Between-classes sorting within schools and test scores

An empirical analysis of the Italian junior secondary schools

Tommaso Agasisti [1,*] and Patrizia Falzetti [2]

1. Politecnico di Milano School of Management. Department of Management, Economics and Industrial Engineering, e. tommaso.agasisti@polimi.it * Corresponding author

2. INVALSI, e. patrizia.falzetti@invalsi.it

ABSTRACT.

This paper suggests that some Italian junior secondary schools are likely to practice sorting between classes, and proposes an indicator to measure this phenomenon. Adopting an appropriate Instrumental Variables (IV) approach, the impact of "informal" sorting on students' achievement is evaluated; the results suggest that this practice harms students' results in reading, as measured through standardized test scores. Heterogeneity of this effect is then explored, by considering different school types, as well as different characteristics of students. Overall, practising sorting within schools contributes to reproduce inequalities through unequal educational opportunities.

KEYWORDS.

Between-classes sorting, Instrumental Variables (IV), educational evaluation, equality

JEL CODES.

I24, I21, J24

Outline

Acknowledgments.

…

1. Introduction

The expressions “equality of opportunities” and “homogeneity” have been the keywords for the Italian educational system since the emanation of the Constitution Law in 1948. As a consequence, all the schools are subjected to the same national regulations, and have limited autonomy: thus, this system has always been supposed to produce quite homogenous results across schools, in terms of students’ achievement. Therefore, many observers were shocked to observe how different is nowadays the average student’s performance in the various Regions. Standardized test scores administered by the Italian National Evaluation Committee for Education (hereafter, INVALSI) have been extended to all primary and junior secondary schools only in 2007/08; before that moment, there was only anecdotal suspicion of profound differences in the students’ achievement level between the areas of the country, but limited empirical evidence of them. One of the latter was the data about fifteen years old students emerged from the OECD Programme for the International Student Assessment (PISA), in the 2009 edition, revealed that the average test score in Reading for the students in the North-West and North-East of the country was 511 and 504, above the OECD mean (conventionally set at 500), while the same figures for students in South and South/Isles were 468 and 456, respectively. On the PISA 2009 scale, 39 points is approximately the measure of one year of formal schooling, so the gap between the areas should be judged as huge. The results from the national INVALSI standardized tests are coherent with this picture: at grade 8, in the year 2012/13, the difference in literacy test score between the average student in the North and her counterpart in the South is 20 points on a [0;200] scale. Contextually, statistical analyses provided by INVALSI itself demonstrated that, even within the geographical macro-areas, there are significant differences between schools, and between classes in the same school: the figure 1 reveals that, despite around 80% of variance in students’ achievement is attributable to differences between students, nearly 15% of it id sue to structural differences between schools, and further 4% to differences between classes (INVALSI, 2013). Once these data have been analysed and discussed by politicians, institutions' leaders and academics, those actors claim the necessity to conduct deeper research on the effectiveness and equality of the country's educational systems.

In this context, understanding the determinants of students' performances at junior secondary school level[1] is a priority in Italy. Indeed, this particularly segment of the educational system is characterized by two peculiar features. The first is that this is a "weak link in the chain": the average level of skills and competencies, as measured by international tests, is pretty high at the end of primary schooling, and it declines rapidly during junior secondary schooling (see, for instance, the results of the 2011 edition of the study "TIMSS" – Trend in International Mathematics and Science Study, in Mullis *et al.*, 2012). The second reason is that at the end of junior secondary schools, students and their families must decide which track of upper secondary schooling to attend (academic, technical or vocational), and this choice has a very strong effect on subsequent educational opportunities and work life (Brunello & Checchi, 2007). The same way the choice of upper secondary schooling influences the educational and professional future of children, in the identical manner, junior secondary education has a strong impact on upper secondary one, and can prospectively reduce its dependence upon parental background, which still plays a too strong role as documented by Checchi & Flabbi (2007). In this perspective, past research about Italian junior secondary schools showed that their characteristics can be positively or negatively related to the probability of succeed (Mocetti, 2008); if it is the case, then, it is relevant to investigate those schools' organizational attitudes and characteristics that can improve or harm students' achievement.

In this paper, we focus on a practice that we define as "sorting" between classes, within the same school. The most similar concept that has been defined in the past literature is that of tracking as defined by Oakes (1986; 2005), who refer to it as "(…) dividing students into separate classes for high, average and low achievers" (Oakes, 1986; p. 13). The phenomenon studied here, thus, is tracking *within* schools and not tracking *between*

[1] The Italian educational system is currently organised in three sequential steps: primary schooling (grades 1-5, age 6-10), junior secondary schooling (grades 6-8, age 11-13), and high secondary schooling (grades 9-13, age 14-18).

schools, which is another important feature of the educational system studied by Brunello & Checchi (2007). The main difference between the two concepts is that, while tracking students into different classes is a decision made by the school, the existence of different educational tracks (i.e. between-schools tracks) is an institutional feature. In Italy, junior secondary schools (JSS) are not tracked into different streams, while upper secondary schools are (namely, the three tracks cited above exist: academic, technical and vocational); as a consequence, JSS are supposed to provide similar educational curricula, contents and experiences. Nevertheless, some recent literature challenges this assumption, and actually provides evidence that some schools are likely to produce internal tracking, which means sorting students between classes according to their characteristics (Ferrer-Esteban, 2011). The present research is inspired by this preliminary evidence, and provides two contributes: first, we argument that this type of between-classes sorting results in a segmentation based on students' socioeconomic background, and second we propose an indicator for empirically measuring if and how schools practice socioeconomic sorting between classes. After all, the research interest lies in understanding how diffuse is this practice among Italian junior secondary schools, and to assess its impact on students' performances. Thus, our main research question is: *does the school's practice of sorting students between classes have an effect on its students' academic achievement?* Also, we would also explore if such a potential effect is heterogeneous across different students and schools' characteristics. In the remainder of the paper, we use the words "tracking", "sorting", "grouping" and "segmentation" quite interchangeably, and these must be interpreted as referring to the school practice described above.

The paper is organised as follows. In the section 2, we discuss some related literature, present the institutional setting of the Italian educational system, and also introduce the theoretical background. Methodology and data are presented in section 3; a specific point is devoted to illustrate the Instrumental Variables (IV) approach adopted here for handling the endogeneity between the practice of sorting and test scores. Section 4 contains the results, plus provides sensitivity tests and extensions; lastly, section 5 discusses the implications and concludes.

2. Related literature, theoretical framework and background

2.1. How classes are composed: when (and how) sorting can arise

In the Italian educational system, the groups created through the composition of classes are of crucial importance for the experience of children. Indeed, students are assigned to a specific class for the entire length of an educational cycle (primary – 5 years, junior secondary – 3 years) and the classmates are the same for all subjects, curricular and extracurricular activities, etc.; moreover, neither between-classes formative exchanges are realised, nor formal interactions during the academic activities. As a consequence, and given the importance of peer contaminations, the process of composing the classes within a specific school, in the first year of an educational cycle, is dramatically relevant and likely to exert a strong effect on subsequent students' performances.

In terms of decision-making, the power of defining the criteria for the formation of the classes are in the hand of the School Board, which is representative of teachers, non-teaching staff, parents, local authorities, etc. The Italian Law only prescribes the minimum and maximum number for composing a class[2], even if exceptions are allowed for particular situations (for instance, schools in very disadvantaged areas, or located in the municipalities in small isles or mountains, or specific dispositions for the proportion of disabled students, etc.). The vast majority of schools decided that classes must be composed following the criterion of "equal-heterogeneity", that is to say by guaranteeing an equilibrium between classes of the same school concerning several dimensions such as the average academic ability of the students, their nationality (native *vs* immigrants), the presence of disabled students, etc. Such theoretical ideal is generally accepted by the

[2] In the year 2011/12, for junior secondary schools, the minimum/maximum numbers for composing a class were 18 and 27, respectively [Presidential Decree n. 81/2009].

educational community as a whole, and since a long time it is considered as the most coherent with the principles of Italian educational legislation (Poletto, 1992).

Despite the consensus about the criterion of "equal heterogeneity" of class compositions, several facts can occur that impede its concrete application. First, schools can deliberately choose not to follow this principle, explicitly or implicitly; in other words, schools can behave by grouping students between classes according to some dimensions (i.e. prior ability, socioeconomic status, nationality, etc.). Although this practice is contrary to the principles that inspire the Italian legislation about education, the practical difficulty of demonstrating this behaviour, and measuring it, makes it theoretically possible to schools operating this way; however, schools are not obliged to adhere to the equality principle, and can make different autonomous choices in this respect. Second, parents can exert pressures on school administrators and teachers to create classes, which are explicitly or implicitly sorted, for instance by replicating existing networks of friendships, relationships, etc. Third, it has been showed that Italian teachers decide to move from one school to another in search for easier conditions, i.e. when students' socioeconomic conditions are better (Barbieri *et al.*, 2011); it can be the case that similar dynamics happen within schools (not only between them) and in certain circumstances this would lead to between-classes socioeconomic sorting. Lastly, it can be the case that some external conditions of the school generate higher probability of sorting students, for instance because of a particularly high proportion of immigrants or disadvantaged students, with less chances of creating equal-heterogeneous groups across classes, so leading teachers to prefer segmented classes for educational purposes. Obviously, all these sources of influence can have an impact on the real, observed level of sorting between classes within each school; and given their very school-specific nature, it is logical to expect that the degree of sorting between schools is different across the schools in the country[3].

[3] Since the choice of sorting or not, and more generally the ability of managing the process of composing the classes, is a major task for the school principal, it could also be interesting to understand if and how sorting is affected by the principal's skills and practices; the work by Di Liberto *et al.* (2013) can be promising in this direction.

In this paper, we assume that some schools deliberately decide to sort their student between classes (or at least this is the leading option among those listed above), and consequently the observed level of sorting is the result of an *intentional* process, not casual. As there are no schools, which explicitly declare to sort students between classes, we refer to this phenomenon as an “informal” tracking (or sorting) process. To illustrate the concrete mechanism that we have in mind, it is useful to describe how schools operationally proceed in forming their classes; even though there is not a prescribed protocol that must be adopted by all schools, the most part of schools acts as follows. At the end of the scholastic year (t-1), some appointed teacher of each junior secondary school collect information about prospective students who enrolled to the school for the scholastic year (t). Such information is essentially based on the judgments expressed by teachers of primary schools, who probably mix together direct considerations about students’ academic results and indirect sketches about their socioeconomic background, to provide a context about how and where each student is growing up; as the sociology of education literature demonstrated extensively, these two dimensions (i.e. socioeconomic status and achievement) tend to be correlated, especially in the early stages of educational path (Haveman & Wolfe, 1995; Sirin, 2005). Also, primary school teachers’ evaluations can deal with more unobservable factors, such as behaviour at school, social skills, etc. In addition to the collection of this indirect information, when necessary or required junior secondary teachers and/or school principal can also meet the parents of prospective students, especially when specific requests or clarifications can be important. Once also this source is acquired, the group of selected teachers assigns students to future classes, based on the criteria of “equal-heterogeneity” (when established) and other considerations. This is the phase in which sorting can happen; indeed now teachers can decide to sort students according to some observable (i.e. grades at the end of primary schools) or unobservable (i.e. students’ behaviour) characteristics.

The main idea proposed in this paper is that, when such sorting happens, it ends in a socioeconomic sorting; indeed, either done by ability or by unobservable factors such as behaviour, these factors are correlated with students’ socioeconomic background. The relationships between socioeconomic status and achievement, and their effect on sorting, are well documented in the sociological perspective in the US, where practices like "ability grouping" are widespread: "(…) one of the most widely documented patterns in sociological research on skill grouping is that low-SES and non-Asian minority students are disproportionately placed into groups for lower-skilled students while high-SES, white and Asian students are placed higher" (Condron, 2007; p. 142). Overall, the practice of sorting students usually tends to reproduce social inequality: "Through tracking, schools continue to replicate existing inequality along lines of race and social class and contribute to the intergenerational transmission of social and economic inequality" (Oakes, 2008; p. 705).[4]

This study proposes a quantitative indicator for measuring sorting between classes (within each school) that is based on the values of an index that captures the socioeconomic condition of each student (see section "Methodology"). As can be easily understood from the description of the process for composing classes, much endogeneity can arise between sorting and students' academic results, consequently in the empirical analysis this eventuality is properly taken into account by means of an instrumental variables approach (see, again, the section 3).

2.2. Why is sorting potentially affecting students' performances?

In this paragraph, we argue that sorting students between classes is likely to have some negative/positive effects on their results. The starting point for our theoretical framework is the well-known relationship between a student's achievement level and her family socioeconomic status (SES). As pointed out by Haveman & Wolfe (1995) in their review, SES exerts both direct and indirect influences on students' performances, through economic (i.e. income), social (i.e. education) and psychological (i.e. motivation) channels. Overall, the literature consistently shows that "(…) parents' location in the socioeconomic structure has a strong impact on students' academic achievement" (Sirin,

[4] Nevertheless, for empirical purposes and with the aim of testing this major assumption, the annex C discusses the results obtained when assuming that sorting is along ability and not socioeconomic status.

2005; p. 438). This research takes this influence as granted, and uses an indicator about parental SES as a control in the empirical analysis, which is conducted through the specification of an educational production function (EPF).

On a different perspective, the presence of peer effects between classmates is more relevant for policy and managerial purposes. Indeed, previous research showed that not only individual-level SES is a factor associated with academic performance, but also the socioeconomic background of the peers (classmates and schoolmates) have an *independent* effect on a student's academic experience and results (Ammermueller & Pischke, 2009). In other words, attending a school and/or being part of a class with a low/high average socioeconomic level can have a negative/positive effect on the educational results, on top of the impact of own SES. The mechanism through which this peer effect related to SES acts is still not completely clear, and complex altogether; nevertheless, its important role has been acknowledged by the literature since a long time (for a description of the main theoretical arguments, see Caldas & Bankston, 1997 and Schneider, 2013; Perry & McConney, 2010 provide international evidence about the effects related to school-average SES)[5]. Therefore, our study does not deal with the effects of attending a school with a particularly high or low average SES level; to properly consider its impact, a measure of class-average SES is included as a covariate in the empirical model.

