2011- 22 / november 20

Document de travail

PEER HETEROGENEITY, SCHOOL TRACKING AND STUDENTS' PERFORMANCES: EVIDENCE FROM PISA 2006

Michele Raitano
Sapienza University of Rome

Francesco Vona
OFCE-Sciences-Po

ofce

Peer Heterogeneity, School Tracking and Students' Performances: Evidence from PISA 2006

Michele Raitano
Sapienza University of Rome

Francesco Vona *
OFCE-Sciences-Po

Abstract

This paper analyses the interaction between school tracking policies and peer effects in OECD countries. Using the PISA 2006 dataset, we show that the linear peer effects are stronger and more concave-shaped in the early-tracking educational system than in the comprehensive one. Second, and more interestingly, the effect of peer heterogeneity goes in opposite directions in the two systems. In both student- and school-level estimates, peer heterogeneity reduces students' achievements in the comprehensive system while it has a positive impact in the early-tracking one. For late tracking countries, this result appears driven by pupils attending vocationally-oriented programs. Finally, peer effects are stronger for low ability students in both groups of countries.

JEL Codes: I21, I28, J24.

Keywords: peer heterogeneity, peer effects, schooling tracking, educational production function.

* Corresponding author: francesco.vona@sciences-po.fr. We wish to thank Maurizio Franzini, Edoardo Di Porto and seminar participants at IARIW General Conference in San Gallen, at EALE conference in Paphos and at the department of Public Economics of Sapienza University for useful comments. Francesco Vona acknoledges financial support of the POLHIA project (Seventh Framework Program of the European Commission).

1. Introduction

The quality of the educational system is commonly recognized to affect economic growth and the equalization of students' opportunities. International assessment programs (i.e. PISA, TIMMS, PIRLS) constitute a unique instrument to analyse the determinants of educational quality in a comparative perspective, circumventing problems of skill comparability. With the aim of explaining the failure of a resource-based approach to schooling production, shown by the lack of correlation between educational inputs and outputs (Hanushek 1986, 2003), several studies emphasize the institutional dimension of the educational system and, specifically, the role played by school autonomy and well-designed accountability practices (e.g. Woessmann et al. 2009). In these studies, much less attention has been devoted to the social aspects of the learning environment in terms of peer composition, social context and broad neighbourhood effects.

The importance of peer variables in explaining differences in educational achievements has been emphasized by a large theoretical literature[1], whereas at the empirical level the identification of peer effects remains cumbersome as exhaustive determinants of students' sorting to schools are often unavailable to the econometrician (Nechyba 2006). However, since peer group composition depends on policies affecting sorting, exploiting cross-country differences in these policies can provide interesting insights to better understand the way in which the peer group composition affects educational outcomes. Among sorting policies, school tracking is likely to be particularly important as it determines the allocation of students to schools offering completely different programs. Following the empirical strategy presented in section 2, the main novelty of this paper consists in extending current cross-country analyses on the determinants of students test scores inquiring how the interplay between peer composition and school tracking policies affects students' outcomes.

It is well understood that tracking policies, both between and within-schools[2], play a key role in determining the sorting of students from different background and abilities (Epple et al. 2003, Brunello and Checchi 2007). Especially early school tracking between vocational and academic-oriented schools creates relatively more homogeneous learning environment along observable (e.g. parental background) and students' characteristics that are often unobservable (ability, motivation, attitudes, etc.). In Germany, for instance, low ability students are sorted into the vocational track increasing the peer-group homogeneity along an unobservable dimension (Dustmann 2004, Checchi and Flabbi 2007). By the same token, since less-advantaged students have a higher likelihood of being enrolled in the vocational track, the few who enter the academic track are probably particularly motivated and/or talented. An early differentiation of educational programs often implies a substantial diversification of learning patterns, i.e. more or less oriented towards practical and theoretical knowledge, and hence should lead to an increase the degree of heterogeneity

1. See De Bartolome (1990), Benabou (1996), Fernandez and Rogerson (1996), Hoxby (2000), Lazear (2001), Durlauf (2004).
2. Within-school tracking consists in grouping students into different classes according to their ability and prevails in Anglo-Saxon countries; between-schools tracking assign students to completely different segments of the education process at early ages, generally offering general or vocational programs such as in Germany and in many central European countries (Brunello and Checchi 2007).

between- rather than within-schools. Finally, comprehensive and early-tracking systems are inspired by different educational philosophies and hence differ in many other aspects affecting students' sorting, such as the number of slots in vocational programs, the share of students tracked by ability, school's admission policies.

