0.06]	[1.80]*	[2.18]**			
Proportion of computers	4.209	20.021	20.626	0.70	0.62	0.75
connected to internet	[0.59]	[2.21]**	[1.97]*			
Teacher/student ratio	1.211	-0.074	1.354	9.09	10.04	10.64
	[0.97]	[0.28]	[2.57]**			

(segue)

TABLE 4 – *(continua)*

	EPFs			means **x**, **s**, **q**		
	North	Centre	South	North	Centre	South
Highest parental education in years of school - school average	-0.020 [0.01]	2.898 [0.55]	-0.782 [0.15]	12.84	13.10	11.99
Highest parental occupational status – school average	0.090 [0.12]	2.168 [2.18]**	2.514 [1.93]*	47.88	48.75	44.64
No. books at home school average	0.076 [1.55]	-0.056 [0.72]	-0.203 [1.78]*	188.44	199.84	129.27
Student-teacher relations at school (*sturel*)	-4.475 [2.93]***	-5.772 [2.88]***	-5.137 [3.08]***	-0.44	-0.46	-0.14
Disciplinary climate in maths lessons (*disclim*)	6.464 [3.97]***	6.312 [2.73]***	7.191 [3.70]***	-0.11	-0.29	0.06
Fraction of students held back	-89.845 [4.68]***	32.826 [0.86]	-64.479 [1.76]*	0.14	0.16	0.13
Temporary-permanent teacher ratio x 100	0.288 [0.81]	-1.954 [1.53]	-0.036 [0.06]	19.68	11.16	12.71
% buildings unfit to school and improperly adapted		0.288 [0.81]		8.14	14.29	19.93
% school buildings rented for schooling		-0.043 [0.19]		10.06	21.76	36.06
Maintenance of buildings - 1st factor		-1.226 [0.59]		-0.45	0.22	0.89
Log govrn. exp. on capital account per student		4.596 [0.80]		11.33	10.99	10.75
Log govrn. exp. on intermediate inputs per student		-22.310 [1.55]		14.22	14.21	14.12
Employment probability (100-unemployment rate, at province lvl)		0.790 [1.02]		94.73	91.89	75.75
% irregular work		-0.697 [2.54]**		16.71	20.91	34.79
N. crimes with unknown author per 100,000 inhabitants		0.000 [0.02]		3,054.52	2,703.51	2,377.45
% without formal education (on pop > 6 years)		-2.014 [0.96]		7.42	9.10	12.75
% rented houses		-1.557 [2.99]***		20.60	17.46	19.72
% foreign-born in resident population (2001)		1.410 [0.60]		2.92	3.18	1.10
social capital - 1st factor		3.482 [1.99]*		-0.72	0.42	-0.60
Observations		8,410				
R-squared		0.49				
Log likelihood		-47,136.50				

Note: Absolute value of t-statistics in brackets. t-statistics are computed using 80 balanced repeated replications and 5 plausible values. The reference individual is male, lives in a town (15,000-100,000) in Northern Italy, comes from an intact family and he is enrolled in *other schools* in the public education sector (see Section 5.1). The number of observations may differ across columns due to missing data. Errors are clustered by province.
* significant at 10%; ** significant at 5%; *** significant at 1%.

From this table we observe that when allowing the 'effectiveness' of schools (i.e. the contribution of school inputs over and above those of student and local variables) to be different across regions, many local variables loose statistical significance, while we observe significant differences across schools by macro-areas. A possible interpretation is that the local environment acts so as to increase (or reduce) the 'effectiveness' of schools with the same characteristics in terms of producing mathematical literacy.

The performance of students in the scientific high school track in Northern Italy is by far the highest, followed by technical schools and humanities high schools (*licei classici*) in the same area. To be noted is that none of the school type coefficients in the Centre and in the South is positive, which suggests that students in these areas perform worse than most of Northern students, irrespective of the school type in which they are enrolled in.[45] Attending a high school (*liceo*) is much more effective in Northern regions than in Southern ones. Note also that failure seems to be a stronger signal in the former than in the latter regions: in fact both areas have similar fractions of repeaters, but being in a school with a larger fraction of them creates a disadvantage only in the former regions. Combined with the fact that students in private schools have a lower level of mathematical literacy only in the South, we propose a possible interpretation, which is consistent with our evidence. While in the North student sorting occurs according to ability and is based on school tracking and grade retention (if you are less talented you are readdressed towards technical/vocational schools and/or are held back one or more years), in the South students are less sorted among tracks (the dummies identifying school types are all insignificant except for professional schools), and in case of low performance they are readdressed towards the private sector.