Another stream of partially related literature is that about between-schools segregation. International evidence exists that students tend to be sorted (or parents tend to sort their children) between schools according to some of their observable characteristics, among which the socioeconomic background stands as one of the main factors. Despite the fact that "(…) there is very little evidence internationally that having pupils with similar characteristics clustered in the same schools produce any improvement in overall levels

[5] There is a significant bulk of recent empirical evidence that shows and measures the relevance of peer effects on educational and other social outcomes. See, for instance, Gaviria & Raphael (2001) on peer group influences with respect to an array of behaviours such as drug use, smoking, alcohol abuse, drop out school, etc.), and Vardardottir (2013) who reports an empirical study in the Iceland setting, where peer effects seem to be the main factor explaining differences between high and low-ability classes. Also, peer effects are related to the average level of ability in the classroom, as Kang (2007) demonstrated in an international perspective, using TIMSS data.

of attainment" (Gorard & Cheng, 2011; p. 328) this pattern of segmentation between schools still exists in many countries. However, this paper is not primarily focused on this topic, and we limit ourselves to include some school-level variables to account for between-schools stratification, because we are interested in within-schools variations.

Lastly, this paper does not directly addresses the idea of within-school competition, which has been proposed as a key for reading some sorting phenomena within schools (see Adnett & Davies, 2005, and Borland *et al.* 2006): indeed, in Italy students are sorted between fixed classes by the school, so there are not phenomena of internal segmentation due to the students' choice of following one or more specific subjects/groups; instead, it may be the case that effects of this kind – that is, students being exposed to different schooling quality – arise because of sorting practiced by the school.

Our main point of attention is the effect, for the generic ith student, to attend a school that practices between-classes sorting. This average effect is surely dependent upon the distribution of high-ability and low-ability students into the school. Consider a student a_i, whose ability (as measured by prior achievement score $Y_{(a)(t-1)}$) is high, as well as her socioeconomic status (as measured by an indicator called SES). If the school that she attends is practising sorting, she will be placed in a class (A) where $avg(SES_A)$ is higher than in the second class B, attended by the student b_i, whose ability and SES is lower; thus, $avg(SES_A)>avg(SES_B)$. The reasons listed above suggest that peer effects play a role in influencing the level of (current) performance of the students in a class, so that $Y_{at}=f(Y_{a(t-1)};SES_A)$ and $Y_{bt}=f(Y_{b(t-1)};SES_B)$. If the low socioeconomic conditions of the class B exert a negative effect on Y_b, and the high average SES of the class A positively influences Y_a instead, then the "net" effect on the generic ith student is due to the relative impact of the forces acting in different directions. It is remarkable that what is interesting here is not only the specific impact of class-level SES on the achievement of the high or low ability student, but the average impact of practising sorting on the generic student. There are good reasons to believe that such an effect is independent by the class-specific effect exerted by the average SES in the classroom. For instance, it can be the case that segmentation between classes in the same school harms the overall school climate, by reducing teachers' propensity to collaborate and share good practices. Also, these schools could experience worse between-students relationships, especially if they perceive that classes have been created according to some of their characteristics (and most notably their family background). The interactions between these dimensions can play a role, too; for example, less motivated or less skilled teachers could be assigned to more complicated classes, and this would amply the differences between classes' academic results. In addition, it can happen that classes with more disadvantaged students obtain different instructional stimulus, such as partly different curricula, less extracurricular activities, less quality and/or quantity of resources (this mechanism is at the heart of the critic promoted by Oakes, 2005 in the US context). Lastly, as suggested above, the negative effects on disadvantaged students can simply overcome the positive ones on better-off individuals – albeit the usual justification for this practice is that teaching more internally homogenous classes is helpful for targeting students with different characteristics – so generating a negative net "average" effect on the student population attending the school.

In the next sections, we propose an indicator for measuring the extent to which schools practice between-classes sorting, and evaluate the impact of this indicator on students' performances. It is important to recall that in what follows, we do not distinguish between "based-on-ability" sorting and socioeconomic sorting, by having assumed that whatever the reasons and practices behind sorting, they conduct to a socioeconomic sorting.

3. Methodology and data

3.1. Measuring socioeconomic sorting: introducing ESCS and ESCS_Var_Within

After having assumed that intentional behaviour of schools can lead to socioeconomic sorting between classes, an important methodological step is to measure this phenomenon. For this purpose, we use an indicator that measures the socioeconomic background of students, in analogy with the OECD indicator of Economic, Social and Cultural Status (ESCS) (for a detailed description of how OECD calculates it, see OECD, 2009; for the application in the Italian context, see Campodifiori *et al.*, 2010). The variable takes into account several dimensions, among which parents' occupational status and education, as well as home possessions (goods, books, etc.). By construction, the mean is set equal to zero, and standard deviation equal to one. The figure 2 reports the distribution of ESCS for the over 450,000 grade 6 students analysed in this work. It is worth noting that, while following the general pattern of a normal distribution, there are several peaks around some unit values, due to the particular way of calculate the student-specific value taken by the indicator ESCS.

With the aim of deriving a measure of between-classes socioeconomic sorting within each school, the variance of ESCS that is accounted by statistical differences between classes ($ESCS\_Var\_Within_k$) is calculated, through the following formula (which reproduces the decomposition of ESCS variance between classes):

$$ESCS\_Var\_Within_k = \frac{1}{N_k}\sum(ESCS_{jk} - \overline{ESCS_k})^2 \qquad (1)$$

where N_k is the number of students in the kth school, $ESCS_{jk}$ is the average ESCS of the jth class in that school, and $ESCS_k$ is the kth school's average ESCS. This measure is the main focus of this paper, and in the next steps its impact on students' performances is estimated. The values taken by this variable can be interpreted as percentages, i.e. which proportion of variance between students' socioeconomic status can be attributed to structural differences in the composition of classes within each school, as measured at school level. The figure 3 contains a graphical illustration of the distribution of $ESCS\_Var\_Within_k$; the table 1, instead, tabulates the value it assumed in the distribution by percentiles. The mean value is around 8.5% (with a standard deviation of 7.9, and a median around 6.4%). The most part of the distribution is between 3% and 11%, and more than 80% of the schools have a value that is below 15%. Around 300 schools seem to heavily sort their students, with the indicator assuming values >75%; it can be the case that measurement errors or peculiar distributions of students drive this figure; in the results, we check how sensitive are the empirical analyses to the exclusion of this group of institutes (to anticipate the finding, the results are not affected at all).

Overall, it seems that only few Italian schools do sort their students by socioeconomic status, and also with different intensity among them: as evidenced by exploring the dataset, the number of schools for which the variance between classes in ESCS composition is higher 50% is definitely low. Nonetheless, it seems important to study the effect of this particular school's feature on the performances of their students, in those cases where sorting of this kind does actually occur.

3.2. Estimating the impact of ESCS_Var_Within on students' performances

An educational production function (EPF) is specified, in a value added (VA) fashion; considering i = 1,…, I students, j = 1, … J classes, k = 1, …K schools, at the period t, the EPF is mathematically expressed as:

$$Y_{ijkt} = \alpha_0 + \lambda Y_{ijk(t-1)} + \alpha_1 X_{1ijkt} + \alpha_2 X_{2jkt} + \alpha_3 X_{3kt} + \varepsilon_{ijkt} \qquad (2)$$

where X_1 is a vector of student-level variables, X_2 is a vector of class-specific characteristics, and X_3 is a vector of school-level features; $Y_{ijk(t-1)}$ controls for prior achievement (i.e. test score at grade 5), and α_0, α_1, α_2 and α_3 are vectors of parameters to be estimated. Robust standard errors are clustered at school-level, which is the highest level of data aggregation in our dataset[6]. One noteworthy characteristic of the EPF in (2) is that it does not only include school-level covariates, but it also tries to correctly specify the determinants of student achievement at class level – where the most of instructional activity really happens.

The variable of interest, that is ESCS_Var_Within, is included among the variables at school level (vector X_3). A specific discussion is worth regarding this variable, due to its likely endogeneity that must be controlled for. Indeed, as discussed in the section about 2,

[6] Although we strongly believe to the opportunity of clustering S.E. at school level, we also tested how the results change when clustering at class-level, and we did not found any sensible difference.

it is certainly the case that some schools may decide to sort student between classes according to their prior achievement. The inclusion of $Y_{ijk(t-1)}$ among the covariates avoid the problem of endogeneity of ESCS_Var_Within to the current level of performance: if socioeconomic segmentation is actually driven by its correlation with performance, it should be controlled for. However, it can also be the case that ESCS_Var_Within is instead driven by different (unobservable) phenomena, such as students' behaviour, or the teachers or principals' impressions from discussions with students' parents before composing the classes, or the pressure exerted by parents for getting their children in some specific classes, etc. If such events occur, ESCS_Var_Within can be the result of these (endogenous) forces, and not of a random process; these underlying unobservable variables then have an impact on test scores, which is masked by the value of ESCS_Var_Within itself. In order to deal with this potential endogeneity, an Instrumental Variables (IV) approach is proposed. As an instrument, we followed the intuition of Collins & Gan (2013) in using the value of between-classes segmentation for another grade in the same school. The crucial assumption is that schools voluntarily choose the level to which sorting students according to their socioeconomic status; if it is the case, then such process should be present also in other grades, and using the value of segmentation in that context would be a good instrument[7] – as surely related to the variable of interest, and by definition unrelated to the target population of students, as they are different ones. Operationally, the choice in this paper is to use the value of ESCS_Var_Within in the same school the year before that used in the empirical analysis; in other words, $\text{ESCS\_Var\_Within}_{(t-1)}$ has been computed for the year 2010/11, and then used as an instrument for $\text{ESCS\_Var\_Within}_{(t)}$ (year 2011/12). A graphical illustration of the relationship between the variable of interest ($\text{ESCS\_Var\_Within}_{k(t)}$) and the instrument ($\text{ESCS\_Var\_Within}_{k(t-1)}$) is provided in the figure 4; their pairwise correlation is 0.27 and is statistically significant at 1% level.

[7] To the extent that the school changed policy from one year to another, the instrument does not capture the phenomenon of interest. For instance, individuating if there was a school principal's turnover could be an (indirect) indicator of it; unfortunately we do not have available data to check for this eventuality. Nevertheless, the practice of sorting is much likely to be fruit of the agreement of the wider community of teachers, and not only a school principal's decision, so it is probably more persistent over time.

Mathematically, we estimated a two-stage equation; the first stage provides estimates of the variable of interest, ESCS_Var_Within$_{(t)}$ and is computed as follows:

$$ESCS\_Var\_Within_{(t)} = \beta_0 + \delta Y_{ijk(t-1)} + \beta_1 X_{1ijkt} + \beta_2 X_{2jkt} + \beta_3 X_{3kt} + \eta Z_k + \varepsilon_{1ijkt} \qquad (3)$$

where the vectors of variables are as in the equation (2); Z_k is the instrument (ESCS_Var_Within $_{(t-1)}$) that acts as the exclusion restriction[8]; η is the instrument's parameter to be estimated to judge empirically the quality of the instrument. The predicted values of the variable of interest, ESCS_Var_Within$_{(t)}$ are then used in the equation (2) for obtaining reliable estimates of its impact on the output Y_{ijkt}. Annex A reports the results from the first-stage regression, for both using Reading and Mathematics as the output. As can be easily seen, the instrument is well correlated with the variable of interest, and its z-value is high (around |8|)[9].

3.3. Data

In this paper, much information about students' performances, and their classes and schools' characteristics, is available. All data refers to all grade 6 students (almost 500,000) who took the standardized test in the year 2011/12; students in grade 6 are those who are enrolled in the first year of a junior secondary school. The original dataset comes from INVALSI, and contains some variables, which are collected in collaboration with the Ministry of Education. The variables at disposal for the research can be classified in three groups.

The first group contains student-level information: gender, immigrant status (Italian, first-generation or second-generation immigrant), age (students who went to school one year before the suggested age are called "early enrolled", while those who entered the school

[8] For a methodological description of the IV functioning, see Angrist & Piscke (2009).

[9] Albeit high in magnitude, and highly statistically significant, the value is slightly lower than |10| which is the target suggested by Staiger & Stock (1997) to fully guarantee the (empirical) reliability of the instrument.

one year after[10] – or repeated one or more years, are called “late enrolled”), family structure (students who live with both parents or not, students with siblings or not). Moreover, the indicator about the Economic, Social and Cultural Status (ESCS) of each student is calculated (see section 3.1). Lastly, for the first time in the Italian context, data about prior achievement (at grade 5) has been tracked at the level of individual student, so allowing a value-added modelling of the EPF. Unfortunately, the matching procedure encountered some technical and administrative problems, so this information is actually available only for a subsample of students, namely 47% of students for scores in Reading, and 53% for scores in Mathematics. In the Annex B, some characteristics of the subsample of students are reported and compared with those of the entire population. Despite some slight differences, the sample must be considered as generally representative, thus the empirical analysis is conducted on the sample; such strategy is preferable because, given that prior achievement is the strongest predictor of current performance level, excluding this variable would create severe problems of omission bias. In the Annex B, it is also reported a comparison between the main results obtained in the paper and those obtained when excluding prior achievement and focusing on the entire population of students: the results are qualitatively very similar, but the goodness-of-fit of the model is considerably lower.

The second group of variables aims at measuring the main characteristics of the classes attended by the students. For this purpose, the class-average ESCS is computed, as the mean of the indicator ESCS calculated for each student. Moreover, the following indicators were derived: proportion of females, first-generation and second-generation, early and late-enrolled, and disabled students[11]. The number of students who compose the classroom controls for size effects; the proportion of students who took the test controls for potential strategic behaviours such as inviting worse students to stay at home, or for

[10] It is quite frequent that second-generation immigrants are enrolled in one or two lower grades than their age, as for facilitating their educational progress (limiting the problems with the Italian language and/or compensating the inadequate instruction that they received in their country of origin).

[11] What is remarkable here is that, while we have not data about disabled students’ achievement (as they do not took the INVALSI test, or when taken it is not included in the data), we control for the proportion of disabled students in the class to check whether spillover effects exist – potentially related to different educational activities and strategies in presence of higher proportions.

particularly unusual situations (high absence rates the day of the test). A dummy is included for those classes that can be classified as *"tempo pieno"* (full-time); those are classes which schedule is organized in entire days (8am-4pm usually) instead that only in the mornings. Two variables that deserve specific attention are those that face a particular drawback of the Italian system of standardized testing. Indeed, since the very beginning of INVALSI exercise of administering the standardized tests, many schools adopted "cheating" behaviours, through suggesting to the students the right answers. The real reasons of this behaviours are still not clear, but are probably related to the fear that test scores can be used in the future for accountability purposes (even if the Italian Government always clarifies that this is not the aim and the perspective of the INVALSI work). Nevertheless, cheating phenomenon is problematic and significantly affects the reliability of test scores (Bertoni *et al.*, 2013). Therefore, a statistically based solution has been developed by INVALSI to purge the data from this problem. As there is a representative sample of classes and schools that takes the test in a controlled setting (with specifically-trained external examiners), the statistical properties of the distribution of answers are used to calculate an indicator for all the Italian classes called "cheating propensity", which represents the probability that the students of that specific ith class were influenced by cheating behaviours. This indicator is then included in the empirical analysis, together with a dummy indicating if the class was selected as part of the sample of classes that took the test in the controlled setting.

