

IZA DP No. 2603

**Territorial Differences in Italian Students'  
Mathematical Competencies:  
Evidence from PISA 2003**

Massimiliano Bratti  
Daniele Checchi  
Antonio Filippin

February 2007

# **Territorial Differences in Italian Students' Mathematical Competencies: Evidence from PISA 2003**

**Massimiliano Bratti**

*University of Milan and IZA*

**Daniele Checchi**

*University of Milan and IZA*

**Antonio Filippin**

*University of Milan and IZA*

Discussion Paper No. 2603  
February 2007

IZA

P.O. Box 7240  
53072 Bonn  
Germany

Phone: +49-228-3894-0  
Fax: +49-228-3894-180  
E-mail: [iza@iza.org](mailto:iza@iza.org)

Any opinions expressed here are those of the author(s) and not those of the institute. Research disseminated by IZA may include views on policy, but the institute itself takes no institutional policy positions.

The Institute for the Study of Labor (IZA) in Bonn is a local and virtual international research center and a place of communication between science, politics and business. IZA is an independent nonprofit company supported by Deutsche Post World Net. The center is associated with the University of Bonn and offers a stimulating research environment through its research networks, research support, and visitors and doctoral programs. IZA engages in (i) original and internationally competitive research in all fields of labor economics, (ii) development of policy concepts, and (iii) dissemination of research results and concepts to the interested public.

IZA Discussion Papers often represent preliminary work and are circulated to encourage discussion. Citation of such a paper should account for its provisional character. A revised version may be available directly from the author.

## ABSTRACT

### **Territorial Differences in Italian Students' Mathematical Competencies: Evidence from PISA 2003\***

In this paper we investigate the existence and the size of territorial differences in Italian students' mathematical competencies. Our analysis benefits from a new data set that merges the 2003 wave of the OECD *Programme for International Student Assessment* (PISA) with territorial data collected from several statistical sources and with administrative school data collected by the Italian Ministry of Education. We consider three different groups of educational inputs: individual characteristics (mainly family background), school types and available resources, and territorial features related to labour market, cultural resources and aspirations. In addition to the standard gradient represented by parental education and occupation, we find that student sorting across school types also plays a significant role. Among the local factors measured at province level, we find a significant impact of buildings maintenance and employment probabilities. When accounting for territorial differences, we find that most of the North-South divide (75%) is accounted for by differences in endowments, while the local school production functions account for the remaining fraction.

JEL Classification: J21, J24, H52

Keywords: education, PISA, students, territorial differences

Corresponding author:

Massimiliano Bratti  
Department of Economics, Business and Statistics (DEAS)  
University of Milan  
Via Conservatorio 7  
I-20122 Milan  
Italy  
Email: [Massimiliano.Bratti@unimi.it](mailto:Massimiliano.Bratti@unimi.it)

---

\* 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), at the University of Milan (Milan, June 2006), the conferences 'Economics of Education: Major Contributions and Future Directions' (Dijon, June 2006) and 'V Brucchi Luchino conference' (Padua, November 2006) for useful suggestions. We thank in particular Giuseppe Bertola and Lawrence Kahn. The authors only are responsible for the content of this paper.

# 1 Introduction

In international surveys on students' competencies such as the *Programme for International Student Assessment* (PISA), the *International Mathematics and Science Study* (TIMSS), the *Progress in International Reading Literacy Study* (PIRLS), Italy typically ranks low.<sup>5</sup> This is a common feature among most Mediterranean countries, and could be related to the relatively recent improvements in the level of education of the population.

At the exit 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%.<sup>6</sup>

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. Simply by running an OLS regression of the level of PISA test scores in mathematical literacy on macro-area dummies (North-East, North-West, Centre, South and Islands) we obtain statistically significant differences across the Italian territory. Indeed, the difference between South and Islands and the reference category (North-East) amounts at 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 North-East. North-West does not perform differently from the reference category (see Table 1).

[Table 1 about here]

This has lead us to wonder which are the determinants of such large territorial differences in student performance, 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. 90% 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

---

<sup>5</sup>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 lowest two levels of competencies.

<sup>6</sup>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.

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<sup>7</sup> 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 territorial distribution of social phenomena like 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 available at local level to the 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'*.

Although our paper has descriptive aim, since it estimates statistical associations without giving any causal interpretation, we think that our 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 imply causation, the latter cannot be excluded. Therefore, we believe to have been able to identify territorial variables whose importance should be further investigated in order to assess possible causal effects.

In the sequel, after reviewing the (rather limited) 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 concluding remarks.

## 2 The literature on the argument

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.<sup>8</sup> He investigates the

---

<sup>7</sup>When not stated otherwise, we refer to "student performance" meaning PISA test scores in mathematics.literacy

<sup>8</sup>It should be reminded that the Italian upper secondary school system can be described as tripartite, with a an academic oriented generalist education provided by high schools (5 years,

determinants of secondary school track and finds that parental education and socio-economic status are the main determinants of track choice.

The same exercise has been repeated on the 2003 survey, where additional information on pre-existing ability of students was introduced, 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 and Flabbi, 2006). Checchi (2004) also analysed the role of family background and school level peer effects in affecting students' performances in PISA test scores. His main result was that average parental education and socio-economic prestige measured at school level is a much stronger predictor than individual variables, thus indirectly validating that environmental factors may be important determinants of student performance. Indeed, even when controlling for individual background and self-selection of students in different 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.

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<sup>9</sup> (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. Although the model is quite successful in explaining between-school variance, a consistent part of both individual and regional differences is left 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

---

called *licei*, with further division in humanities, scientific activities, languages, pedagogical sciences), 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 could 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.

<sup>9</sup>Since only this level of information is available in PISA 2000.

sociologists, but they also have drawn attention of the media and the general public. For this reason we report here only their main characteristics.<sup>10</sup> 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 2003 and whose main focus is on 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 a knowledge society.