Therefore, we observe some heterogeneity in educational production functions across macro-areas. In order to assess how much of the difference in student performance is attributable to differences in the educational processes (the coefficients of the EPFs) and how much is due to different students' characteristics we decompose the predicted differential in students' scores according to equation (4) in Table 5. While school 'effectiveness' accounts for one fourth of the differences between North and South, endowment covers the complementary fraction.

[45] In addition, the impact of socio-economic status at school level (proxied by average highest ISEI by school) appears to be stronger and statistically significant in Central and Southern Italy when compared to the North. Similarly, access to the internet is correlated with better numeracy especially in Centre and South Italy. The correlation between good student-teacher relations or the disciplinary climate and mathematical literacy is similar across areas.

TABLE 5 – *Accounting for territorial differences*

	$(\bar{\mathbf{x}}'_N - \bar{\mathbf{x}}'_S)\boldsymbol{\beta}_N$ individual endowment	$(\bar{\mathbf{s}}'_N - \bar{\mathbf{s}}'_S)\boldsymbol{\gamma}_N$ school endowment	$\bar{\mathbf{s}}'_S(\boldsymbol{\gamma}_N - \boldsymbol{\gamma}_S)$ school effectiveness	$(\bar{\mathbf{q}}'_N - \bar{\mathbf{q}}'_S)\boldsymbol{\eta}_N$ local territory endowment	overall difference
$\bar{y}_N - \bar{y}_C$	0.49	3.27	22.31	4.05	30.11
$\bar{y}_N - \bar{y}_S$	3.05	7.42	19.04	46.41	75.92

According to our decomposition, even an ideal situation characterised by equal resources at school level[46] would not correspond to the absence of a territorial gap in students' competencies. However, data available in the PISA survey do not allow us to disentangle the effect of different styles of school management from that of student sorting. Indeed, we have previously noted that student allocation to school types may be more effective in the North. Therefore, what we interpret as 'school effectiveness' could be a composition of better screening of students at lower secondary school level and the outcome of different ways of organising and managing upper secondary schools.

7. CONCLUDING REMARKS

In this paper we have studied the territorial divide in mathematical performance of Italian students as measured in PISA 2003. We argue that the particular concept of performance measured in PISA ('knowledge for life') calls for a significant role of the local environment in explaining territorial disparities.

In order to test this hypothesis we have merged PISA data with school administrative data provided by the Italian Ministry of Public Education, and with territorial data on various social phenomena at province level.

Our results show that province variables are indeed important in partially accounting for the North-South gap in students' mathematical performance. Among the most significant and sizeable correlations are worth citing those of school infrastructures and the state of the local labour market, in terms of both employment probability and extension of the irregular economy. In accordance with the literature, we also find support for the fact that financial resources invested in schools are positively correlated with student performance only when they are spent in equipment and buildings, but not in teachers or auxiliary personnel.

[46] Note that in our sample the teacher/student ratio is highest in Southern schools, similarly the fraction of computers connected to internet.

As we already acknowledged, our study has only descriptive nature. However, we think it is nonetheless interesting since it represents a first attempt to relate geographical differences in student performance in Italy to territorial variables. After further investigation, which would be necessary in order to check the causal nature of the estimated correlations, our analysis could give important suggestions to policy makers. For instance, if the estimated correlations partly reflect causation, a policy focusing only on Southern schools (e.g., increasing school buildings quality) to reduce the North-South gap might not be successful if the local environment surrounding schools in Southern regions deteriorates. For instance, individuals facing high unemployment rates may perceive that putting their effort into study is not worth undertaking, since it will not help them to find a decent job or to obtain higher salaries. Spending their time differently (e.g., working in the irregular sector) could represent a more economically rewarding activity than investing in human capital. A policy simultaneously targeting schools, families and the local socio-economic environment might be much more effective in reducing territorial disparities.