The third and last group of variables refer to schools' characteristics. Specifically, dummies control for the geographical macroarea in which the school is located (North_West, North_East, Central Italy, and South/Isles): given that there is a huge gap between average achievement levels of the different areas (see, for instance, Bratti *et al.*, 2007; Agasisti & Vittadini, 2012), the inclusion of these dummies is essential in the Italian context, and adds much reliability to the overall results of the empirical analysis. The number of students controls for size effects; the inclusion of the number of classes acts both (i) for reducing the dependence of the target variable of interest (ESCS_Var_Within) to the number of classes, and (ii) for controlling for eventual school's practices of keeping particularly small/large classes (also another variable, measuring the average number of students per class, serves this purpose). A dummy variable indicates if the school is public or private; another dummy if the school includes a primary and secondary school *(Istituto comprensivo)*, or is a standing-alone junior secondary school. Lastly, a dummy concerns whether the school is located in one of the four so called PON Italian Regions (all of them are poor Regions in the South, and namely: *Basilicata, Calabria, Campania, Puglia, Sardegna e Sicilia*), which received European funds in the last years for developing specific projects with the objective of improving their students' educational levels.

Descriptive statistics for the whole population of students are reported in the table 2. The test scores are reported as a percentage of correct answers, so they range between [0;100]. In 2011/12, the average score in reading for 6th graders was approximately 65 points, and 46 that in mathematics. The analogous figures for those students when they were 5th graders, in 2010/11, were 74 and 70, respectively; such higher values are justified because tests at primary schools are easier. Around 10% of students are immigrants, quite equally distributed between first and second-generation immigrants. Early enrolled students represent around 2.2% of the whole sample, while late-enrolled ones (many of whom are immigrant students) around 7.3%. The students who do not live with both parents, in a traditional family, are 13.7% of the total; and those without siblings count for the 16% of the population. Turning to the classroom-level variables, the estimated cheating propensity is 7%, but with a wide variation (standard deviation is around 19%). The values of class-average ESCS are coherent with the student-level distribution, but with much less variation (standard deviation is less than half of that among students, and equal to 0.5). The proportion of students who took the test is high, and close to 95% of the total. The number of students in the class is around 23 (a glance to schools' characteristics reveals that in all the country the average number of students per class is indeed 22.6); as explained above, there are some national regulations that constrain the class size within a quite narrow range. The average school size is about 150 students, and – given the number of students per class – the average number of classes is 6.5. Lastly, the geographical distribution of schools is such as 43% in the North, 18% in Central Italy and 39% in the South. Private schools represent only 3% of the entire population.

3.4. Some preliminary insights about the data

Before showing the results of the econometric analysis, it is interesting to report some further descriptive data concerning the statistical correlations between the variable of interest, ESCS_Var_Within, and two dimensions of students' performances, namely the (school-average) score in the standardized tests, and the "dispersion" within each school (variance of test scores, measured at school level). Following the intuition that the sorting phenomenon can be differentiated across different Italian geographical areas, the correlation is reported both as calculated for the overall country and separately by Region (the paper explored this geographical heterogeneity more in detail later, see the section 4.2). The results, as contained in the table 3, can be summarized in two main evidences. The first is that there is a negative correlation between $ESCS\_Var\_Within_k$ and schools' average scores, and a positive one between the variable and the dispersion of scores within schools. In other words, sorting practices seem to be statistically related to lower scores and higher inequality between students; albeit these relationships are statistically significant at 1% level, they are only pairwise correlations, that must be tested through appropriate econometric techniques. The second evidence is that these statistical correlations are stronger in the regions of Southern Italy, suggesting that the practice of sorting is more common between schools in the South, or that the influence of this practice on students' results is stronger for these schools than for those operating in the North. Also this preliminary insight has been tested through the empirical analysis, which results are contained in the next sections. To give a graphical illustration of the North/South gap in the statistical (negative) correlation between ESCS_Var_Within and schools' average performance, we included the information in the figure 5, panel b – together with a picture of the differences in achievement scores, panel a – which is a map of Italy, by Region; data refer to reading scores. In the panel b, the colour represents the intensity of the relationship as measured through the coefficient (the lighter the colour, the higher the coefficient), while the borders are ticker for those Regions where the correlation is statistically significant. The difference between the two areas of the country is clear, and it raises concerns about the overall equity of the educational system: students in the South not only obtain lower average scores, but also are more (negatively) affected by sorting practice.

Another aspect that is worth of attention is that sorting students between classes is not a characteristic of schools which are particularly rich or poor (on average); the figure 6 reports the distribution of ESCS for all the schools in the population, and that of the sample of schools for which there is a “high” measured sorting between-classes (namely $ESCS\_Var\_Within_k$>15%). As can be noted, there are not sensible differences in the distribution, and in the subgroup of sorting schools there are both schools serving advantaged and disadvantaged students. To further consider this eventuality, we also calculated pairwise correlations between $ESCS\_Var\_Within_k$, the average ESCS at school level, and another indirect measure of school’s average SES, namely the proportion of immigrants (here the measure is the sum of first and second generation immigrants): these numbers are all low (<|0.11|) and negative, suggesting that schools that practice sorting are not characterized by particular student populations.

Lastly, a topic that deserves attention deals with the potential relationship between the focal variable, $ESCS\_Var\_Within_k$, and the number of classes of the kth school. Indeed, it can be the case that the way the variable is measures is affected by the number of classes and their dimension of each school; if proved, then our empirical analysis can actually capture dimensional factors more than causal effects of sorting. To check this possible relationship, the figure 7 plots the two variables; the pairwise correlation between the two is around 0.24, but the figure seems suggesting that no meaningful positive correlation exists. Thus, $ESCS\_Var\_Within_k$ measures a phenomenon that is not simply the higher (casual) probability that classes are more socioeconomically segmented when they are more numerous, but an independent attitude at voluntarily practising sorting. While it is not possible to rule completely out the possibility that the correlation between the two variables can disturb the empirical analysis in some way, it does not seem to introduce serious biases in the main results.

4. Results

4.1. Baseline results

The main baseline results of the empirical analysis are in the table 4, which has been organised according to the three groups of variables. Student-level indicators are all statistically significant, and have the expected signs. Prior achievement is the strongest predictor of current performance in test scores; its coefficient tells that the impact is around 0.5 standard deviations (s.d.). Female students perform better on reading, and male on mathematics. Immigrant students perform worse than their Italian counterparts; the magnitude of this gap is similar for first and second generation immigrants, and equal to around 0.05 s.d. in reading, much less in mathematics. Early-enrolled students have performances quantitatively similar to regular classmates, while those students who were classified as late-enrolled lag behind (-0.04 s.d. in reading, -0.02 s.d. in mathematics). Even after having controlled for prior achievement (which partly captures a student's background), the indicator about current economic, social and cultural status (ESCS) still has a positive correlation with the results in the test scores (0.15 and 0.13 s.d., respectively for reading and mathematics). Students who do not live with their parents have a disadvantage in test scores (-0.02 s.d.), as well as those with siblings (but only when considering reading, -0.014 s.d.). Among the class-level factors, some indicators reflecting their composition do contribute to explain students' performances: among them, the average class' ESCS is positively related to test scores (0.01 s.d.); the proportion of late-enrolled students negatively (-0.01 s.d.); the number of students in the class instead is associated with better results (possibly suggesting the idea that there is the necessity of a critical mass to exploit positive peer effects). The proportion of students who took the test is positively related to the score; the students attending more motivated and participating classes are then likely to obtain benefits from it. Lastly, among the school-level variables, the well-known gap between geographical macro-areas is confirmed by our analysis, with students attending a school in the South obtaining, on average, more than 1 point less (0.07 s.d.) than their counterparts in the North, all else equal. The schools in PON area apparently did not benefit of the European resources for restructuring their educational systems, at least the coefficient for reading is negative and statistically significant; an alternative explanation is that being located in such a disadvantaged area is not (still) compensated by investments in the last years.

Moving out attention to the variable of interest, ESCS_Var_Within, it turns out to exert a negative impact on reading test score, while having no effect on mathematics. The magnitude is substantial (-0.05 s.d.): an increase of 8% of ESCS variance between classes diminishes the student's achievement in reading of almost 1 point (or 1.2% when calculated at the mean score observed in the population)[12].

Summing up, the empirical analysis suggests that, after having controlled for students, classes and schools' characteristics, attending a school, which deliberately practices between-classes sorting, has a negative effect on test score in reading. This effect is sizeable in magnitude, and statistically significant. The next section extends this analysis, looking for potential heterogeneity of this effect, and its possible different impact on different students' profiles.

4.2. Extensions: heterogeneity and further analyses

The first additional analysis considers the effect of ESCS_Var_Within as possibly heterogeneous (i) between public and private schools, and (ii) across geographical macro-areas. To explore this possibility, we ran the equation (3) on the reference subpopulations of students; the results of this exercise are reported in the table 5, panel A. What emerges is in line with expectations from descriptive statistics. As the phenomenon of segmenting between-classes is more frequent among public schools than among private ones, its

[12] We also checked if the empirical analysis is sensitive to the exclusion of the group of schools for which ESCS_Var_Within is particularly (maybe too) high, specifically >75%. The results are almost identical; the estimated impact of ESCS_Var_Within on reading score is 0.049 (instead of 0.050) and statistically significant. Also the other coefficients of the educational production function are virtually unchanged.

(negative) effect is concentrated among the former; the coefficient mirrors the one estimated for the whole student population (-0.05 s.d.). Similar considerations can be formulated for the case of heterogeneity between schools located in Northern, Central or Southern Italy. The negative effect of ESCS_Var_Within has been found only for schools operating in the South, and in this case the magnitude of the effect is bigger than average: -0.071 s.d. – in other words, an increase of 9% in ESCS_Var_Within (1 s.d.) causes a decrease in the reading score of around 1.2 points, or almost 2% of the mean average score[13]). This particularly negative effects on students in the South raises further and serious worries about the equality of the Italian educational system; indeed, they are not only unprivileged because of lower educational results, but they are also more likely to be harmed by between-schools segmentation practices adopted by schools.

After having explored heterogeneity, we tried to understand more about the effect of ESCS_Var_Within by looking at its interactions with students' own prior achievement and ESCS. Conceptually, the aim is to understand whether attending a school that practices between-classes segmentation is more harmful/beneficial for students with high/low ability and high/low socioeconomic background. Mathematically, we estimate the following:

$$\begin{aligned} &Y_{ijkt} = \alpha_0 + \lambda Y_{ijk(t-1)} + \alpha_1 X_{1ijkt} + \alpha_2 X_{2jkt} + \alpha_3 X_{3kt} + \beta_1 W_{(z)ijkt} + \varepsilon_{ijkt} \\ &W_{(1)} = ESCS\_Var\_Within_k \\ &W_{(2)} = ESCS\_Var\_Within_k \times Y_{ijk(t-1)} \\ &W_{(3)} = ESCS\_Var\_Within_k \times ESCS_{ijkt} \end{aligned} \tag{4}$$

where each of the three interaction terms, containing ESCS_Var_Within, is instrumented (ESCS_Var_Within$_{(t-1)}$ has been used for this scope[14]). The results are in the table 5, panel B; many interesting points should be underlined here. The first evidence, looking at

[13] The magnitude of this effect has been calculated by consider the specific values of (standard deviations of) reading scores and ESCS_Var_Within for the students attending schools in Southern Italy (16.7 and 9.2, respectively).

[14] In this circumstance, the t-values associated with the instruments in the first stage are very high, well above the threshold suggested by Staiger & Stock (1997).

the test scores in reading, is that the previous finding about the negative effect of ESCS_Var_Within is confirmed; therefore, the estimated magnitude is higher (0.089 s.d., that is increasing the variable of 8% has a negative effect of almost 1.5 points, or 2.2% when calculated at the performance mean). The coefficient for the interaction with ESCS is positive, suggesting that better-off students can benefit from between-classes segmentation; it would be interesting to understand more of this phenomenon, but one potential explanation is more advantaged students are able to obtain “positional” gains from the within-school distribution of classes. At the same time, the coefficient for the interaction with prior achievement is negative; in other words, students with higher levels of ability are more penalised by attending a school that practices between-classes sorting. This finding is coherent with the idea that the main factor behind within-school segmentation (when present) is more likely to be the socioeconomic background than ability. A particularly relevant result is also that the same patterns have been detected when considering mathematics score as output. In this case, the coefficient for ESCS_Var_Within is positive, suggesting that the average student can obtain benefits in mathematics (but not reading) when attending a school that practices between-classes sorting; this difference between subjects can be justified through the differences in pedagogy, and can suggest that sorting could have positive/negative effects on one subject, but not the other. However, the coefficients for the interaction terms are coherent with those for reading, so they can also be interpreted as revealing between-classes segmentation in action, based on SES more than ability.

As a further step, we checked whether the effect of ESCS_Var_Within is particularly high or low at the tails of its distribution. For this purpose, we estimated our baseline model for different subsamples of students who attend schools characterized by a value of ESCS_Var_Within at the 25^{th} and 75^{th} percentile of its distribution (and also in the middle of the distribution itself, excluding the tails). The results are in table 5, panel C, and seem to indicate that there is some relationship with the intensity of sorting (the higher the variance of socioeconomic status between classes, the strongest the negative effect); however, none of these estimates are statistically significant, so we do not draw any specific robust indication from them.