PISA data gathers a wealth of information both on student's and school's characteristics. The latter are collected through a questionnaire answered by the teacher head of each school that entered PISA national samples.

For our purpose of analysing geographical differences, the main drawback of PISA survey in Italy is that its sample design only contains very aggregated information of school geographical location (11 areas), which does not enable an analysis of the role of territorial factors.<sup>11</sup>

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.<sup>12</sup>

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 organised 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.<sup>13</sup> 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

---

<sup>10</sup>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).

<sup>11</sup>The official data set released by the OECD does not even allow an analysis by macroregion, e.g. North, Center and South, since some Center and Southern regions are aggregated in a common macroarea.

<sup>12</sup>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.

<sup>13</sup>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.

permanent contracts, teachers with temporary contracts, type of teacher's qualification, age). MPI finally returned the merged and anonymised dataset to us.

It must be noted that the explicit stratification variables used in the Italian sample design (see OECD 2005a, p. 54) are aggregate geographical regions (11), school types (4) and school size (2), while only one implicit stratification variable is used (public/private school). 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 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 disappears.

With this cautionary note in mind, Figure 1 reports the quintiles of the average students' mathematical literacy by province. The map shows only raw differences, i.e. unadjusted for contextual factors such as family background characteristics or school type of the student population sampled in each province, and 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, but the map is only meant to give a broad picture of territorial differences in students' mathematical literacy. After controlling for contextual factors, such as individual, school and local variables, we will be 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.<sup>14</sup> 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 province variables are likely to capture other unobserved effects that are common within macro-areas. If this happens, we are able to account for the North-South divide using observable information.

---

<sup>14</sup> We have seen in section 1 why this is a relevant question in Italy.



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

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

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

where  $i=1,\dots,N$ ,  $j=1,\dots,J$  and  $p=1,\dots,P$  are the subscripts for individuals, schools and provinces respectively.  $\varepsilon_{ijp}$  is a stochastic 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.

A crucial assumption to obtain unbiased estimates of  $\boldsymbol{\beta}$ ,  $\boldsymbol{\gamma}$  and  $\boldsymbol{\eta}$  is that the error term  $\varepsilon_{ijp}$  must be uncorrelated with the explanatory variables included in the right-hand side (RHS, hereafter) of equation (1), i.e. we did not omit any relevant variable that is correlated with those included in the regression. This assumption is probably not too strong for individual and school characteristics when using PISA data, which gather a wealth of information on individuals and schools (even if not very rich with respect to teachers). However, it must be noted that one important variable is missing in PISA, namely students’ 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 determine biased coefficients (*ability bias*). This represents less of a problem for the analysis in this paper, where we do not aim at estimating causal effects but only robust correlations between local variables and student performance. Indeed, given that we run an analysis at individual level while local variables are measured at province level, we think that the estimated of their coefficients are less prone to the *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 their local environment (such as employment rates) to get unbiased coefficients.

A possible way through which *endogeneity* may still arise 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. We assume that by controlling both for several parents’ characteristics (such as wealth, socio-economic status and education), we are likely to address this form of *self-selection* (or *endogeneity*).<sup>15</sup> As we already said, a full assessment of the ‘effects’ of local variables would require an analysis of causality, which cannot be performed with the information available in our data set.<sup>16</sup>

## 5 Results

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

### 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 progressive addition of control variables.<sup>17</sup> 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 raise multicollinearity problems.<sup>18</sup> In what follows we describe our main findings.

[Table 2 about here]

*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,<sup>19</sup> but at the same time we mainly include factors that are presumably exogenous and for which a problem of reverse causality can

---

<sup>15</sup> That is, we assume selection only on the observables included in the models.

<sup>16</sup> An analysis of causal effects would require finding good ‘instruments’, i.e. variables that are correlated with the explanatory variable of interest but not with the error term.

<sup>17</sup> Table A1 in the Appendix reports sample summary statistics. Notice that we exclude 15-year-old students who were still in lower secondary schools, due to repeated failures (74 observations). In our econometric analysis we include individuals in grades 8, 9 and 10 without controlling for grade. This is done essentially since the grade attended is strongly related to the probability to be held back at school in previous years, which, in turn, is not exogenous with respect to mathematical competencies. In the analysis, we will also estimate our preferred specification only on students attending the 9<sup>th</sup> grade.

<sup>18</sup> The results of this factor analysis are available upon request from the authors.

<sup>19</sup> 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.

be reasonably excluded.<sup>20</sup> 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.<sup>21</sup> 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 index *ISEI*, see Ganzeboom and Treiman, 1996). Also cultural capital (proxied by number of books at home, the availability of computer facilities at home and parents' education) is positively associated with student performance, although the last variable shows a non-linear relation. The use of elaboration learning strategies and competitive learning has 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.*<sup>22</sup> 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 were offered an indicator of the type of school program that is more detailed than that commonly available through PISA.<sup>23</sup> The type of school attended comes out to be a very important determinant 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 classic high schools are not statistically different from the excluded 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 reflect self-selection of

---

<sup>20</sup> 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.

<sup>21</sup> An opposite gender gap is found in reading literacy test scores.

<sup>22</sup> 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 teachers' 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 in the opposite direction.

<sup>23</sup> For 7 schools (6 in the province of Bolzano and 1 in the province of Aosta, corresponding to 199 students) data on school type is 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 prefer to stick to MPI data and drop these schools.

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 and Checchi, 2004). Better student-teacher relations are associated with a lower performance, while a tougher disciplinary climate positively correlates with performance, another well known 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 turns out to be significant only in model 2 while they lose 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 opportunity 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 column 3 we start including some characteristics of school infrastructures.<sup>24</sup> 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 structures. The third variable corresponds to the first factor extracted from the maintenance state of six building's attributes,<sup>25</sup> 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.