With these premises in mind, one should expect that, for a given level of observable peer heterogeneity, schools in early tracking countries are on average characterized by a lower level of unobservable peer heterogeneity. Put it differently, it might be the case that the observable functional relationship between peer variables and achievements substantially differ in the two systems. This difference should become particularly evident when the relationship between peer group composition and final outcomes is nonlinear and changes slope depending on the effective level of peers. The presence of nonlinearities is essential to sketch out policy implications regarding the optimal degree of students' heterogeneity. Were the peer-achievement relationship linear, sorting policies would not have an effect on the aggregate level, but only on the distribution, of human capital[3].

From a theoretical standpoint, peer heterogeneity should affect educational achievements in two ways. On the one hand, having more homogeneous classes implies similar cognitive levels and the sharing of common behavioural codes, so less teaching efforts devoted to equalize students skills. On the other hand, in heterogeneous classes various types of externalities might arise: disruptive due to the presence of students with a particular bad attitude (Lazear 2001), or positive knowledge spill-over from good students to average and/or bad ones (Durlauf 2004). Overall, which effect tends to prevail is an open empirical question that the interaction with school tracking policies can contribute to answer by offering an exogenous variation in the degree of unobservable heterogeneity for a given level of observable peers' characteristics.

Using PISA 2006 data on OECD countries, our empirical study seeks to answer two important questions, which, to the best of our knowledge, no study has jointly addressed so far: 1. Is the relationship between peer composition and achievement non-linear? 2. Is this relationship significantly affected by school tracking policies? The several robustness checks carried out in our empirical analysis lend support to a highly nonlinear peer-achievement relationship and further suggest that peer heterogeneity affects achievement in a completely different way depending on school tracking policies.

The paper is organized as follows. Next section briefly summarizes the literature to which our work is connected to and discusses the empirical strategy adopted. Section 3 describes the data and provides preliminary evidence on some different characteristics of school policies and students' performances in early and late tracking countries. Section 4 presents the main results based on student-level regressions and several robustness checks. Section 5 concludes.

3. To be sure, if peer inputs enter linearly in the educational production function, efficiency is unaffected by the reallocation of students to schools and classes; the opposite occurs in the non-linear case (Benabou 1996). These considerations have relevant policy implications as long as the optimal level of students' heterogeneity is very unlikely to emerge as a market outcome, given the several structural constraints shaping schooling choices: e.g. admission procedures, physical distance, early tracking policies, within-school ability tracking (de Bartolome 1992).

2. Related literature and empirical strategy

Our paper is related to the empirical literature on peer effects using international assessment surveys. As known, two issues render the identification of peer effects particularly cumbersome: 1. the reflexivity problem – i.e. the outcome of each student is affected by the mean outcome of the others and, at the same time, it affects the outcome of the others (Manski 1993); 2. the selectivity problem – i.e. sorting students to schools depends on several factors, such as residence, parental preferences, tracking and admission policies, school quality, that are difficult to jointly observe (e.g. Nechyba 2006).

Because for policy purposes it is relevant to correctly quantify peer effects, rather than to identify their exact source, most of the studies ignore the reflexivity problem and estimate a reduced-form specification of the relationship between student achievements and peers' characteristics where the peers' achievement is not included in the set of explanatory variables and peer variables are measured through parental background (e.g. parental education or occupation) and/or the share of the mates having certain characteristics (e.g. the share of immigrant or female classmates). Actually, since parental background is the single most important determinant of students' achievements, the sign of the coefficient associated with peer levels of family background is usually considered the "relevant" peer effect (Ammermueller and Pischke 2009).

In order to completely settle the selectivity problem, strict data requirements are needed, namely some exogenous variations that randomly assign students to schools (i.e. quasi-natural experiments), which are very difficult to be verified through data collected by international assessment surveys. However, using the PIRLS survey on pupils at the end of primary education, Ammermueller and Pischke (2009) exploit within-school variations in the class composition to identify linear peer effects under the assumption that the allocation of students and resources is random within the school[4]. When controlling for measurement errors, they find modestly large peer effects and a small difference in the estimated coefficient obtained in regressions accounting for the selection bias with respect to the OLS ones.