We also tried to understand more about the mechanism that is driven the results, in other words if the students who are attending classes with a higher ESCS (within schools that practice sorting) are also obtaining higher test scores (this would be also an indication of peer effects in action). Proceeding in this direction, we included the interaction between ESCS_Var_Within and the average ESCS measured at class level (class_avg_ESCS) in the estimation of the EPF; as an instrument for this interaction term, we used the interaction between ESCS_Var_Within$_{(t-1)}$ and class_avg_ESCS. The results are presented in the table 5, panel D, and suggest an interesting story. The effect of attending a class with higher average ESCS, in a school where there is between-classes sorting, actually produces positive effects, as demonstrated by the positive and statistically significant coefficient of ESCS_Var_Within*class_avg_ESCS; the magnitude of this impact is around 0.07 s.d for reading (about 1 point in the INVALSI scale [0;100]). Interestingly, the effect is statistically existent also for mathematics, and not only reading, even if the magnitude is somewhat lower (0.05 s.d., <1 INVALSI point). However, the overall effect at school level is negative (for reading), and the magnitude of this impact is higher than that estimated in the baseline model (around -0.06 s.d.): this means that students in disadvantaged classes obtain performances that are as low as to neutralize the positive effects for better-off students. If it is the case, the mechanism is definitely perverse: not only within-school sorting is ineffective (the overall effect on students' performances is negative), but also it contributes to reproduce inequalities by increasing the achievement gaps between disadvantaged and advantaged pupils.

A final remark is about the potential collinearity of the variable of interest with other factors acting at geographical level; for instance, Ferrer-Esteban (2011) suggests that sorting phenomena are frequent in certain Italian Provinces more than in others. The reasons behind this geographical disparity in the phenomenon are not clear, but it should be controlled for; the variables included in the baseline version of the model (i.e. the dummy for geographical macro-areas) cannot be enough detailed to capture more local phenomena. For instance, it can be the case that a higher density of schools, which in turn can have direct and indirect effects on competition and affecting students' performances, characterizes some Provinces[15]; alternatively, it can be that families have different levels of education across Provinces, or they do care with a different intensity about education, and this have an effect on the schools chosen (i.e. they choose schools, which sort more/less). Whatever the mechanism, we would verify if more detailed information about the geographical location of schools do contribute to explain students' achievement together with schools' sorting behaviour. Given that the effects is not theoretically clear, we included fixed effects instead that covariates at Province level; moreover, we use two different specifications to analyse different role played by regional of provincial distribution of schools. Mathematically, we estimate:

$$Y_{ijk(z)t} = \alpha_0 + \lambda Y_{ijk(z)(t-1)} + \alpha_1 X_{1ijk(z)t} + \alpha_2 X_{2jk(z)t} + \alpha_3 X_{3k(z)t} + \phi_z Z_z + \varepsilon_{ijk(z)t} \tag{5}$$

where z is a subscript for the z=1,…Z Italian Regions (20) or Provinces (103), and ϕ_z is the coefficient for the Region/Province to be estimated. The results are presented in the panel E of table 5, and showed that, when geographical fixed effects are included, the magnitude of the coefficient for ESCS_Var_Within tends to diminish. In addition, it becomes no statistically significant also for reading test scores, when Province-level fixed effects are concerned; this evidence suggests that at least part of the phenomenon under scrutiny is partly related to specific local factors, as suggested by Ferrer-Esteban (2011). The data at-hands, however, does not permit deeper explorations of these patterns; at the same time, the direction of coefficients' signs confirm that, even after controlling for (unobservable) structural differences between Provinces, still the academic effect of schools' practices of sorting students between classes – if any – is negative.

[15] Agasisti (2011) actually showed that different competitive pressures in Italian geographical areas partly account for differentials in achievement.

5. Implications and concluding remarks

The analyses presented in this paper demonstrate that some Italian junior secondary schools do practice between-classes sorting, despite the legislation being inspired by a principle of "equal-heterogeneity" in the composition of classes. Operationally, we propose to calculate an indicator, named ESCS_Var_Within, which measures the variance in the socioeconomic background of students that can be attributable to between-classes structural differences. In 2011/12, for almost 20% of the Italian JSS (around 1,000 institutions), ESCS_Var_Within was higher than 15%. The first by-product of this research is evidencing that the phenomenon is not such widespread, so in general the schools actually pursue equality objectives when composing their classes; therefore, the existence of some schools in which there are strong differences in the average socioeconomic characteristics of students between classes, justifies a specific attention to evaluate its policy consequences.

Assuming that schools intentionally decide if sorting or not, we adopted an adequate IV approach, using the value of ESCS_Var_Within in the year before (2010/11) as an instrument. The results highlight a substantial, negative and statistically significant effect of sorting on test scores in reading (but not mathematics): more specifically, the impact is 0.05 s.d., meaning for example that a student attending a school where ESCS_Var_Within is 16% instead of 8%, obtains a score which is almost 1 point lower (on a 0;100 scale, where 65 is the mean), all else equal. This effect is not uniform across all school types and geographical areas: it is instead determined by public schools, especially those located in Southern Italy. Moreover, sorting seems to impact more negatively the socioeconomically disadvantaged students, and the abler ones (those whose test score in grade 5 was higher), reinforcing the suspect that sorting finishes to be based more on socioeconomic status than ability. Further on this mechanism, we also empirically testified that the benefits for better-off students are not able to compensate the negative effects for disadvantaged students.

The findings have some policy implications. First, the differences between schools in their policies about between-classes sorting should be made public. Given that the legislation assigns the powers to define criteria for composing the classes to the school administrators and teachers, sorting should be regarded as an acceptable practice, on a theoretical ground; what is unacceptable is that sorting is practiced without any information to the parents and students, and even that it could be masked by the application of “equal-heterogeneity”. Second, schools that practice sorting should provide theoretical and empirical justification about their decision; indeed, the pieces of evidence reported in this paper suggest that such strategy actually harms student achievement (at least in reading). Third, if the principle of “equality of educational opportunities” is regarded as a priority, such practice should be discouraged, independently from its independent impact on student achievement: indeed, higher levels of socioeconomic segmentation tend to reproduce social and educational disparities, instead of contributing in closing the gap between disadvantaged and advantaged students. With respect to this latter point, the problem of promoting inequality would be even more dramatic if poor students are more likely to be enrolled in a school that practices sorting and/or in a class in which students’ average socioeconomic background is disadvantaged: this is the case, unfortunately. Indeed, for empirically testing such hypothesis, we defined two dummies, namely (i) one for the student being enrolled in a school that practices sorting ($\text{School\_Sorting\_High}_i$), defining this condition as $\text{ESCS\_Var\_Within}_k > 15\%$, and one ($\text{Disadvantaged\_Class}_i$) for students in classes which average ESCS is lower than the 25th percentile (specifically, -0.19). Simple probit regressions were then performed, using individual-level characteristics as predictors, such as:

$$\Pr(Y_i = 1 \mid X_{ijk}) = \Phi(X_{ijk}'\beta) \qquad (6)$$

where Yi is $\text{School\_Sorting\_High}_i$ or $\text{Disadvantaged\_Class}_i$, alternatively, X_{ijk} are the students’ characteristics as defined in the previous sections, Φ is the distribution function and β is the vector of parameters to be estimated. The results are reported in the table 6, and clearly demonstrate that the probability of being in a poor class and/or in a sorting school is linearly decreasing with the socioeconomic status (ESCS). This relationship holds for being in a disadvantaged class even after having controlled for the school-average ESCS (column b). The marginal effects, computed as first derivative of the coefficient with respect to the probability, signal that the likelihood of being in a disadvantaged class ranges between 5% and 10%. These findings bring further evidence of the channels through which educational inequality persists and is even pursued through the sorting mechanisms discussed in the paper, and claim for a serious reflection about the risky long-term effects of this practice.

This study innovates the existent literature, as it is the first one that looks explicitly at the between-classes segmentation in Italy; moreover, it is also among the few studies that adopt a value-added specification in this country, by including prior achievement in the estimation of the educational production function. In addition, while previous studies in Italy consider the tracking phenomenon in its formal version at high school level, here we explored informal tracking at junior secondary schools, where it is likely to have a stronger and longer-run effect. Our results are specific to the Italian context, and limited to the grade 6; they are neither easily extendable to other contexts or countries, nor to other educational levels. Therefore, this work can provide some empirical and theoretical insights into the effects of sorting students within schools, which can be helpful also for those educational systems where ability sorting and/or tracking is a common practice. An obvious direction for extending current research in the future is to apply our empirical strategy to other cohorts of students, when new wave of data will become available. Also, following students over time could provide evidence about the effects of tracking on subsequent grades and post-secondary decisions. Last but not least, the patterns that we identified as Province-level structural differences in the phenomenon are worthy to be further explored, to understand if there are differences in the educational practices in different areas of the country – and not only the well-known differences in mean performance between Northern and Southern Italy.

References

Adnett N. and Davies P., (2005), Competition between or within schools? Re-assessing school choice, Education Economics, 13(1), 109-121.

Agasisti T., (2011), How competition affects schools' performances: does specification matter?, Economics Letters, 110(3), 259-261.

Agasisti T. and Vittadini G., (2012), Regional economic disparities as determinants of students' achievement in Italy, Research in Applied Economics, 4(1), 33-53.

Ammermueller A and Pischke J.S., (2009), Peer effects in European primary schools: evidence from the Progress in International Reading Literacy Study, Journal of Labor Economics, 27(3), 315-348.

Angrist J.D and Pischke J.S., (2009), Mostly harmless econometrics, Princeton, NJ: Princeton University Press.

Barbieri G., Rossetti C. and Sestito P., (2011), The determinants of teacher mobility: evidence using Italian teachers' transfer application, Economics of Education Review, 30(6), 1430-1444.

Bertoni M., Brunello G. and Rocco L., (2013), When the cat is near, the mice won't play: The effect of external examiners in Italian schools, Journal of Public Economics, 104, 65-77.

Borland M.V., Howsen R.M. and Trawick M.W., (2006), Intra-school competition and student achievement, Applied Economics, 38(14), 1641-1647.

Bratti M., Checchi D. and Filippin D., (2007), Geographical differences in Italian students' mathematical competencies: evidence from PISA 2003, Giornale degli Economisti e Annali di Economia, 66(3), 299-333.

Brunello G. and Checchi D., (1997), Does school tracking affect equality of opportunity? New international evidence, Economic Policy, 22(52), 781-861.

Caldas S.J. and Bankston C., (1997), Effect of school population socioeconomic status on individual academic achievement, The Journal of Educational Research, 90(5), 269-277.

Campodifiori E., Figura E., Papini M. and Ricci R., (2010), Un indicatore di status socio-economico-culturale degli allievi della quinta primaria in Italia (An indicator for students' socio-economic background), INVALSI Working Paper no. 02/2010, INVALSI, Rome (Italy).

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

Collins C.A. and Gan L., (2013), Does sorting students improve scores? An analysis of class composition, NBER Working Paper, n. 18848.

Condron D.J., (2007), Stratification and educational sorting: explaining ascriptive inequalities in early childhood reading group placement, Social Problems, 54(1), 139-260.

De Simone G. and Gavosto A., (2013), Patterns of value-added creation in the transition from primary to lower-secondary education in Italy, FGA Working Paper n. 47 (04/2013).

Di Liberto A., Schivardi F. and Sulis G., (2013), Managerial practices and students' performance, FGA Working Paper n. 49 (07/2013).

Ferrer-Esteban G., (2011), Beyond the traditional territorial divide in the Italian education system. Effects of system management factors on performance in lower secondary schools, FGA Working Paper n. 42 (12/2011).

Gaviria A. and Raphael S., (2011), School-based peer effects and juvenile behaviour, The Review of Economics and Statistics, 83(2), 257-268.

Gorard S. and Cheng S. C., (2011), Pupil clustering in English secondary schools: one pattern or several? International Journal of Research & Method in Education, 34(3), 327-339.

Haveman R. and Wolfe B., (1995), The determinants of children's attainments: a review of methods and findings, Journal of Economic Literature, XXXIII, 1829-1878.

INVALSI – *Istituto nazionale per la valutazione del sistema educativo di istruzione e di formazione* (2013), Rilevazioni nazionali sugli apprendimenti 2012/13 (National standardized test scores 2012/13), Roma: INVALSI.

Kang C., (2007), Academic interactions among classroom peers: a cross-country comparison using TIMSS, Applied Economics, 39(12), 1531-1544.

Loveless T., (2013), How well are American students learning? The 2013 Brown Center Report on American Education, Brookings Institution: Washington (DC), USA.

Mocetti S., (2008), Educational choices and the selection process before and after compulsory schooling, Bank of Italy Temi di Discussione, n. 691 (September, 2008).

Mullis I.V.S., Martin M.O., Foy P. and Arora, A. (2012), TIMSS 2011 International Results in Mathematics, Chestnut Hill, MA: TIMSS & PIRLS International Study Center, Boston College.

Oakes J. (1986), Keeping track, part 1: the policy and practice of curriculum inequality, Phi Delta Kappa, 68(1), 12-17.

Oakes J., (2005), Keeping track: how schools structure inequality, 2nd edition, New Haven, CT: Yale University Press.

Oakes J. (2008), Keeping track: structuring equality and inequality in an era of accountability, Teachers College Record, 110(3), 700-712.

OECD – Organization for Economic Co-operation and Development, (2012), PISA 2009 Technical Report, Paris: OECD.

Perri L.B. and McConney A., (2010), Does the SES of the school matter? An examination of socioeconomic status and student achievement using PISA 2003, Teachers College Record, 112(4), 1137-1162.

Poletto S., (1992), La formazione delle classi (The process of composing classes), in Bartolini, D. (ed.), L'uomo – la scuola (The man, the school), Nuova Dimensione, Portogruaro (Italy).