*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 consumption and other personnel, and expenditures on capital account.

We observe a highly significant positive correlation between the educational expenditure on capital account<sup>26</sup> and student performance, which reinforces the idea of a potentially important role of schooling infrastructures. By contrast, our estimates show a negative and significant correlation between performance and

---

<sup>24</sup> When we use these data, which are provided by MPI, we lose 2,500 observations relating to the provinces of Trento, Bolzano and Aosta corresponding to 69 schools.

<sup>25</sup> Roofing, flooring, water system, sewerage plant, electric system, heating-plant.

<sup>26</sup> These data refer to 1998 and 1999. The expenditure in 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.

the expenditure in intermediate consumptions and other personnel.<sup>27</sup> Both effects are quite robust across specifications. The most surprising result, although in line with the international literature, is the absence on any positive correlation with the expenditures for teachers in model 4.<sup>28</sup> It must be noted that the same effect emerged when controlling for student-teacher ratios provided by school head teachers. Alternatively, when we consider the average class size we do not find any significant effect. 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 subsequent 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 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 what they directly measure.<sup>29</sup> However, since we control for family background and school factors, we are tempted to suggest a possible ‘causal’ interpretation. Individuals living in areas with a well functioning local labour market experience higher employment probabilities; anticipating higher expected returns to education they invest more in education. By contrast, the larger the extension of the irregular sector, which mainly employs unskilled work, 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 devoting time to alternative activities.

*Local cultural factors.* In model 6 we make an attempt to introduce a variables proxying 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<sup>30</sup> 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 consumption of TV programmes, and we did not find any statistically significant correlation, and therefore they do not appear in our estimated model. Vice versa, the fraction of

---

<sup>27</sup> Which is obtained residually from total expenditures after subtracting expenditures in capital account and for teachers.

<sup>28</sup> The expenditures for teachers is provided by MPI and is given by the teachers salaries plus the *indennità integrativa*. The province of Udine did not provide these data.

<sup>29</sup> We did not include in these specification the (log of) GDP per capita at province level given its high correlation with the employment rate.

<sup>30</sup> 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 the this variable may also be a proxy of tax evasion.

population without formal schooling is negatively correlated with student performance. This evidence is partial surprise, especially in the light of the fact that our model already controls for an individual parents' education and for the average parental education at school level. To have a rough idea of its effect, 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.

*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 from the first variable, which may approximate the municipality size, complementing the categorical variables already included in the individual data section (reclassifying the categorical variable provided by PISA).<sup>31</sup> 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 column 8 we add some proxies for social capital. Once again we experimented with several variables, among which only two seemed worth including in our regressions. We only retained 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 no-profit organisations (per 100 inhabitants) captures the degree of solidarity within each province. Both variables turn out being statistically insignificant (see model (8)). We also tried other variables existing in the literature as alternative proxies for social capital (such as participation to political elections, blood donations, etc. – alternatively the factor extracted using principal component analysis: see Micucci and Nuzzo, 2003) but all were not significant (see model (9)).

We also estimated the model on the population of individuals who were attending the 9<sup>th</sup> grade only (i.e. who were regular in their previous educational career)<sup>32</sup>, in order to control potential sample distortions introduced by different policies of student failures 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 effects are both qualitatively and quantitatively very similar to model (9), the only notable exception being the negative effect of being in a single parent family, which disappears when considering 9<sup>th</sup> graders only. This indirectly suggests that being in a non-intact family mostly increase 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

---

<sup>31</sup> 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).

<sup>32</sup> 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.

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' performance in IQ tests (see Plomin and 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. Column 1 shows that individual variables are not sufficient to explain regional differences. Indeed the difference between the North-East and South (including also Islands) is almost 70 test points. Including school variables in column 2 improves the explanatory power of the model but does not help explaining the North-South divide in students' mathematical performance. Including information on school infrastructures in column 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 inclusion of the variables related to the local labour market. In this case the coefficient on the South dummy falls by 50%, becoming not statistically significant at 5%. Inclusion of other control variables induces further loss of significance in the coefficient. The coefficient on the Centre Italy dummy remains instead statistically significant also in the most general specification, although the inclusion of territorial variables determines a reduction of its size by about 50%.

[Table 3 about here]

We interpret these findings as evidence that allocating students from Southern Italy to schools endowed with the same resources and located in similar social environments as Northern schools would reduce most of the gap in mathematical literacy.<sup>33</sup>

This concept can be illustrated by comparing figures 1 and 2. Figure 1 reports 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 excellent performance both in the North and in the South of Italy, including several provinces of Southern Sicily and Sardinia.

---

<sup>33</sup> In section 6 we see that school effectiveness accounts for the remaining differences, particularly in Centre Italy.

[Figure 1 and 2 about here]

## 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

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

we can obtain the following decomposition

$$\begin{aligned} \bar{y}_N - \bar{y}_S = & \underbrace{(\alpha_N - \alpha_S)}_{\text{unobservables endowment}} + \underbrace{(\bar{\mathbf{x}}'_N - \bar{\mathbf{x}}'_S) \boldsymbol{\beta}_N}_{\text{individual endowment effect}} + \underbrace{\bar{\mathbf{x}}'_S (\boldsymbol{\beta}_N - \boldsymbol{\beta}_S)}_{\text{individual returns}} + \\ & + \underbrace{(\bar{\mathbf{s}}'_N - \bar{\mathbf{s}}'_S) \boldsymbol{\gamma}_N}_{\text{school endowment effect}} + \underbrace{\bar{\mathbf{s}}'_S (\boldsymbol{\gamma}_N - \boldsymbol{\gamma}_S)}_{\text{school effectiveness}} + \\ & + \underbrace{(\bar{\mathbf{q}}'_N - \bar{\mathbf{q}}'_S) \boldsymbol{\eta}_N}_{\text{local territory endowment effect}} + \underbrace{\bar{\mathbf{q}}'_S (\boldsymbol{\eta}_N - \boldsymbol{\eta}_S)}_{\text{local returns}} \end{aligned} \quad (3)$$