Our final section raises more questions than it provides answers. Indeed, we show that the difference in mathematical literacy between North and South Italy is attributable for 25% to different school management across regions, while the remaining share of variance is related to factor endowment. PISA survey does not provide sufficient information about teacher behaviour in classes, and we suspect that part of this difference has to do with teachers' way of conduct in classes and in schools. Since most of the school heads were formerly teachers, the information they provide may not be insightful on this topic, and further investigation is required to assess the reasons for the territorial differences in the educational processes.

References

BRATTI, M., CHECCHI, D. AND FILIPPIN, A. (2007), "Da dove vengono le competenze degli studenti?", Bologna, il Mulino.

BRUNELLO, G. AND CHECCHI, D. (2004), "School vouchers Italian style", *Giornale degli Economisti e Annali di Economia*, vol. 63, pp. 357-399.

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. (2006), "Mobilità intergenerazionale e decisioni scolastiche in Italia", in *Scelte individuali e vincoli strutturali. Sistema scolastico e disuguaglianza sociale*, ed. by G. Ballarino - D. Checchi, Bologna, il Mulino, pp. 27-56.

CIPOLLONE, P. AND SESTITO, P. (2007), "Quanto imparano gli studenti italiani: I divari nord-sud", *mimeo*.

GANZEBOOM, H.B.G. AND TREIMAN, D.J. (1996), "Internationally comparable measures of occupational status for the 1988 International Standard Classification of Occupations", *Social Science Research*, vol. 25, pp. 201-239.

GREENWALD, R., HEDGES, L.V AND LAINE, L.R. (1996), "The effect of school resources on student achievement", *Review of Educational Research*, vol. 66, pp. 361-396.

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.

LAZEAR, E.P. (2001), "Educational production", *Quarterly Journal of Economics*, vol. 116, pp. 777–803.

MICUCCI, G. AND NUZZO, G. (2003), "La misurazione del capitale sociale: evidenza da una analisi del territorio italiano", *Papers and proceedings of the Conferences on "Economie locali, modelli di agglomerazione e apertura internazionale. Nuove ricerche della Banca d'Italia sullo sviluppo territoriale",* Bologna 20th November 2003, pp. 207-250.

MONTANARO, P. (2007), "La qualità dell'istruzione italiana", *mimeo*.

OAXACA, R. (1973), "Male-female wage differentials in urban labor markets", *International Economic Review*, vol. 14, pp. 693–709.

OECD (2003), Education at a glance, Paris, OECD.

OECD (2004), Learning for tomorrow's world - First results from PISA 2003, Paris, OECD.

OECD (2005a), PISA 2003 technical report, Paris, OECD.

OECD (2005b), PISA 2003 Data Analysis Manual. SPSS Users, Paris, OECD.

PLOMIN, R. AND PETRILL, S. (1997), "Intelligence: What's new?", *Intelligence*, vol. 24, pp. 53-77.

TODD, P.E. AND WOLPIN, K.I. (2003), "On the specification and estimation of the production function for cognitive achievement", *Economic Journal*, vol. 113, pp. F3-F33.

TRAMONTE, L. (2004), "Nell'Italia dei quindicenni, un'esplorazione delle differenze territoriali nelle competenze alfabetiche funzionali. Come si modellano le abilità nella lettura rilevate dalla ricerca OECD-PISA 2000 secondo le componenti economiche, culturali e sociali del background degli studenti", unpublished PhD dissertation, University of Milan.