Unfortunately, a similar identification strategy is not feasible with the PISA dataset, because it does not provide classroom information (apart from the class size) and can at most consent to estimate broad school-level peer effects. However, for the main scope of this paper, the PISA dataset presents the main advantage of interviewing students aged 15, then above the usual age of first track in the large majority of OECD countries.

Absent data requirements for completely settling the selectivity problem, the few papers that estimate peer effects using PISA have resorted to a second-best identification strategy, based on the claim that the omitted variable bias is the most important source of selectivity bias. Accordingly, the selectivity bias is reduced by adding among the regressors school-level policies and further controls that are likely to affect the sorting of students to schools[5]. In a study on Denmark, Rangvid (2007) uses the parental academic and cultural interest to mitigate the selectivity problem. Quantile regressions allow her to find stronger peer effects for low ability students, whereas social heterogeneity, measured as the standard deviation of parental education, has an

4. A similar strategy is followed by Hoxby (2000), Hanushek et al. (2003), McEwan (2003).
5. It is worth noticing that not including peer variables leads to a misspecification of the correct educational production function, whereas including it using a second-best strategy leads at most to an incorrect quantification and (perhaps) interpretation of what the peer effects capture, i.e. broad contextual effects or true effects on learning (Zimmer and Toma 2000).

insignificant impact on students' achievements along the entire distribution of test scores. Similar results are found for Austria by Schneeweis and Winter-Ebmer (2007), which exploit the large number of possible school types that can be chosen by Austrian students to control for unobservable factors affecting sorting. The cross-country study of Vandenberghe (2002), based on the TIMMS dataset, displays a negative and significant effect of class heterogeneity (measured with the standard deviation of a composite index of family background) on students' attainments. In a slightly different paper with a focus on immigrants, Entorf and Lauk (2006) attempt to distinguish the peer effect of immigrants and natives in different tracking systems and show stronger peer effects in countries with an earlier tracking. For the U.S., Fertig (2003) follows a different strategy – instrumenting peer heterogeneity with proxies of the caring behaviour of parents at home and of school admission procedures – and finds a negative effect of heterogeneity (measured with the coefficient of variation in the achievement of schoolmates) on literacy scores.

In our empirical analysis, we mainly follow the omitted variable argument used by Rangvid (2007) and Schneeweis and Winter-Ebmer (2007). To be more precise, we extend the common specification of the schooling production function used in the literature on PISA scores (e.g. Fuchs and Woessmann 2007) including peer variables and their interaction with school tracking policies:

$$A_{isc} = \alpha + \beta X_c + \gamma X_s + \delta X_i + \theta \overline{X}_s + (\mu + \varphi D_{T=1}) f(\overline{FB}_s) + u_s + u_c + u_{is} + u_{isc} \qquad (1).$$

The achievement of student i in school s and country c (A_{isc}) is the resultant of a vector of individual (X_i, including family background), school (X_s), country (X_c) controls to which we add school compositional variables: the share of schoolmates with certain characteristics, such as gender, immigrant status, grade, etc. ($\overline{X_s}$), and a flexible functional form to capture the influence of peer characteristics related to family background $f(\overline{FB}_s)$.

The flexible functional specification is required as – apart from a linear peer effect summarized by the mean schoolmates' family background – non-linear peer effects can be either captured by the mean peer squared, or by the peers' standard deviation, or by both. Specifically, the square of the mean captures the concavity in the relationship between student's achievement and peer characteristics, whereas the standard deviation shows the extent to which this concavity is associated to a greater school heterogeneity[6]. To capture differential peer effects depending on school tracking, we also introduce an interaction term between peers and a dummy equal 1 if the country tracks students earlier than 13.

Finally, the error is decomposed in a country effect u_c, a school effect u_s and a correlated school-individual effect u_{is} plus an independent error term u_{isc}. The school-individual interaction might be a source of selectivity bias as it is correlated with both peer variables and achievements. It is worth recalling that this bias hinges upon unobservable variables affecting students' sorting. Under the assumption that these unobservable variables are mainly captured by

6. A specification with both the square of the mean peer and the peers' standard deviation is used by Vandenberghe (2002) and Zimmer and Toma (2000) for similar reasons. Interestingly, literature on the environmental Kuznets curve often includes both the income per capita squared and an index of inequality to jointly assess the curvature of the emission-income relationship and to investigate the scope for income redistributions favouring emission reduction (see, e.g., Ravallion et al. 2000).

other socio-demographic school-composition variables (e.g. the share of immigrants or of students enrolled in vocational programs within the school) and school characteristics, peer effects appear correctly assessed using "school-level clustering-robust" OLS regressions and controlling for countries' characteristics.