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.<sup>34</sup> In addition, we are mostly interested in the effect of variables that in principle can be modified by the educational authorities (like

<sup>34</sup> If we estimate directly equation (3) by interacting all individual and local variables with geographical dummies, we are unable to identify part of the effect since many variables drop from the model due to perfect multicollinearity. When considering the surviving variables, we find that, contrary to our expectations, the impact of family background (parental education and occupation, books and computers at home) is stronger in Northern regions than in Southern ones. On the contrary, the gender gap gets stronger and stronger the more southward we go in the peninsula. 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.



school endowments and/or their impact on competences formation). As a consequence, the territorial differences are decomposed according to the following equation

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

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.

From this table we observe that when allowing the effectiveness of schools 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 in the same area. To be noted are the negative coefficients of all the other school type-area interactions, which suggest that students in all these schools perform worse than Northern students, irrespective of the school type in which they are enrolled in.<sup>35</sup> Thus we observe that student performance is much less differentiated across school types in the South (given the absence of statistical significance of any school type in this region) compared to the other areas of Italy. Attending a high school is much more effective in Northern regions than in Southern ones. Notice also that failure seems to be a stronger signal in the former than in the latter regions: in fact both regions have similar fractions of repeaters, but being in a school with a larger fraction of them creates a disadvantage only in the former regions. Combining with the fact that students in private schools have a lower level of mathematical literacy only in the South, we propose the following interpretation. While in the North student sorting occurs according to ability and is based on school tracking and repetition (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 a lower 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 students' performance is attributable to differences in the educational processes (the

---

<sup>35</sup> In addition, the impact of socio-economic status at school level (proxied by average HISEI 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.

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, the endowment effect covers the complementary fraction. Most of it is due to local effects: 46 score points over 76 (61.1%) of North-South divide would disappear if Southern schools would ideally be transferred in the North. A more significant percentage of the North-Centre gap in student performance could be closed by adopting Northern "styles" in managing schools in Centre Italy (74% of the North-Centre difference).

According to our decomposition, even equalising resources at school level<sup>36</sup> would not close the territorial gap in students' competencies. This is not good news for the Italian educational authorities. However, data available in the PISA survey do not allow us to disentangle the effect of school management styles from that of student sorting. Indeed, we have previously noticed that student allocation across school types may be more effective in the North. Therefore, what we interpret as "school effectiveness" (i.e. value added) 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 Italian students' mathematical performance 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 explaining the North-South gap in students' mathematical performance. Among the most significant and sizeable effects 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 and illegal economies. In accordance with the literature, we also find support to the fact that financial resources matter for student performance only when they are spent in equipment and buildings, but not in teachers or auxiliary personnel.

As we already acknowledged, our study is only of a 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

---

<sup>36</sup> Notice that in our sample the teacher/student ratio is highest in Southern schools. Similarly the fraction of computers connected to internet.

suggestions to policy makers. For instance, if the estimated correlations partly reflect causation, a policy focusing only on Southern schools (e.g., increasing the buildings quality) to reduce the North-South gap might not be successful if the local environment surrounding individuals 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 finding a decent job or having higher salaries. Spending their time differently (e.g., working in the irregular sector) could represent a more economically rewarding activity than investing in their human capital. A policy simultaneously targeting schools, families and the local socio-economic environment might be much more effective in reducing territorial disparities.

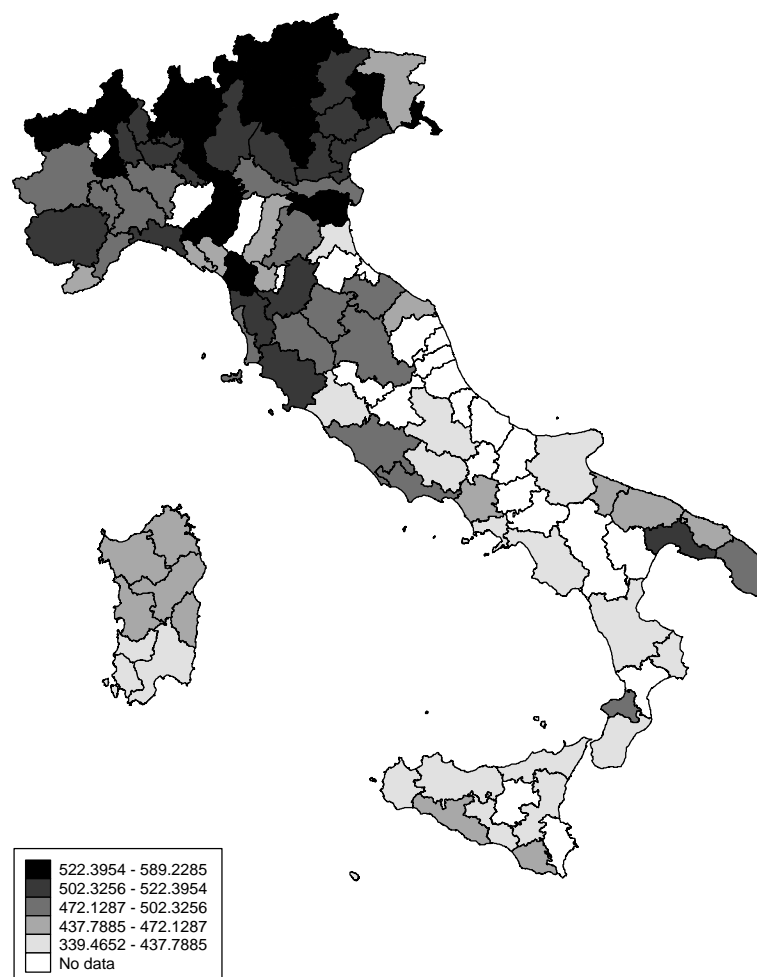
Our final section raises more questions than it provides answers. Indeed, we show that the difference in mathematical literacy between North and South Italy is attributable for 25% to different school processes across regions, while the remaining share of variance is related to factor endowment. PISA 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., D. Checchi and A. Filippin (2007). *Qualità della formazione scolastica, uso delle risorse e fattori ambientali. L'indagine PISA 2003 in Italia*. forthcoming by Fondazione per la Scuola della Compagnia di San Paolo.
- Brunello, G. and D. Checchi (2004). 'School vouchers Italian style', *Giornale degli Economisti e Annali di Economia*, 63: 357-399
- Checchi, D. (2004). 'Da dove vengono le competenze scolastiche? L'indagine PISA 2000 in Italia', *Stato e Mercato*, 72: 413-453.
- Checchi, D. and L. Flabbi (2006). 'Mobilità intergenerazionale e decisioni scolastiche in Italia' in G. Ballarino and D. Checchi (eds) 2006, *Scelte individuali e vincoli strutturali. Sistema scolastico e disuguaglianza sociale*, Bologna: Il Mulino.
- Ganzeboom, H.B.G. and D. J. Treiman (1996). 'Internationally Comparable Measures of Occupational Status for the 1988 International Standard Classification of Occupations', *Social Science Research*, 25: 201-239.
- Lazear, E. P. (2001). "Educational production", *Quarterly Journal of Economics* 116: 777-803.
- Micucci, G. and G. Nuzzo. (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.
- Oaxaca, R. (1973). 'Male-Female Wage Differentials in Urban Labor Markets', *International Economic Review* 14:693-709.
- OECD (2003). *Education at a glance*. Paris: OECD.
- OECD (2004). *Learning for tomorrow's world - First result from PISA*. Paris: OECD.
- OECD (2005a). *PISA 2003 technical report*. Paris: OECD.
- OECD (2005b). *PISA 2003 Data Analysis Manual. SPSS Users*. Paris: OECD.
- Plomin, R. and S. Petrill (1997). 'Intelligence: What's New?', *Intelligence*, 24: 53-77.
- 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.