APPENDIX

TABLE A1 – *Dataset summary statistics*

Variable	Obs.	Mean	Std.Dev.	Min	Max
Individual information					
PISA score mathematics (mean of 5 Plausible Values)	11,565	467.974	90.120	109.160	771.179
North-West	11,565	0.216	0.412	0.000	1.000
Centre	11,565	0.188	0.391	0.000	1.000
South and Islands	11,565	0.444	0.497	0.000	1.000
Small town (< 15,000)	11,565	0.131	0.337	0.000	1.000
City (> 100,000)	11,565	0.312	0.463	0.000	1.000
Female	11,565	0.523	0.500	0.000	1.000
Age	11,565	15.707	0.285	15.250	16.250
Single parent	11,410	0.155	0.362	0.000	1.000
Highest parental occupational status (0-100)	11,325	47.038	16.686	16.000	90.000
Highest parental education in years of schooling	11,490	12.568	3.672	0.000	17.000
Computer facilities at home (*comphome*)	11,532	-0.139	0.957	-1.676	1.051
No. books at home	11,405	161.593	203.289	5.000	750.000
Index of home possessions (*homepos*)	11,548	-0.074	0.948	-3.787	1.939
Memorisation strategies (*memor*)	11,495	0.032	0.881	-3.483	3.292
Elaboration strategies (*elab*)	11,499	0.029	0.964	-3.262	3.263
Competitive learning (*complrn*)	11,480	0.093	0.945	-2.844	2.450
Cooperative learning (*cooplrn*)	11,483	0.137	0.974	-3.134	2.742
School information					
Proportion of computers connected to internet	11,311	0.711	0.329	0.000	1.000
Student-teacher ratio	11,465	10.007	5.494	1.758	88.250
Highest parental education in years of schooling - school average	11,565	12.568	1.671	8.793	17.000
Highest parental occupational status – school average	11,565	46.935	8.421	29.176	70.469
No. books at home - school average	11,565	161.018	87.741	25.652	537.500
Student-teacher relations at school (*sturel*)	11,497	-0.301	0.947	-3.090	2.855
Disciplinary climate in maths lessons (*disclim*)	11,467	-0.094	1.039	-2.738	2.353
Fraction of students held back	11,565	0.047	0.142	0.000	1.000
High school oriented towards humanities (*liceo classico*)	11,565	0.083	0.276	0.000	1.000
High school oriented towards science (*liceo scientifico*)	11,565	0.214	0.410	0.000	1.000
Technical school (*istituto tecnico*)	11,565	0.355	0.478	0.000	1.000
Vocational school (*istituti professionali*)	11,565	0.218	0.413	0.000	1.000
Private school	11,565	0.043	0.203	0.000	1.000
Temporary-permanent teacher ratio x 100	9,254	16.160	8.402	4.870	53.730

(segue)

TABLE A1 – *(continua)*

Variable	Obs.	Mean	Std.Dev.	Min	Max
Territorial information (province level)					
% buildings unfit to school and improperly adapted	78	14.018	9.188	2.740	38.890
% school buildings rented for schooling	78	19.576	14.085	0.000	59.700
Maintenance of buildings - 1st factor	78	0.123	1.261	-2.025	4.104
Log govrn. exp. on capital account per student	77	11.147	0.517	8.938	11.975
Log govern.exp. on teachers per student	78	15.369	0.085	15.180	15.623
Log govrn. exp. on intermediate inputs per student	77	14.222	0.198	13.694	14.776
Class size (no. students)	78	21.531	0.963	18.670	23.660
Employment probability (100-unemployment rate, province lvl)	81	88.499	9.489	62.979	97.904
% irregular work	81	25.519	11.536	8.000	53.000
N. crimes with unknown author per 100,000 inhabitants	81	2,381.55	822.96	1,158.00	5,250.38
% without formal education (on pop > 6 years)	81	9.517	2.820	4.989	15.965
% rented houses	81	18.159	4.661	9.412	36.001
% foreign-born in resident population (2001)	81	2.311	1.130	0 .451	5.493
No. suicides attempted per 100,000 inhab.	81	7.602	4.458	1.072	27.427
No. volunteers in non-profit organisations per 100 inhab.	81	6.870	3.005	0.278	19.507
social capital - 1st factor	81	0.017	1.472	-4.42	3.73

Note: This table reports data summary statistics. Statistics for the variables in the original PISA 2003 data set were weighted using student final weights. The summary statistics refer to the variables in the data set and consider all observations with non-missing values.