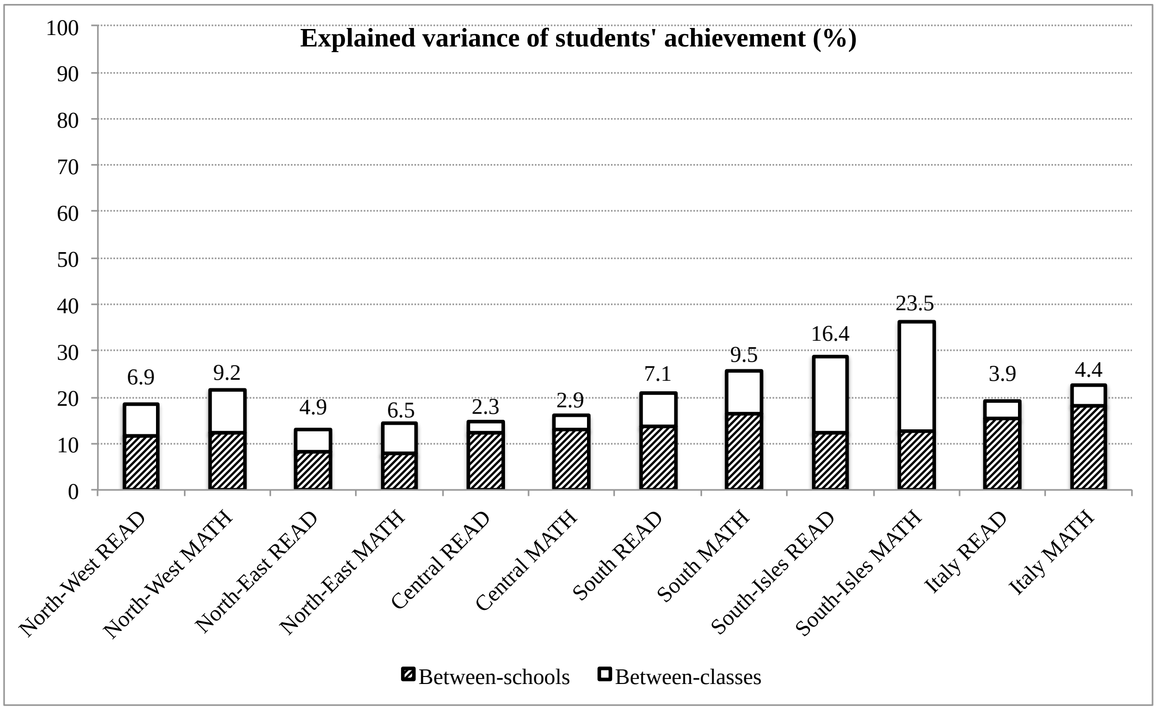
Schneider T., (2003), School class composition and student development in cognitive and non-cognitive domains: longitudinal analyses of primary school students in Germany, in Windzio M. (ed.), Integration and inequality in educational institutions, 167-190, Dordrecht (The Netherlands): Springer.

Sirin S.R., (2005), Socioeconomic status and academic achievement: a meta-analytic review of research, Review of Educational Research, 75(3), 417-453.

Staiger D. and Stock J.H., (1997) Instrumental variables regression with weak instruments, Econometrica, 65(3), 557 – 586.

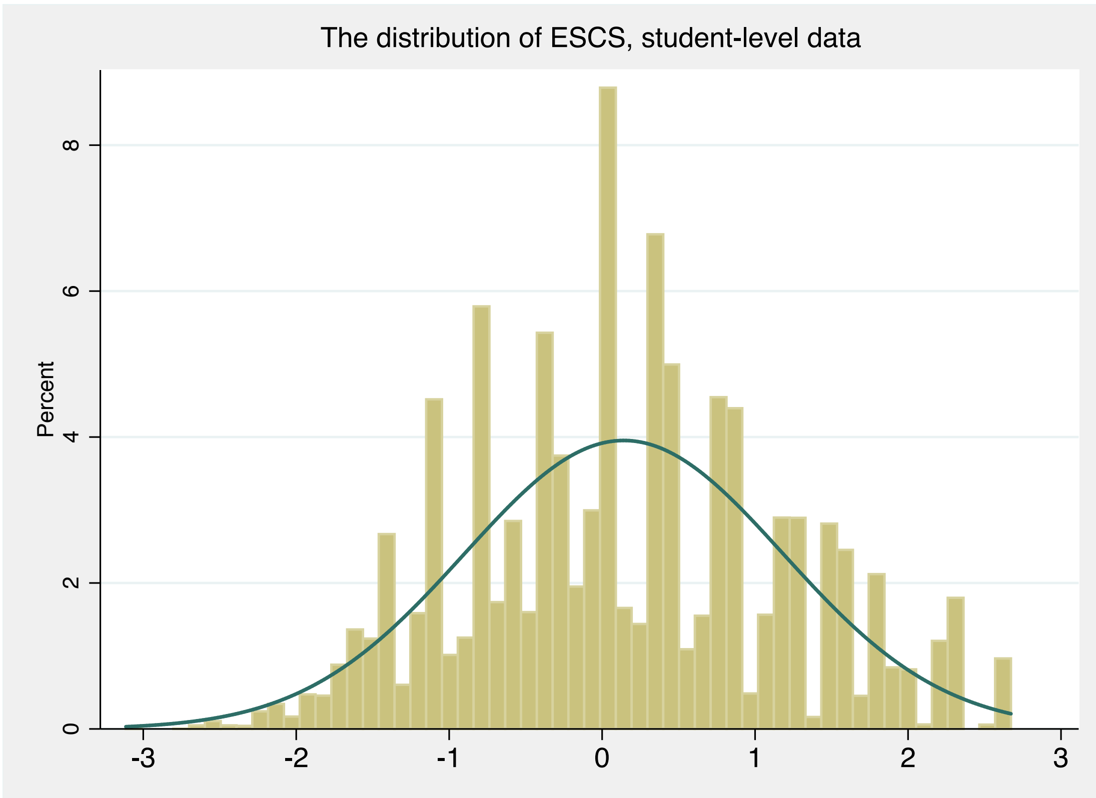
Vardardottir A., 2013, Peer effects and academic achievement: a regression discontinuity analysis, Economics of Education Review, 36(1), 108-121.

Figure 1.
Decomposition of variance in students' achievement: between-schools and between-classes (within the same school)
Grade 6 students, 2012/13



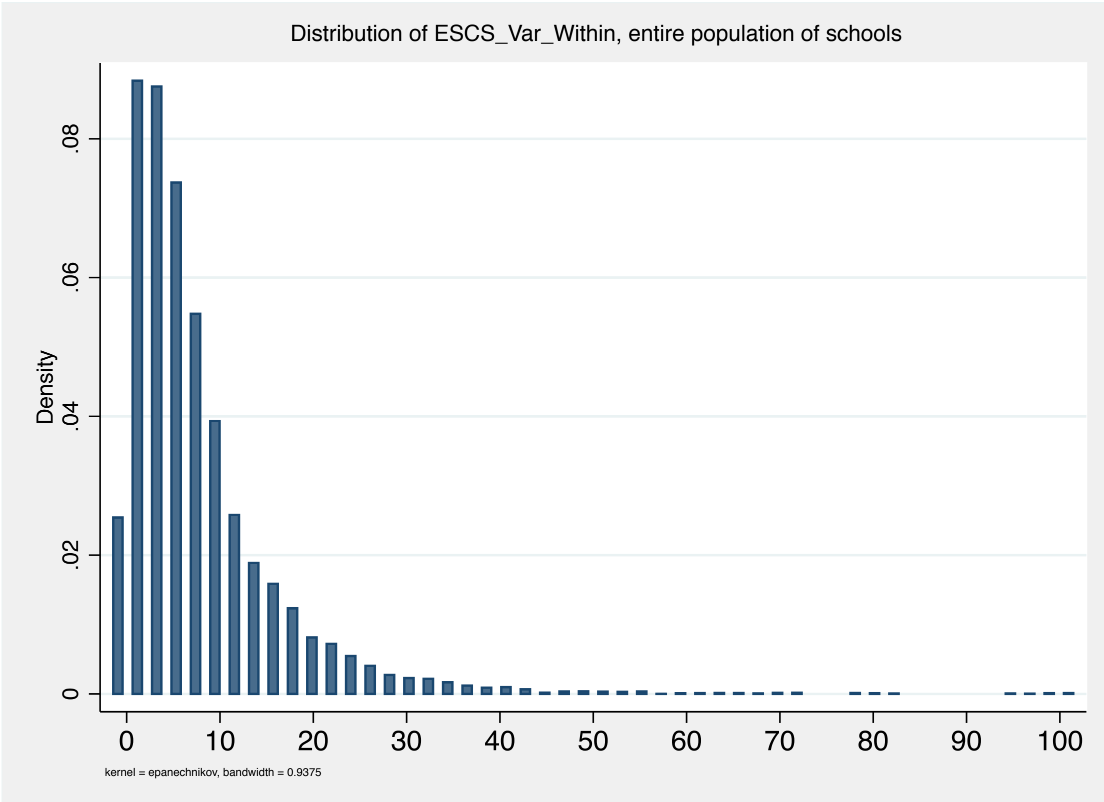
Notes. Authors' adaptation from INVALSI (2013; p. 36).

Figure 2.
The distribution of ESCS (the variable measuring students' Economic, Social and Cultural Status)



Notes. By construction, the variable has been built for having a distribution (0;1). The solid line represents the normal distribution of data. Student-level data (# observations: 467,121).

Figure 3.
The distribution of ESCS_Var_Within$_k$ (the variance of students' socioeconomic status, between classes within each school, as measured for each k schools)



Notes. The numbers on the horizontal axis can be interpreted as percentages. #schools = 5,008. Details about how the variable is calculated are provided in the section "Methodology".

Table 1.
The distribution of $ESCS\_Var\_Within_k$ (the variance of students' socioeconomic status, between classes within each school, as measured for each k schools)

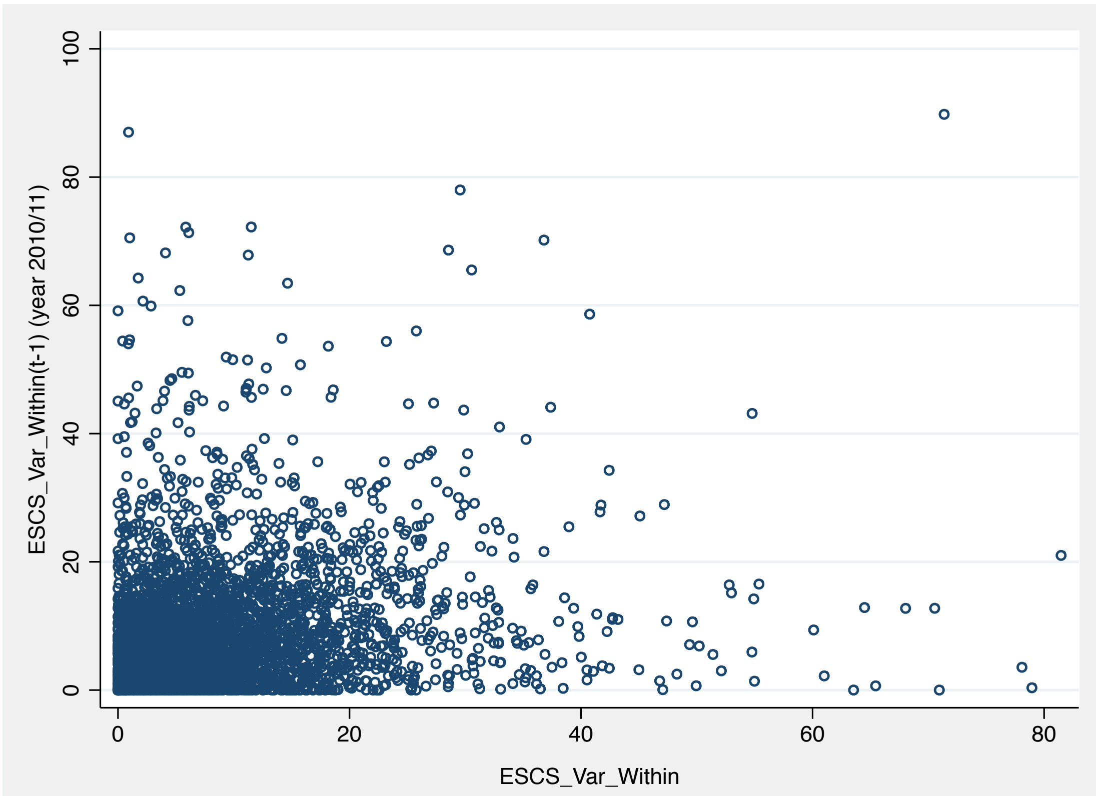
Panel A. Distribution by percentiles

Percentiles	Value of ESCS_Var_Within
1%	0.030
5%	0.637
10%	1.398
25%	3.291
50%	6.414
75%	11.278
90%	18.266
95%	24.044
99%	37.493
Mean	8.588
Std. Dev.	7.965

Panel B. The number of schools with a given value of ESCS_Var_Within

ESCS_Var_Within	Number of schools	% of the total
<5%	2,374	44.7%
bw 5% and 10%	1,380	26.0%
bw 10% and 15%	565	10.6%
bw 15% and 20%	312	5.9%
bw 20% and 25%	162	3.0%
bw 25% and 30%	86	1.6%
bw 30% and 35%	50	0.9%
bw 35% and 40%	28	0.5%
bw 40% and 45%	16	0.3%
bw 45% and 50%	10	0.2%
bw 50% and 75%	19	0.4%
>75%	313	5.9%
Total	5,315	100.0%

Figure 4.
Evaluating the IV approach: the relationship between the variable of interest ($ESCS\_Var\_Within_{k(t)}$) and the instrument ($ESCS\_Var\_Within_{k(t-1)}$)



Notes. Both variables are measured at school level; the pairwise correlation between the two is 0.24, and it statistically significant at 1%.

Table 2.
Descriptive statistics, all the variables

Variable	Mean	Std. Dev.	N
Student-level characteristics	Mean	Std. Dev.	N
Achievement in Reading	65.237	16.161	510,933
Prior achievement (grade 5) - Reading	74.216	13.834	241,955
Achievement in Mathematics	46.164	17.966	509,371
Prior achievement (grade 5) - Mathematics	70.289	16.426	272,614
Female student	0.489		510,032
1st generation immigrant	0.055		469,517
2nd generation immigrant	0.049		469,517
Early-enrolled student	0.022		510,019
Late-enrolled student	0.073		510,019
Socioeconomic background (index ESCS; mean=0, stdev=1)	0.137	1.045	468,203
Student who does NOT live with both parents	0.137		490,272
Student who has siblings	0.842		490,722
Classroom-level characteristics	Mean	Std. Dev.	N
Cheating propensity	0.074	0.193	510,873
Classroom selected to be part of the "controlled" sample	0.077	0.266	510,933
Class-average socioeconomic background (index ESCS)	0.136	0.553	469,822
Proportion of females in the classrom	0.436		510,757
Proportion of 1st generation immigrants in the classroom	0.049		470,367
Proportion of 2nd generation immigrants in the classroom	0.043		470,367
Proportion of Early-enrolled students in the classroom	0.020		510,757
Proportion of Late-enrolled students in the classroom	0.064		510,757
Proportion of disabled students in the classroom	0.054		510,933
Number of students in the classroom	23.145	3.574	510,933
Proportion of students who took the test	94.262	7.601	510,933
Class with *"tempo-pieno"*	0.024	0.153	470,367
School-level characteristics	Mean	Std. Dev.	N
Number of classrooms in the school	6.482	3.079	510,933
Number of students in the school	149.638	77.554	510,933
Average number of students per class, in the school	22.695	2.976	510,933
School located in Northern Italy	0.432		510,933
School located in Central Italy	0.177		510,933
School located in Southern Italy	0.391		510,933
School located in PON area	0.333		510,933
Istituto Comprensivo	0.606		510,933
Private school	0.029		510,933
Variance in the ESCS index between classes (ESCS_Var_Within)	8.758	8.290	480,751

Notes. Standard deviation is reported only for continuous variables.