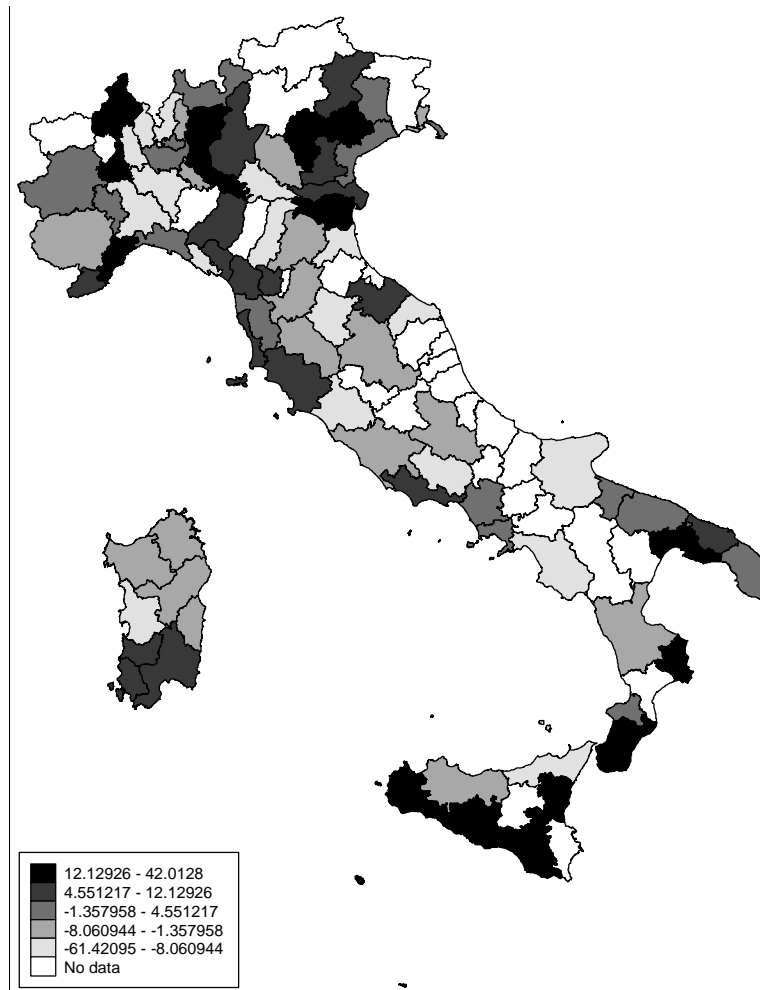
## Tables and Figures

**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, Biella, Benevento, Campobasso, Chieti, Catanzaro, Enna, Forlì, Isernia, Macerata, Matera, Piacenza, Pescara, Prato, Potenza, Reggio Emilia, Rieti, Rimini, Siracusa, Teramo, Terni.

**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. 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).

**Table 1.** Raw territorial differences in students' mathematical literacy (PISA 2003) in Italy

Area	Coef.
North-West	0.492 [0.05]
Centre	-39.075 [3.95]***
South and Islands	-82.797 [8.47]***
Observations	11565
R <sup>2</sup>	0.16
Log likelihood	-67936.7

*Note.* This table reports the estimates of a regression of individual students' mathematical literacy (PISA 2003) on macro-area dummies. The reference group is North-East. Standard errors and t-statistics are computed using 80 balanced repeated replications and 5 plausible values.