In particular, among the many school variables we can control for (see tab. A1), we have information on two aspects that are strongly connected with students' sorting: school admission procedures and school types (private vs. public). In the robustness checks, we also add pseudo-school fixed effects – built as the quintile of the distribution of the school average family background in each country –, dummies for the evaluation procedures followed by the school and proxies of ability tracking within the school. All these variables should mitigate the selectivity problem as they are information upon which parents select schools. For instance, pseudo-school fixed effect can be seen as a proxy of what kind of school environment parents expect to find in a certain school rather than in another, whereas the share of pupils enrolled in vocational programs represent a proxy of school type and quality.

As a refinement of our empirical strategy, school-level regressions allow to improve the identification of the peer heterogeneity. In particular, averaging by school equation (1) we obtain an expression where the correlated term school-individual is indistinguishable from the school fixed effect u_s. More precisely we obtain:

$$\overline{A}_{sc} = \alpha + \beta X_c + (\theta + \gamma + \delta)\overline{X}_s + (\mu + \varphi D_{T=1})SD(\overline{FB}_s) + u_s' + u_c \quad (2).$$

In eq. 2 the linear peer effect is indistinguishable from the family background effect, which was one of the individual characteristics in equation 1. So, only the standard deviation of peer's background can be identified, but in a better way; specifically, the estimation bias is further mitigated by taking the average of the unobservable individual characteristics within the school and hence reducing the effect of outliers on the estimated coefficients. As a result, the impact of background heterogeneity is more likely to be correctly assessed and the functional relationship between peer heterogeneity and achievement correctly specified if the sign and the significance of the associated estimated coefficient remain substantially unchanged once moving from individual- to school-level estimates.

As a final caveat, school tracking itself should contribute to the identification of peer effects. Because tracking represents an additional constraint for the sorting process, less unobservable heterogeneity should characterize schools in the early tracking system. All else equal, the differential impact of peer heterogeneity on outcomes in the two systems gives us additional information on the shape of the effective peer-achievement relationship.

3. Dataset and descriptive statistics

In our empirical analysis we use the PISA 2006 survey that covers all OECD countries and has a target population of 15-year-old students, regardless of the grade they currently attend. The PISA dataset records several variables both at the school and at the student level (see table A1 for a full list)[7]. More specifically, it contains detailed information on student's home background, school resources and a wide range of

7. Since the PISA dataset does not provide information at the class-level, we are not able to infer if possible peer effects come from an actual interaction between students rather than from a broad contextual effect.

institutional variables capturing the degree of school autonomy, accountability practices and variables affecting the students' sorting to schools. A rich set of basic individual controls is also available, whereas policy variables at the national level are usually integrated by other datasets (OECD and UNESCO, see tab. A1). Since PISA-2006 is mainly focussed on science, we consider the achievement in science as our main dependent variable and then checks whether results change for reading and math.

Some of the variables considered here are indexes built by PISA experts in order to summarize questions on specific school or individual characteristics. For instance, the degree of autonomy in managing resources at the school level is captured either by a vector of highly interdependent dummies (autonomy in within-school allocation, in hiring and firing teacher, etc.) or by a synthetic indicator built upon these dummies (see table A1 for details). The same holds for indexes of school resources and family background[8].

The synthetic background index built by the OECD and called the Economic Social Cultural Status "escs" is our preferred measure of family background as it combines various dimensions shaping the impact of family characteristics on the student's attainment: highest parental years of education ("pared"), the highest occupational level quantified with the index of occupational status ("hisei"; Ganzeboom et al. 1992), the "number of books at home" and the resources available at home to study (the "homepos" index). Indeed, by construction, all these variables are highly correlated with the escs.

As natural measures of peer variables, we take the average level of the schoolmates' escs (net of the individual one) for the linear peer effect, whereas possible nonlinearities in the peer-achievement relationship are captured by the squared average escs and by the standard deviation of students' escs. Moreover, in order to partially account for cross-country differences in the sorting process, we normalize the standard deviation at the school level with the one at the country level[9]. Finally, other variables capturing the socio-economic school environment are included as explanatory variables (i.e. the shares of females, immigrants, students speaking a foreign language at home or attending a vocational program).