Table 3.
Statistical correlations between ESCS_Var_Within$_k$ (the variance of students' socioeconomic status, between classes within each school, as measured for each k schools) and schools' results

Panel A. The correlation between ESCS_Var_Within$_k$ and school-average test score

	Reading	Mathematics
Italy	-.168**	-.133**
Northern Italy		
Valle D'Aosta	.087	-.219
Piemonte	-.119*	-.189**
Liguria	.057	.005
Lombardia	-.039	.017
Prov. Aut. Bolzano (l. it.)	-.481	-.205
Prov. Aut. Trento	-.200	.033
Veneto	-.086	-.230**
Friuli-Venezia Giulia	-.260*	.082
Emilia-Romagna	-.056	-.049
Central Italy		
Toscana	.025	.076
Umbria	-.253*	.079
Marche	-.106	-.030
Lazio	-.127*	-.051
Southern Italy		
Abruzzo	-.078	.071
Molise	.004	.109
Campania	-.155**	-.070
Puglia	-.178**	-.145**
Basilicata	-.247*	-.057
Calabria	-.202**	.053
Sicilia	-.173**	.004
Sardegna	-.027	-.145

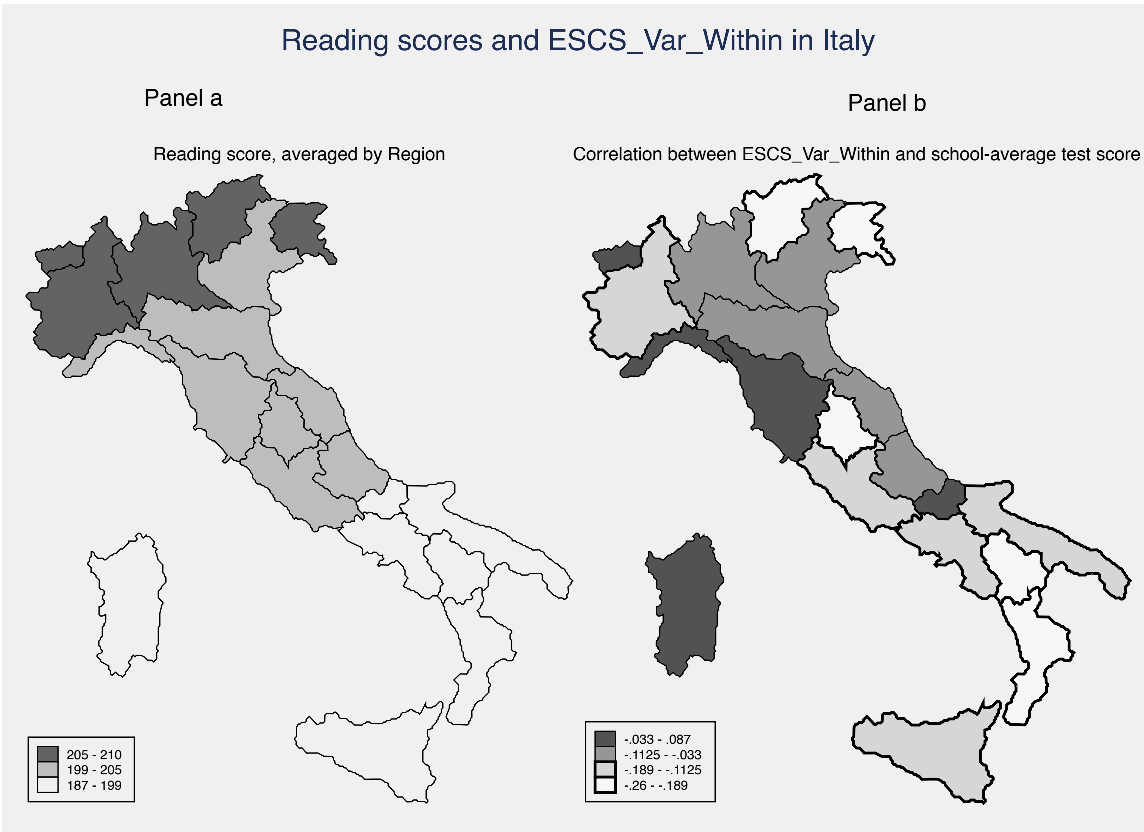
Notes. * means statistically significant at 5% level; ** 1%

Panel B. The correlation between ESCS_Var_Within$_k$ and within-school variance in test scores

	Reading	Mathematics
Italy	.169**	.286**
Northern Italy		
Valle D'Aosta	-0.22	-0.01
Piemonte	.088	.467**
Liguria	-.007	.094
Lombardia	.071	.169**
Prov. Aut. Bolzano (l. it.)	.379	-.098
Prov. Aut. Trento	.145	.075
Veneto	.129**	.066
Friuli-Venezia Giulia	.286**	.224*
Emilia-Romagna	.084	.198**
Central Italy		
Toscana	.038	.041
Umbria	.302**	0.066
Marche	.171*	.207*
Lazio	.184**	.143**
Southern Italy		
Abruzzo	.182*	0.091
Molise	-.036	.015
Campania	.248**	.345**
Puglia	.283**	.352**
Basilicata	.303**	.279*
Calabria	.221**	.217**
Sicilia	.248**	.179**
Sardegna	.079	.088

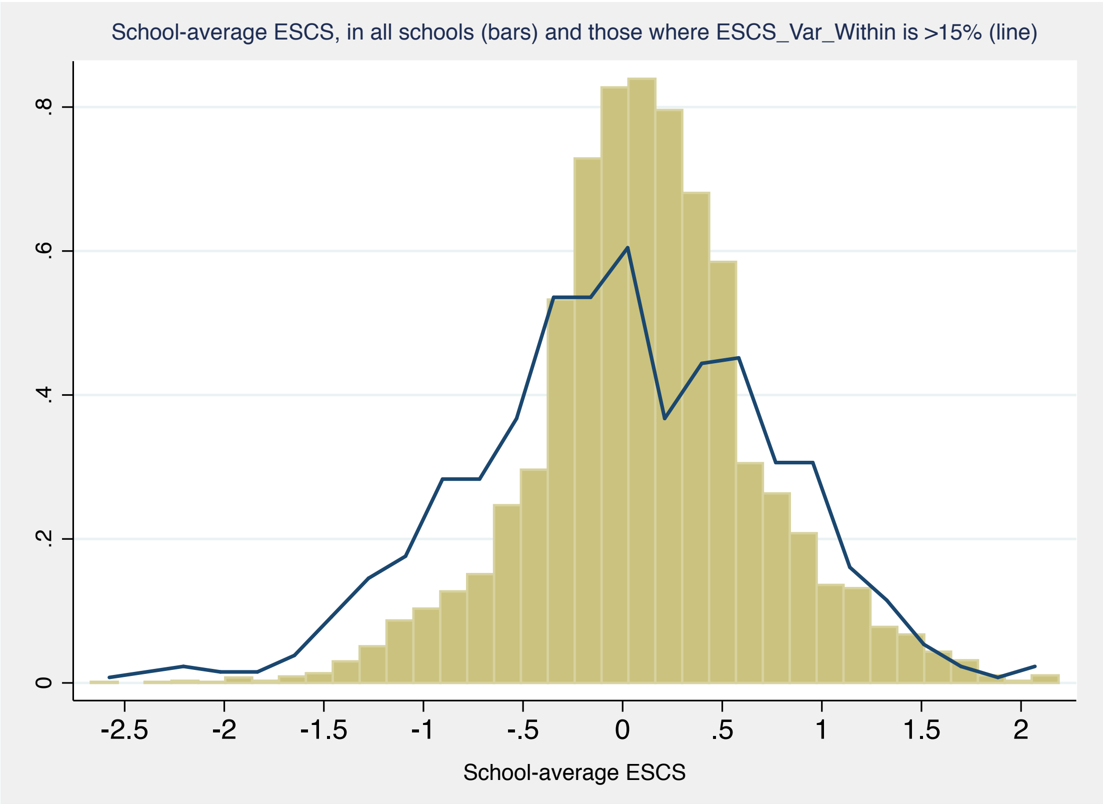
Notes. * means statistically significant at 5% level; ** 1%

Figure 5.
The correlation between ESCS_Var_Within$_k$ and school-average test score: geographical distribution, by Region



Notes. The colours represent the estimated coefficient of the correlation between ESCS_Var_Within and school-average score in Reading – the lighter the colour, the higher the coefficient. The Regions delimited by ticker borders are those for which the correlation is statistically significant.

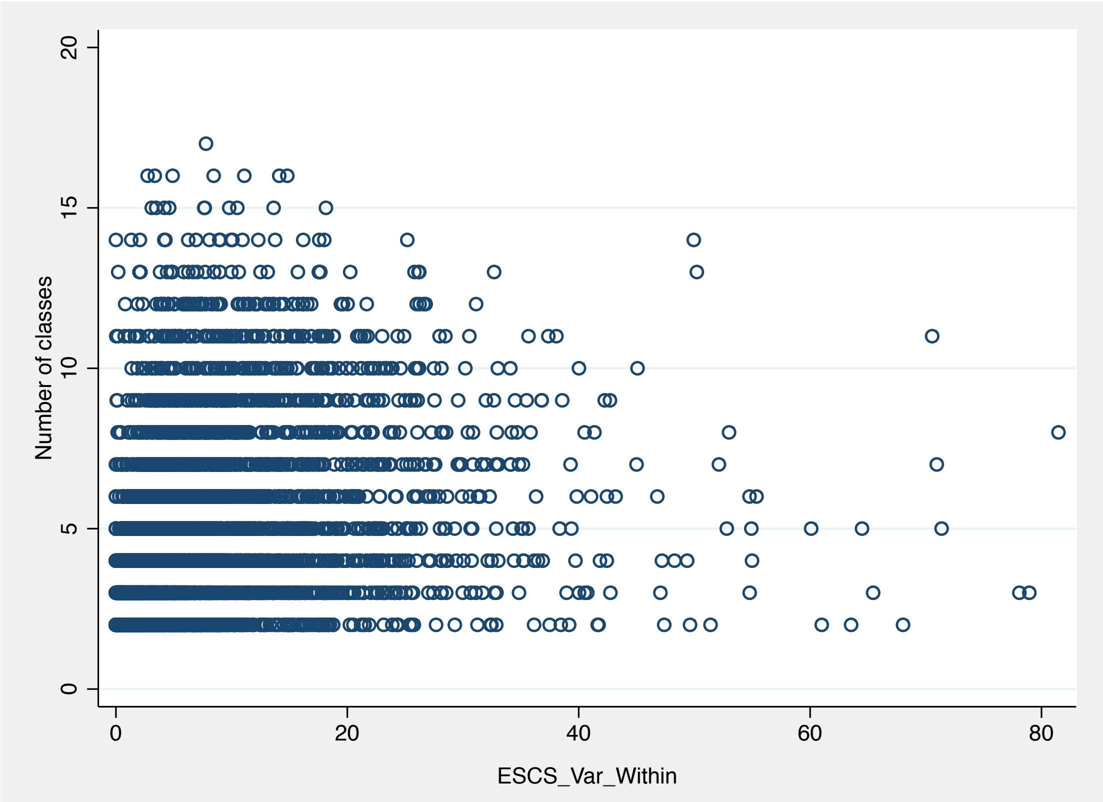
Figure 6.
School-average ESCS in the entire population of schools and those practising more socioeconomic sorting between classes



Notes. School-average ESCS is calculated as the average of ESCS of all the students attending the school. The histogram shows the distribution for the entire population of students; the blue line is the distribution of school-average ESCS for those schools that are practising more socioeconomic sorting between classes (operationally they are those schools for which $ESCS\_Var\_Within_k$ is >15%).

Figure 7.

The relationship between ESCS_Var_Within$_k$ (the variance of students' socioeconomic status, between classes within each school, as measured for each k schools) and the number of classes – measures at school-level



Notes. Both variables are measured at school level; the pairwise correlation between the two is 0.24, and it statistically significant at 1%.

Table 4.
Baseline results: educational production function (EPF) estimates

	Reading	Math
Student-level characteristics	b-coeff	b-coeff
Prior achievement (grade 5)	0.487***	0.500***
	(97.76)	(102.59)
Female student	0.074***	-0.064***
	(42.98)	(-37.21)
1st generation immigrant	-0.055***	-0.016***
	(-24.54)	(-8.82)
2nd generation immigrant	-0.053***	-0.026***
	(-28.92)	(-16.06)
Early-enrolled student	-0.009***	-0.006***
	(-4.83)	(-3.62)
Late-enrolled student	-0.042***	-0.025***
	(-17.47)	(-13.12)
Socioeconomic background (index ESCS; mean=0, stdev=1)	0.155***	0.136***
	(66.33)	(61.64)
Student who does NOT live with both parents	-0.024***	-0.027***
	(-13.53)	(-16.42)
Student who has siblings	-0.016***	0.004*
	(-9.43)	(2.15)
Classroom-level characteristics	b-coeff	b-coeff
Cheating propensity	0.084***	-0.120***
	(19.71)	(-30.52)
Classroom selected to be part of the "controlled" sample	0.007**	-0.008**
	(2.60)	(-2.80)
Class-average socioeconomic background (index ESCS)	0.016**	0.019***
	(2.92)	(4.62)
Proportion of females in the classrom	-0.002	0.010**
	(-0-50)	(2.68)
Proportion of 1st generation immigrants in the classroom	0.014***	0.005
	(3.64)	(1.28)
Proportion of 2nd generation immigrants in the classroom	0.005	-0.003
	(1.70)	(-0.86)
Proportion of Early-enrolled students in the classroom	0.009*	-0.002
	(1.97)	(-0.42)
Proportion of Late-enrolled students in the classroom	-0.014**	-0.013**
	(-3.47)	(-3.25)
Proportion of disabled students in the classroom	0.001	0.006
	(0.27)	(1.59)
Number of students in the classroom	0.024***	0.016**
	(5.23)	(3.17)
Proportion of students who took the test	0.033***	0.010*
	(7.99)	(2.54)
Class with *"tempo-pieno"*	-0.003	0.000
	(-0.78)	(0.03)

	Reading	Math
School-level characteristics	b-coeff	b-coeff
Number of classrooms in the school	-0.016	-0.047
	(-0.40)	(-1.17)
Number of students in the school	0.039	0.066
	(0.92)	(1.56)
Average number of students per class, in the school	-0.012	-0.011
	(-1.26)	(-1.07)
School located in Northern-West Italy	0.020***	0.004
	(5.22)	(0.83)
School located in Central Italy	-0.021***	-0.040***
	(-4.73)	(-8.78)
School located in Southern Italy	-0.070***	-0.082***
	(-6.25)	(-7.34)
School located in PON area	-0.052***	0.005
	(-4.85)	(0.47)
Istituto Comprensivo	0.003	0.004
	(0.36)	(0.57)
Private school	-0.021***	-0.016**
	(-4.05)	(-2.94)
Variance in the ESCS index between classes (ESCS_Var_Within)	-0.050*	-0.019
	(-2.20)	(-0.77)
N	221,224	247,351
adj. R^2	0.382	0.360

Notes. *** is statistically significant at 1%; ** 5%, * 10%. p-values in parentheses.
ESCS_Var_Within is instrumented with ESCS_Var_Within (t-1); the coefficients of first-stage regression are in the Annex 1. Robust standard errors are clustered at school-level.