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

Variable	Model									
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
	individual	+ school	+ infrastru ctures	+ expendi tures	+ labour mkt	+ cultural	+ demo graphic	+ social capita	(7) + social cap. indx	(9) only on 9 <sup>th</sup> grade students
Small town (< 15,000)	11.116 [1.32]	2.947 [0.34]	-10.433 [1.40]	-7.191 [0.93]	-5.115 [0.76]	-6.335 [0.97]	-3.945 [0.61]	-4.863 [0.76]	-3.905 [0.61]	-5.979 [0.87]
City (> 100,000)	2.436 [0.30]	-3.119 [0.45]	-13.966 [3.03]***	-9.404 [2.31]**	-3.564 [0.85]	-3.483 [0.79]	-2.387 [0.55]	-2.682 [0.60]	-2.197 [0.50]	-3.035 [0.64]
Female	-18.254 [4.20]***	-23.292 [7.53]***	-22.234 [8.12]***	-22.323 [8.09]***	-22.541 [8.32]***	-22.675 [8.48]***	-22.172 [8.34]***	-22.129 [8.30]***	-22.045 [8.27]***	-24.034 [8.02]***
Age	15.035 [3.10]***	8.100 [1.79]*	9.135 [2.10]**	8.792 [1.99]*	8.372 [1.91]*	8.394 [1.91]*	8.988 [2.03]**	8.882 [2.00]**	9.064 [2.04]**	4.139 [0.90]
Single parent	-11.021 [3.60]***	-6.636 [2.15]**	-6.129 [2.10]**	-6.180 [2.22]**	-5.798 [2.14]**	-5.853 [2.16]**	-5.598 [2.07]**	-5.620 [2.07]**	-5.537 [2.04]**	-2.368 [0.75]
Highest parental occupational status	0.786 [7.56]***	0.295 [3.94]***	0.311 [4.16]***	0.335 [4.63]***	0.335 [4.62]***	0.334 [4.59]***	0.335 [4.62]***	0.335 [4.61]***	0.336 [4.63]***	0.216 [2.86]***
Highest parental education in years of schooling	17.340 [8.64]***	10.095 [5.77]***	8.051 [4.78]***	7.995 [4.49]***	7.499 [4.07]***	7.427 [4.08]***	7.134 [3.97]***	7.111 [3.95]***	7.188 [3.99]***	7.535 [3.48]***
Highest parental education in years of schooling squared	-0.733 [8.33]***	-0.499 [6.64]***	-0.405 [5.66]***	-0.407 [5.36]***	-0.385 [4.93]***	-0.382 [4.93]***	-0.369 [4.82]***	-0.368 [4.79]***	-0.371 [4.83]***	-0.387 [4.15]***
Computer facilities at home ( <i>compome</i> )	12.051 [6.77]***	9.564 [5.67]***	7.478 [4.70]***	7.800 [5.16]***	7.890 [5.38]***	7.625 [5.31]***	7.384 [5.27]***	7.401 [5.26]***	7.335 [5.26]***	7.045 [4.35]***
No. books at home	0.056 [7.39]***	0.026 [3.77]***	0.027 [4.03]***	0.027 [4.08]***	0.027 [4.02]***	0.027 [4.00]***	0.027 [4.03]***	0.027 [4.04]***	0.027 [4.04]***	0.024 [3.73]***
Index of home possessions ( <i>homepos</i> )	14.702 [6.61]***	3.704 [1.97]*	3.099 [1.61]	2.376 [1.39]	2.207 [1.31]	2.458 [1.47]	2.481 [1.50]	2.456 [1.48]	2.486 [1.50]	2.549 [1.54]
Memorisation strategies ( <i>memor</i> )	-9.122 [5.01]***	-8.129 [4.63]***	-8.158 [4.89]***	-8.342 [5.28]***	-8.286 [5.36]***	-8.165 [5.31]***	-8.053 [5.21]***	-8.043 [5.20]***	-8.048 [5.22]***	-9.138 [4.99]***
Elaboration strategies ( <i>elab</i> )	0.905 [0.63]	3.638 [2.45]**	5.229 [3.84]***	5.234 [3.86]***	5.747 [4.37]***	5.611 [4.23]***	5.505 [4.12]***	5.559 [4.14]***	5.507 [4.09]***	6.102 [4.10]***
Competitive learning ( <i>complm</i> )	0.142 [0.08]	1.887 [1.18]	4.410 [2.90]***	4.561 [2.82]***	5.020 [3.23]***	5.088 [3.27]***	5.133 [3.34]***	5.149 [3.34]***	5.147 [3.34]***	5.278 [3.44]***
Co-operative learning ( <i>cooplm</i> )	-6.285 [4.92]***	-5.069 [4.92]***	-4.297 [4.06]***	-4.395 [4.04]***	-4.232 [3.81]***	-4.345 [3.89]***	-4.318 [3.85]***	-4.358 [3.89]***	-4.286 [3.81]***	-4.495 [3.77]***
High school oriented towards humanities ( <i>liceo classico</i> )		-27.064 [2.14]**	-3.831 [0.36]	-0.547 [0.05]	4.825 [0.46]	5.116 [0.49]	4.510 [0.45]	5.952 [0.59]	5.982 [0.60]	6.266 [0.58]
High school oriented towards sciences ( <i>liceo scientifico</i> )		2.939 [0.28]	27.627 [2.81]***	30.942 [3.24]***	39.864 [4.21]***	40.324 [4.31]***	39.936 [4.33]***	39.589 [4.28]***	40.511 [4.38]***	40.272 [4.00]***
Technical school ( <i>istituto tecnico</i> )		25.469 [3.40]***	25.104 [3.80]***	25.190 [3.47]***	20.419 [3.07]***	20.068 [3.01]***	19.081 [2.90]***	18.716 [2.85]***	19.488 [2.95]***	22.188 [3.20]***
Vocational school ( <i>istituto professionali</i> )		-6.467 [0.69]	-11.308 [1.49]	-14.158 [1.63]	-23.130 [2.82]***	-22.925 [2.81]***	-25.395 [3.27]***	-25.283 [3.22]***	-25.563 [3.26]***	-24.947 [2.99]***
Private school		-37.848 [3.26]***	-19.503 [1.50]	-19.645 [1.47]	-20.148 [1.48]	-21.148 [1.58]	-22.926 [1.71]*	-22.813 [1.70]*	-22.606 [1.68]*	-23.374 [1.57]
Proportion of computers connected to internet		17.177 [2.15]**	12.709 [1.95]*	14.221 [2.21]**	10.724 [1.73]*	11.805 [1.90]*	13.701 [2.14]**	13.107 [2.03]**	13.584 [2.12]**	14.519 [2.22]**
Teacher/student ratio		0.498 [1.50]	0.337 [0.99]	0.328 [0.92]	0.565 [1.23]	0.505 [1.13]	0.489 [1.05]	0.499 [1.06]	0.469 [1.00]	0.444 [0.95]
Highest parental education in yrs of schooling - schl average		9.548 [2.26]**	4.337 [1.54]	3.422 [1.17]	0.849 [0.32]	0.584 [0.21]	-0.447 [0.16]	-0.234 [0.08]	-0.297 [0.10]	-0.777 [0.27]
Highest parental occupational status - school average		0.619 [0.69]	0.927 [1.40]	1.130 [1.78]*	0.830 [1.29]	0.845 [1.33]	1.146 [1.66]	1.163 [1.72]*	1.122 [1.62]	1.256 [1.74]*
No. books at home - school average		0.211 [3.74]***	0.098 [2.21]**	0.075 [1.61]	0.063 [1.54]	0.055 [1.35]	0.031 [0.70]	0.022 [0.48]	0.028 [0.61]	0.029 [0.61]
Student-teacher relations at school ( <i>sturel</i> )		-5.700 [4.29]***	-4.557 [3.89]***	-4.676 [4.21]***	-4.613 [4.35]***	-4.708 [4.44]***	-4.693 [4.49]***	-4.682 [4.50]***	-4.683 [4.46]***	-3.989 [3.42]***
Disciplinary climate in maths lessons ( <i>disclim</i> )		6.825 [4.49]***	7.428 [5.83]***	7.585 [6.17]***	7.921 [6.83]***	7.870 [6.92]***	7.796 [6.69]***	7.820 [6.77]***	7.812 [6.68]***	6.740 [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 teacher ratio x 100	1.185	-0.021	0.410	-0.082	-0.163	-0.373	-0.281	-0.256	-0.114
	[3.98]***	[0.08]	[1.40]	[0.33]	[0.63]	[1.41]	[1.02]	[0.85]	[0.38]
% buildings unfit to school and improperly adapted to schools		-0.937	-0.891	-0.195	-0.180	-0.346	-0.446	-0.374	-0.526
		[3.48]***	[3.20]***	[0.79]	[0.72]	[1.30]	[1.68]*	[1.38]	[1.76]*
% school buildings rented for schooling		-0.877	-0.762	-0.482	-0.373	-0.277	-0.096	-0.211	-0.059
		[5.00]***	[3.96]***	[2.60]**	[2.03]**	[1.50]	[0.43]	[1.04]	[0.28]
Maintenance of buildings – 1st factor		-11.458	-9.499	-2.717	-3.936	-3.228	-2.270	-3.256	-3.462
		[5.74]***	[6.13]***	[1.42]	[2.03]**	[1.84]*	[1.34]	[1.90]*	[1.81]*
Log govern. exp. on capital account per student			16.748	11.603	11.268	10.696	10.561	11.006	10.251
			[2.98]***	[2.21]**	[2.08]**	[2.00]**	[1.97]*	[2.02]**	[1.73]*
Log govern. exp. on intermediate inputs per student			-31.993	-32.905	-25.810	-37.332	-40.261	-38.492	-40.590
			[2.00]**	[2.96]***	[2.30]**	[3.33]***	[3.73]***	[3.40]***	[3.52]***
Log govern.exp. on teachers per student			73.339						
			[1.73]*						
Class size (no. students)			3.092						
			[0.98]						
Employment probability (1-unempl.rate, at province level)				1.943	1.444	1.389	1.491	1.196	1.332
				[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 author per 100,000 inhab.				-0.008	-0.010	-0.005	-0.003	-0.004	-0.004
				[3.02]***	[3.27]***	[1.77]*	[1.03]	[1.26]	[1.03]
% without formal education (on pop > 6 years)					-3.192	-5.240	-4.630	-5.432	-6.024
					[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 population (2001)						-4.096	-3.398	-2.976	-2.479
						[1.99]*	[1.51]	[1.33]	[1.03]
social capital – 1st factor								1.391	1.760
								[0.85]	[0.96]
No. attempted suicides per per 100,000 inhab.							0.654		
							[1.29]		
No. volunteers in non-profit organisations per 100 inhab.							1.190		
							[1.19]		
Observations	10,894	8,468	8,468	8,410	8,410	8,410	8,410	8,410	7,045
R <sup>2</sup>	0.20	0.40	0.45	0.46	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%.