Table 5.

Panel A. The heterogeneous effect of ESCS_Var_Within on students' performances, by school type and location (geographical macro-areas)

Impact of ESCS_Var_Within	Reading			Math		
Public versus private schools	N	Coefficient	R^2	N	Coefficient	R^2
Public schools	214,495	-0.051*	0.382	238,846	-0.020	0.358
Private schools	6,729	0.130	0.348	8,505	0.283	0.322
By geographical macro-area	N	Coefficient	R^2	N	Coefficient	R^2
Northern Italy	114,418	0.006	0.460	125,595	-0.001	0.455
Central Italy	39,252	-0.820	0.220	44,025	-0.411	0.207
Southern Italy	67,554	-0.075*	0.283	77,731	-0.015	0.248

Notes. *** is statistically significant at 1%; ** 5%, * 10%.
ESCS_Var_Within is instrumented with ESCS_Var_Within (t-1). Robust standard errors are clustered at school-level.

Panel B. Extending the understanding of the effects of ESCS_Var_Within: considering interactions with students' prior achievement and ESCS

	Reading	Math
ESCS_Var_Within	-0.089***	0.072**
	-3.665	*2.833*
ESCS_Var_Within * Prior Achievement	-0.154***	-0.207***
	-6.120	*-8.559*
ESCS_Var_Within *ESCS	0.057***	0.044**
	3.555	*3.140*
N	221,224	247,351
Adj R2	0.378	0.3584

Notes. *** is statistically significant at 1%; ** 5%, * 10%. t-values in italics.
(i) ESCS_Var_Within, (ii) ESCS_Var_Within * Prior Achievement and (iii) ESCS_Var_Within * ESCS are instrumented with (i) ESCS_Var_Within(t-1), (ii) ESCS_Var_Within(t-1) * Prior Achievement and (iii) ESCS_Var_Within(t-1) * ESCS. Robust standard errors are clustered at school-level.

Panel C. The impact of ESCS_Var_Within on students' achievement, at different points of its distribution

	ESCS_Var_Within<25%	ESCS between 25% and 75%	ESCS_Var_Within>75%
	(a)	(b)	(c)
Reading score	0.636	-0.051	-0.080
	0.64	*-1.74*	*-1.08*
Mathematics score	0.609	-0.008	-0.059
	0.71	*-0.25*	*-0.80*

Notes. *** is statistically significant at 1%; ** 5%, * 10%. t-values in italics.
ESCS_Var_Within is instrumented with ESCS_Var_Within (t-1). Robust standard errors are clustered at school-level. In the column (a), only those students who attend schools with ESCS_Var_Within lower than 3.3% (25th percentile are considered); in the column (c) only those students attending a school with ESCS_Var_Within is higher than 11.2% (75th percentile); in the column (b), the other students.

Panel D. The impact of ESCS_Var_Within and the class-average ESCS

	Reading	Math
ESCS_Var_Within	-0.060**	-0.034
	-2.600	*-1.248*
ESCS_Var_Within * class_avg_ESCS	0.068*	0.055*
	2.534	*2.012*
N	221,224	247,351
Adj R2	0.381	0.359

Notes. *** is statistically significant at 1%; ** 5%, * 10%. t-values in italics.
(i) ESCS_Var_Within and (ii) ESCS_Var_Within * class_avg_ESCS are instrumented with (i) ESCS_Var_Within(t-1), (ii) ESCS_Var_Within(t-1) * class_Avg_ESCS. Robust standard errors are clustered at school-level.

Panel E. The impact of ESCS_Var_Within – check when including Region or Province-level fixed effects

		Reading	Math
Baseline model		-0.050*	-0.019
		(-2.20)	(-0.77)
	N	221,224	247,351
	R^2	0.382	0.360
Model w/ Region-level FE		-0.047*	-0.013
		(-2.13)	(-0.56)
	N	221,224	247,351
	R^2	0.384	0.362
Model w/ Province-level FE		-0.041	-0.009
		(-1.81)	(-0.37)
	N	221,224	247,351
	R^2	0.387	0.366

Notes. *** is statistically significant at 1%; ** 5%, * 10%. t-values in parentheses.
ESCS_Var_Within is instrumented with ESCS_Var_Within (t-1). Robust standard errors are clustered at school-level. Region-level FE considers fixed effects for the 20 Italian Regions; Province-level FE considers fixed effects for the 103 Italian Provinces.

Table 6.
The determinants of probability to be enrolled in a school where ESCS_Var_Withink is >15% (School_Sorting_High$_i$) or in a class where average ESCS is lower than 25th percentile (Disadvantaged_Class$_i$)

	Dep variable Disadvantaged_ Class	Dep variable Disadvantaged_ Class	Dep variable School_Sorting_ High
	(a)	(b)	(c)
Prior achievement (grade 5)	-0.0001	-0.0001	0.0015***
	0.0002	*0.0003*	*0.0003*
Female student	0.0089	-0.0011	0.0049
	0.0063	*0.0076*	*0.0066*
1st generation immigrant	-0.2759***	-0.0525	-0.2227***
	0.0167	*0.0197*	*0.0196*
2nd generation immigrant	-0.2708***	-0.0690***	-0.2268***
	0.0146	*0.0172*	*0.0170*
Early-enrolled student	0.2190***	0.0608**	0.3632***
	0.0240	*0.0309*	*0.0228*
Late-enrolled student	0.0399**	0.0677***	0.0471**
	0.0199	*0.0238*	*0.0226*
Socioeconomic background (index ESCS)	-0.5144***	-0.3061***	-0.0311***
	0.0035	*0.0044*	*0.0034*
Student who does NOT live with both parents	-0.1282***	-0.0042	-0.0061
	0.0097	*0.0119*	*0.0101*
Student who has siblings	0.1396***	0.0577***	0.0811***
	0.0090	*0.0110*	*0.0092*
School-average ESCS		-3.3251***	
		0.0164	
Constant	-0.9180***	-0.8245***	-1.1837***
	0.0091	*0.0112*	*0.0095*
Log_likelihood	-103,657.28	-68,561.13	-88,421.01
N	232,474	232,474	232,474
Marginal effect (dx/dy) - ESCS	-0.1283***	-0.0506***	-0.0065***
	0.0008	*0.0007*	*0.0007*

Notes. *** is statistically significant at 1%; ** 5%, * 10%. Standard errors in italics. All the models are estimated through a Probit Regression.

Annex A

First-stage regression in the IV approach

(dependent variable: ESCS_Var_Within$_{(t)}$; instrument: ESCS_Var_Within$_{(t-1)}$)

	Reading		Mathematics	
Variables	Coefficient	St. Err.	Coefficient	St. Err.
Prior achievement (grade 5)	0.002*	0.001	0.003***	0.001
Female student	-0.024	0.027	-0.001	0.025
1st generation immigrant	-0.006	0.075	0.011	0.071
2nd generation immigrant	0.049	0.065	0.031	0.062
Early-enrolled student	0.015	0.108	-0.034	0.096
Late-enrolled student	0.023	0.089	0.024	0.084
Socioeconomic background (index ESCS; mean=0, st.dev=1)	-0.032*	0.015	-0.046***	0.014
Student who does NOT live with both parents	0.175***	0.040	0.147***	0.038
Student who has siblings	0.004	0.036	-0.002	0.034
Cheating propensity	1.528***	0.187	5.052***	0.255
Classroom selected to be part of the "controlled" sample	0.863***	0.049	0.611***	0.046
Class-average socioeconomic background (index ESCS)	0.673***	0.036	1.042***	0.033
Proportion of females in the classroom	0.008***	0.001	0.007***	0.001
Proportion of 1st generation immigrants in the classroom	-0.000	0.003	0.001	0.003
Proportion of 2nd generation immigrants in the classroom	-0.007**	0.003	-0.009***	0.002
Proportion of Early-enrolled students in the classroom	0.036***	0.005	0.066***	0.004
Proportion of Late-enrolled students in the classroom	0.010***	0.003	0.006**	0.003
Proportion of disabled students in the classroom	0.032***	0.003	0.029***	0.002
Number of students in the classroom	0.038***	0.006	0.039***	0.005
Proportion of students who took the test	-0.044***	0.002	-0.039***	0.002
Class with *"tempo-pieno"*	-0.983***	0.088	-0.813***	0.083
Number of classrooms in the school	0.858***	0.042	0.255***	0.037
Number of students in the school	-0.020***	0.002	0.005***	0.002
Average number of students per class, in the school	-0.033***	0.011	-0.130***	0.011
School located in Northern-West Italy	-0.136***	0.038	0.074**	0.036
School located in Central Italy	0.652***	0.042	0.609***	0.039
School located in Southern Italy	1.514***	0.066	1.224***	0.062
School located in PON area	1.759***	0.065	2.103***	0.061
Istituto Comprensivo	-1.148***	0.043	-1.040***	0.040
Private school	-1.443***	0.098	-1.568***	0.088
Variance in the ESCS index between classes (ESCS_Var_Within) the year before (instrument)	0.160***	0.002	0.162***	0.002
Constant	7.101***	0.328	8.620***	0.307
R^2	0.189		0.205	

Annex B

Table B.1. Comparing the entire population of 6th graders with those for whom information about prior achievement (test score in grade 5) is available

	Entire population		Subsample	
	Mean	s.d.	Mean	s.d.
(selected) Student-level variables				
Achievement in Reading	65.24	16.16	66.44	15.13
Achievement in Mathematics	46.16	17.97	48.17	17.60
Female student	0.489		0.497	
1st generation immigrant	0.055		0.044	
2nd generation immigrant	0.049		0.049	
Late-enrolled student	0.073		0.030	
Socioeconomic background (index ESCS; mean=0, stdev=1)	0.137	1.045	0.204	1.032
Student who does NOT live with both parents	0.137		0.128	
(selected) Class-level variables				
Class-average socioeconomic background (index ESCS)	0.136	0.553	0.163	0.501
Proportion of 1st generation immigrants in the classroom	0.049		0.054	
Proportion of 2nd generation immigrants in the classroom	0.043		0.046	
Proportion of Late-enrolled students in the classroom	0.064		0.063	
Proportion of students who took the test	0.943		0.951	
(selected) School-level variables				
Number of classrooms in the school	6.482	3.079	6.231	3.053
Number of students in the school	149.638	77.554	143.501	76.603
Average number of students per class, in the school	22.695	2.976	22.652	2.954
School located in Northern-West Italy	0.253		0.256	
School located in Central Italy	0.177		0.181	
School located in Southern Italy	0.391		0.313	
Private school	0.029		0.030	
ESCS_Var_Within	8.758	8.290	7.953	7.306

Notes. Standard deviation is reported only for continuous variables.

Table B.2. Comparing the results between preferred specification (restricted sample) and an analysis with all students

	All students		Restricted sample (preferred specification)	
	Reading (a)	Math (b)	Reading (c)	Math (d)
Student-level characteristics	b-coeff	b-coeff	b-coeff	b-coeff
Prior achievement (grade 5)			0.487***	0.500***
			(97.76)	(102.59)
Female student	0.080***	-0.081***	0.074***	-0.064***
	(54.99)	(-52.49)	(42.98)	(-37.21)
1st generation immigrant	-0.107***	-0.045***	-0.055***	-0.016***
	(-51.71)	(-25.31)	(-24.54)	(-8.82)
2nd generation immigrant	-0.073***	-0.043***	-0.053***	-0.026***
	(-43.12)	(-26.73)	(-28.92)	(-16.06)
Early-enrolled student	-0.011***	-0.010***	-0.009***	-0.006***
	(-8.38)	(-7.40)	(-4.83)	(-3.62)
Late-enrolled student	-0.146***	-0.107***	-0.042***	-0.025***
	(-81.77)	(-69.51)	(-17.47)	(-13.12)
Socioeconomic background (index ESCS; mean=0, stdev=1)	0.244***	0.227***	0.155***	0.136***
	(136.79)	(122.57)	(66.33)	(61.64)
Student who does NOT live with both parents	-0.036***	-0.043***	-0.024***	-0.027***
	(-23.69)	(-28.54)	(-13.53)	(-16.42)
Student who has siblings	-0.027***	0.000	-0.016***	0.004*
	(-18.96)	(0.11)	(-9.43)	(2.15)
Classroom-level characteristics	b-coeff	b-coeff	b-coeff	b-coeff
Cheating propensity	0.166***	-0.168***	0.084***	-0.120***
	(53.61)	(-45.88)	(19.71)	(-30.52)
Classroom selected to be part of the "controlled" sample	0.005*	-0.008**	0.007**	-0.008**
	(2.06)	(-2.83)	(2.60)	(-2.80)
Class-average socioeconomic background (index ESCS)	0.003	0.011*	0.016**	0.019***
	(0.54)	(1.98)	(2.92)	(4.62)
Proportion of females in the classrom	-0.003	0.009**	-0.002	0.010**
	(-1.16)	(2.68)	(-0-50)	(2.68)
Proportion of 1st generation immigrants in the classroom	0.012***	-0.003	0.014***	0.005
	(3.95)	(-0.89)	(3.64)	(1.28)
Proportion of 2nd generation immigrants in the classroom	0.013***	0.007	0.005	-0.003
	(4.26)	(1.71)	(1.70)	(-0.86)
Proportion of Early-enrolled students in the classroom	0.013***	0.000	0.009*	-0.002
	(3.50)	(-0.07)	(1.97)	(-0.42)
Proportion of Late-enrolled students in the classroom	-0.029***	-0.027***	-0.014**	-0.013**
	(-8.52)	(-7.26)	(-3.47)	(-3.25)
Proportion of disabled students in the classroom	-0.005	0.000	0.001	0.006
	(-1.93)	(0.04)	(0.27)	(1.59)
Number of students in the classroom	0.029***	0.017***	0.024***	0.016**
	(7.96)	(4.10)	(5.23)	(3.17)
Proportion of students who took the test	0.063***	0.023***	0.033***	0.010*
	(15.91)	(5.65)	(7.99)	(2.54)
Class with *"tempo-pieno"*	-0.001	0.001	-0.003	0.000
	(-0.56)	(0.31)	(-0.78)	(0.03)

	All students		Restricted sample (preferred specification)	
	Reading (a)	Math (b)	Reading (c)	Math (d)
School-level characteristics	b-coeff	b-coeff	b-coeff	b-coeff
Number of classrooms in the school	-0.002	-0.018	-0.016	-0.047
	(-0.07)	(-0.48)	(-0.40)	(-1.17)
Number of students in the school	0.018	0.032	0.039	0.066
	(0.56)	(0.82)	(0.92)	(1.56)
Average number of students per class, in the school	-0.003	0.002	-0.012	-0.011
	(-0.41)	(0.19)	(-1.26)	(-1.07)
School located in Northern-West Italy	0.018***	0.006	0.020***	0.004
	(5.79)	(1.32)	(5.22)	(0.83)
School located in Central Italy	-0.010**	-0.036***	-0.021***	-0.040***
	(-2.78)	(-7.67)	(-4.73)	(-8.78)
School located in Southern Italy	-0.068***	-0.103***	-0.070***	-0.082***
	(-8.62)	(-10.78)	(-6.25)	(-7.34)
School located in PON area	-0.049***	0.021*	-0.052***	0.005
	(-6.29)	(2.17)	(-4.85)	(0.47)
Istituto Comprensivo	0.001	0.012	0.003	0.004
	(0.23)	(1.93)	(0.36)	(0.57)
Private school	-0.025***	-0.015***	-0.021***	-0.016**
	(-6.53)	(-3.36)	(-4.05)	(-2.94)
Variance in the ESCS index between classes (ESCS_Var_Within)	-0.024	0.023	-0.050*	-0.019
	(-1.16)	(0.85)	(-2.20)	(-0.77)
N	423,513	424,087	221,224	247,351
R^2	0.220	0.161	0.382	0.360
adj. R^2	0.220	0.161	0.382	0.360

Notes. *** is statistically significant at 1%; ** 5%, * 10%. p-values in parentheses.
ESCS_Var_Within is instrumented with ESCS_Var_Within (t-1).
Robust standard errors are clustered at school-level.

Annex C

As discussed extensively in the paper, we believe in a sorting mechanism that is based on students' SES more than their ability; and however we consider them not only equivalent, but also likely to happen together.

In this annex, however, our aim is to show that our main results are unchanged when considering the alternative sorting force in action. The theoretical rationale is to look at the results obtained in test scores at grade 5, and assuming that they reflect the same kind of ability that is valued and judged by the teachers of primary schools; in other words, we assume that those students who obtained higher scores in INVALSI tests at grade 5 are also those that the primary school's teachers report as better students to junior secondary schools' teachers responsible for composing the classes. In these circumstances, if a school decides to "sort by ability", then we should observe a great structural differentiation between classes in terms of prior achievement, with best students in the same class as well as the worst, etc. To the extent to which prior achievement is related to socioeconomic status (SES), the resulting sorting is similarly based on ability and SES.

Operationally, we calculated the within-school (between-classes) variance when considering Prior Achievement as the focal variable (the indicator was named $PriorAchievement\_Var\_Within_k$). The paper clearly described that the information about prior achievement is, unfortunately, not available for almost 50% of the students; consequently, we opted for a selection procedure that would guarantee that $PriorAchievement\_Var\_Within_k$ is calculated only for those schools where the proportion of students for which the information is available is high enough (the threshold was set at 75%, in other words, only those schools for which we have prior achievement for at least 75% of the students were analysed). At the end of this further data restriction, we have a sample of around 140,000 students for reading scores (33% of the original population) and 180,000 for mathematics (42%), in 2,056 schools. At this stage, we first certify that the pairwise correlation between the two variables, although not high in magnitude (around 0.19) is statistically significant at 1% conventional level. The figure C.1

graphically illustrates the relationships between the two indicators; the table C.2. cross-tabulates this relationship.

Then, we estimated the follow two alternative specification of the EPF, as a sensitivity test for our baseline model:

$$Y_{ijkt} = \alpha_0 + \alpha_1 X_{1ijkt} + \alpha_2 X_{2jkt} + \alpha_3 X_{3kt} + \left[\alpha_p X_{pkt}\right] + \varepsilon_{ijkt} \qquad \text{(C1)}$$

where X_{pkt} is PriorAchievement_Var_Within$_k$ as calculated at school level, and the other variables are as in (2). The objective is to check that coefficients' estimates are not too different from those obtained in the baseline specification, and most importantly to see if the results obtained for the variable of interest (PriorAchievement_Var_Within$_k$) are coherent with those reported for our preferred indicator. Two cautions must be expressed here: as we do not have adequate instruments for this variable, we estimated two different models with two different underlying assumptions: (i) one without any instrument (which results are likely to be affected by endogeneity between sorting and student achievement), and one with ESCS_Var_Within$_{(t-1)}$ as instrument, and the reliability of the results in this last case critically relies on the assumption that the same underlying phenomena are behind "sorting by ability" and "sorting by SES" (this is indeed our main assumption in this paper). The results are reported in the table C1. Two main facts must be noticed, and are both good new for the robustness of the results obtained through the empirical analyses of this paper. First, all the coefficients about student, class and school levels are estimated as practically identical to the baseline ones reported in table 4 – the only notable exception is that prior achievement seems to have a slightly higher effect (around 0.52/53 s.d. instead of 0.49/50). Second, the effect of PriorAchievement_Var_Within$_k$ is very similar in magnitude and sign to that estimated for ESCS_Var_Within$_k$, even though in the former case it does not gain statistical significance (it does in the model which does not use IV, but we tend to believe less to this specification). Summarizing: the variable measuring "sorting by ability" is correlated with our preferred indicator of "sorting by SES"; when included in the empirical analyses, the use of one of one of these two indicators is quite interchangeable and do not alter the main results, so it can be assumed that the underlying forces at action are similar.

Table C.1. The results when using PriorAchievement_Var_Within$_k$ instead of ESCS_Var_Within$_k$

	Reading		Math	
	(a) No IV	(b) IV	(c) No IV	(d) IV
Student-level characteristics				
Prior achievement (grade 5)	0.520***	0.522***	0.536***	0.536***
	82.38	*76.82*	*93.66*	*89.79*
Female student	0.076***	0.076***	-0.061***	-0.062***
	37.45	*36.67*	*-31.07*	*-30.57*
1st generation immigrant	-0.060***	-0.061***	-0.018***	-0.017***
	-22.65	*-22.06*	*-8.59*	*-7.98*
2nd generation immigrant	-0.056***	-0.057***	-0.026***	-0.027***
	-26.09	*-25.58*	*-14.29*	*-14.21*
Early-enrolled student	-0.008***	-0.008***	-0.004	-0.004*
	-3.63	*-3.64*	*-1.95*	*-2.03*
Late-enrolled student	-0.039***	-0.039***	-0.022***	-0.023***
	-14.13	*-13.86*	*-10.26*	*-9.99*
Socioeconomic background (index ESCS; mean=0, stdev=1)	0.150***	0.149***	0.129***	0.129***
	52.87	*50.30*	*50.20*	*48.37*
Student who does NOT live with both parents	-0.024***	-0.023***	-0.027***	-0.027***
	-11.36	*-10.71*	*-14.15*	*-14.01*
Student who has siblings	-0.014***	-0.014***	0.003	0.003
	-7.08	*-6.86*	*1.64*	*1.68*
Classroom-level characteristics				
Cheating propensity	0.069***	0.068***	-0.091***	-0.092***
	16.16	*15.63*	*-20.34*	*-13.17*
Classroom selected to be part of the "controlled" sample	0.007*	0.008*	-0.008*	-0.009*
	2.32	*2.36*	*-2.36*	*-2.44*
Class-average socioeconomic background (index ESCS)	0.003	0.002	0.022***	0.020**
	0.46	*0.33*	*3.81*	*3.03*
Proportion of females in the classrom	-0.007	-0.008	0.010*	0.010*
	-1.73	*-1.88*	*2.27*	*2.07*
Proportion of 1st generation immigrants in the classroom	0.013**	0.013**	0.011*	0.007
	2.95	*2.76*	*2.24*	*1.29*
Proportion of 2nd generation immigrants in the classroom	0.005	0.004	-0.002	-0.002
	1.28	*1.06*	*-0.41*	*-0.57*
Proportion of Early-enrolled students in the classroom	0.002	0.004	-0.005	-0.006
	0.46	*0.89*	*-0.94*	*-1.00*
Proportion of Late-enrolled students in the classroom	-0.004	-0.003	-0.012**	-0.013*
	-0.90	*-0.52*	*-2.58*	*-2.40*
Proportion of disabled students in the classroom	-0.002	-0.002	0.010*	0.011*
	-0.40	*-0.37*	*2.49*	*2.30*
Number of students in the classroom	0.026***	0.025***	0.014*	0.013*
	4.69	*4.41*	*2.34*	*2.11*
Proportion of students who took the test	0.022***	0.023***	0.012**	0.015**
	4.64	*4.67*	*2.70*	*2.80*
Class with *"tempo-pieno"*	-0.002	-0.002	0.000	-0.001
	-0.52	*-0.47*	*-0.06*	*-0.17*

	Reading		Math	
	(a)	(b)	(c)	(d)
	No IV	IV	No IV	IV
School-level characteristics				
Number of classrooms in the school	0.006	0.009	-0.016	0.006
	0.13	*0.17*	*-0.32*	*0.10*
Number of students in the school	0.010	0.009	0.022	-0.015
	0.19	*0.17*	*0.40*	*-0.20*
Average number of students per class, in the school	-0.005	-0.012	-0.008	0.016
	-0.50	*-0.76*	*-0.68*	*0.43*
School located in Northern-West Italy	0.020***	0.020***	0.004	0.004
	4.32	*4.10*	*0.80*	*0.66*
School located in Central Italy	-0.017**	-0.016	-0.040***	-0.049***
	-3.05	*-1.64*	*-7.27*	*-4.01*
School located in Southern Italy	-0.070***	-0.070***	-0.077***	-0.086***
	-5.87	*-4.28*	*-5.97*	*-4.38*
School located in PON area	-0.043***	-0.044***	0.008	0.000
	-3.83	*-3.67*	*0.63*	*-0.02*
Istituto Comprensivo	0.015	0.015	0.003	0.000
	1.76	*1.66*	*0.27*	*0.02*
Private school	-0.012	-0.011	-0.020**	-0.021**
	-1.78	*-1.61*	*-3.02*	*-2.95*
Variance in the index of prior achievement between classes (PriorAchievement_Var_Within)	-0.027***	-0.043	-0.018**	0.053
	-3.47	*-0.78*	*-2.83*	*0.55*
Adjusted R2	0.404	0.406	0.383	0.380
N	149,910	143,180	180,185	172,447

Notes. *** is statistically significant at 1%; ** 5%, * 10%. p-values in italics.
In columns (b) and (d) PriorAchievement_Var_Within is instrumented with ESCS_Var_Within (t-1).
Robust standard errors are clustered at school-level.

Table C.2.

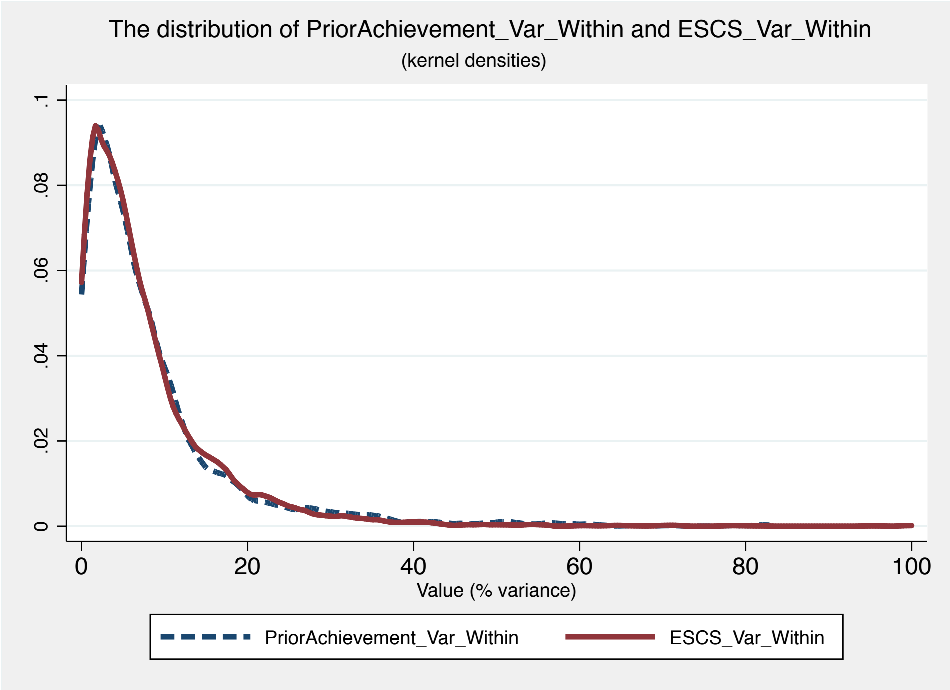
Cross-tabulation of ESCS_Var_Within and PriorAchievement_Var_Within

	PriorAchievement_Var_Within>15%	PriorAchievement_Var_Within<15%
ESCS_Var_Within>15%	50	115
	2.4%	*5.6%*
ESCS_Var_Within<15%	249	1642
	12.1%	*79.9%*

Notes. number of schools=2,056.

Figure C.1. ESCS_Var_Within and PriorAchievement_Var_Within: correlations

Panel A.



Panel B.