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

Area	Model									
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
	individual	+ school	+ infrastru ctures	+ expen ditures	+ labour market	+ cultural	+ demo graphic	+ social capital	(7) + social capital index	(9) only on 9 <sup>th</sup> grade students
North-West	-0.308 [0.04]	-1.048 [0.22]	-0.233 [0.04]	2.383 [0.40]	4.636 [0.67]	3.801 [0.57]	2.835 [0.45]	2.603 [0.42]	3.048 [0.51]	4.793 [0.73]
Center	-37.143 [5.04]***	-32.419 [5.44]***	-23.218 [4.00]***	-21.257 [3.79]***	-16.965 [2.66]***	-16.486 [2.43]**	-15.585 [2.46]**	-14.845 [2.39]**	-18.329 [2.69]***	-17.257 [2.25]**
South and Islands	-69.436 [9.51]***	-68.277 [11.63]***	-48.824 [5.76]***	-44.721 [4.78]***	-23.942 [1.82]*	-22.464 [1.61]	-15.221 [1.05]	-13.188 [0.90]	-17.689 [1.22]	-16.455 [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%.

**Table 4.** Models of students' mathematical literacy (Model (9), Table 2) with interactions between school variables and macro-area (North, Centre and South)

Variable	EPFs			means X		
	North	Centre	South	North	Centre	South
Small town (< 15,000)		1.599 [0.25]		0.18	0.07	0.12
City (> 100,000)		-2.824 [0.59]		0.34	0.31	0.22
Female		-22.721 [9.62]***		0.52	0.52	0.52
Age		8.345 [1.89]*		15.71	15.71	15.70
Single parent		-5.489 [2.07]**		0.15	0.17	0.14
Highest parental occupational status		0.331 [4.55]***		47.83	48.74	44.63
Highest parental education in years of schooling		7.166 [3.90]***		12.85	13.14	12.05
Highest parental education in years of schooling squared		-0.370 [4.77]***		177.10	185.08	159.69
Computer facilities at home ( <i>comphome</i> )		7.694 [5.59]***		0.04	-0.01	-0.30
No. books at home		0.027 [4.06]***		187.51	198.89	129.78
Index of home possessions ( <i>homepos</i> )		2.393 [1.52]		0.07	0.11	-0.22
Memorisation strategies ( <i>memor</i> )		-8.139 [5.28]***		-0.04	-0.02	0.14
Elaboration strategies ( <i>elab</i> )		5.639 [4.08]***		-0.11	-0.05	0.23
Competitive learning ( <i>complrn</i> )		5.321 [3.59]***		-0.11	-0.06	0.35
Co-operative learning ( <i>cooplrn</i> )		-4.020 [3.71]***		0.03	0.08	0.25
High school oriented towards humanities ( <i>liceo classico</i> )	25.302 [1.87]*	-100.63 [1.17]	-82.28 [1.04]	0.08	0.13	0.04
High school oriented towards science ( <i>liceo scientifico</i> )	51.613 [4.34]***	-72.973 [0.86]	-52.523 [0.70]	0.20	0.17	0.27
Technical school ( <i>istituto tecnico</i> )	28.003 [3.32]***	-94.654 [1.24]	-83.524 [1.17]	0.39	0.37	0.34
Vocational school ( <i>istituti professionali</i> )	-34.412 [3.62]***	-164.73 [2.24]**	-110.97 [1.62]	0.20	0.20	0.23
Other schools	-	-122.34 [1.58]	-82.575 [1.18]	0.13	0.14	0.12
Private school	-0.850 [0.06]	-50.451 [1.80]*	-38.303 [2.18]**	0.05	0.03	0.03
Proportion of computers connected to internet	4.209 [0.59]	20.021 [2.21]**	20.626 [1.97]*	0.70	0.62	0.75
Teacher/student ratio	1.211 [0.97]	-0.074 [0.28]	1.354 [2.57]**	9.09	10.04	10.64
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 ( <i>sturel</i> )	-4.475 [2.93]***	-5.772 [2.88]***	-5.137 [3.08]***	-0.44	-0.46	-0.14
Disciplinary climate in maths lessons ( <i>disclim</i> )	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 govern. exp. on capital account per student	4.596 [0.80]	11.33	10.99	10.75
Log govern. exp. on intermediate inputs per student	-22.310 [1.55]	14.22	14.21	14.12
Employment probability (1-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	-47136.50			

*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%.

**Table 5.** Accounting for territorial differences

	$(\bar{\mathbf{x}}'_N - \bar{\mathbf{x}}'_S)\boldsymbol{\beta}_N$ individual endowment effect	$(\bar{\mathbf{s}}'_N - \bar{\mathbf{s}}'_S)\boldsymbol{\gamma}_N$ school endowment effect	$\bar{\mathbf{s}}'_S(\boldsymbol{\gamma}_N - \boldsymbol{\gamma}_S)$ school effectiveness	$(\bar{\mathbf{q}}'_N - \bar{\mathbf{q}}'_S)\boldsymbol{\eta}_N$ local territory endowment effect	overall effect
$\bar{y}_N - \bar{y}_C$	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

## Appendix

**Table A1. Dataset summary statistics**

Variable	Obs.	Mean	Std.Dev.	Min	Max
<i>Individual information</i>					
PISA score mathematics (mean of 5 PVs)	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 ( <i>comphome</i> )	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 ( <i>homepos</i> )	11,548	-0.074	0.948	-3.787	1.939
Memorisation strategies ( <i>memor</i> )	11,495	0.032	0.881	-3.483	3.292
Elaboration strategies ( <i>elab</i> )	11,499	0.029	0.964	-3.262	3.263
Competitive learning ( <i>compplr</i> )	11,480	0.093	0.945	-2.844	2.450
Cooperative learning ( <i>cooplrn</i> )	11,483	0.137	0.974	-3.134	2.742
<i>School information</i>					
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 school - 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 ( <i>sturel</i> )	11,497	-0.301	0.947	-3.090	2.855
Disciplinary climate in maths lessons ( <i>disclim</i> )	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 ( <i>liceo classico</i> )	11,565	0.083	0.276	0.000	1.000
High school oriented towards science ( <i>liceo scientifico</i> )	11,565	0.214	0.410	0.000	1.000
Technical school ( <i>istituto tecnico</i> )	11,565	0.355	0.478	0.000	1.000
Vocational school ( <i>istituti professionali</i> )	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
<i>Territorial information (province level)</i>					
% 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 govern. 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 govern. 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
probability (1-unemployment rate, at 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.