The classification of countries into early tracking and late tracking constitutes another important issue to address. Similarly to Hanushek and Woessmann (2006), we measure between-school tracking policies with a 0/1 dummy variable that intends to capture not only the direct effect of tracking on students' outcomes, but also broader differences in the design of the educational system. Differently from Hanushek and Woessmann (2006), we assume that tracking displays its effects with a lag of two years and hence a country tracks "early" if students choose before they are 13 years old (see table 2)[10].

Our sample includes students interviewed by PISA 2006 in all OECD countries, apart from France, as school variables are not collected in PISA 2006, and Mexico and Turkey as they display an average escs that is a full

8. See OECD 2009 and the PISA 2006 Technical Report for a detailed explanation about how these indexes have been computed.

9. This is because a high heterogeneity at the school level can be due to a high heterogeneity in the country rather than to randomly sorting students to schools. However, all results are robust to the inclusion of the 'no-normalized' standard deviation of backgrounds at the school level. Results using this further measure of heterogeneity are available upon request by the authors.

10. Literature provides several alternative measures of school tracking policies: Hanushek and Woessmann (2006) uses the age of the first tracking choice, Ammermueller (2005) the number of tracks experienced by the student before enrolling in upper secondary education, Waldinger (2006) the minimum school grade where a significant share of students is allocated in different tracks. Our results are robust to these different ways of measuring tracking.

standard deviation below the OECD average. Following many studies using PISA surveys, we also exclude from the sample those very few students (around 1,350) enrolled in grades lower than 8 or higher than 11. Finally, as we intend to analyze the effect of social interactions at school, we follow Rangvid (2007) and restrict the sample to schools with at least15 interviewed students, i.e. we drop schools with less than 15 interviewed students (around 6,400 students). Our final sample includes 202,817 students clustered in 6,728 schools (see table 1).

Table 1 shows the summary statistics of PISA science scores for the countries considered in the econometric analysis. To emphasize the differences between the early tracking and the comprehensive system, we report summary statistics for each group separately. Country means and standard deviations in science scores are, respectively, higher and lower in early tracking countries than in late tracking ones. More importantly, the dispersion of average school performances is significantly higher in countries tracking students earlier. Whereas this result was theoretically expected, the incidence of tracking on school heterogeneity is substantial and very large. In fact, between-schools variation in test scores explains more than half of the overall variation in the early-tracking countries, compared to roughly a quarter in late tracking ones.

Differences between the two systems are much broader than this, as it appears evident looking at the mix of other policies affecting students' sorting (see table 2). In early tracking countries, schools consider less frequently students' residence as a prerequisite or a high priority for enrolment, whereas a major role is assigned to students' performances in previous educational steps. In some late tracking countries (especially in Anglo-Saxon ones), the large majority of schools groups students into different classes according to their abilities, at least in some subjects. Moreover, early tracking countries display a significantly higher fraction of students enrolled in the vocational stream compared to the comprehensive group. In sum, this evidence seems to suggest that differences in tracking are accompanied by a different mix of complementary sorting policies and justify our choice of using a dummy to capture it.

On the contrary, no differences among the two groups of countries emerge with respect to the distribution of the index of economic social condition status and to the share of foreign students (see table 3). The only difference concerns the country mean of escs which is higher in early tracking countries, due to the presence in the late tracking group of three outliers with very low average escs, i.e. Portugal, Spain and Poland. By excluding these three countries, the weighted average escs in late tracking countries increases up to 0.17 (a value only slightly below the one in early-tracking countries), while the negative cross-country correlation between the mean and the standard deviation of escs decreases from -0.511 to a much lower and acceptable value of -0.157.

4. Econometric analyses

According to the empirical strategy outlined in section 2, we analyze the determinants of students' scores in science. We first estimate a basic specification of eq. 1 without peer effects as done in the existing literature on cross-country studies (Fuchs and Woessmann 2007). Moving from this benchmark, we estimate an augmented specification with peer variables and then analyse in detail the role of school tracking policies: first interacting peer variables and the "early tracking dummy" and then running separated regressions for the two groups of countries. Finally, we carry out school-level regressions (section 4.2) and quantile regressions, which allow to observe how the impact of peer variables change along the distribution of students' abilities (section 4.3).

In all specifications of the schooling production functions estimated in this section, four blocks of explanatory variables are included (see table A1 for details).

1. *Students' Characteristics (only in student-level regressions)*: age, sex, grade, two dummies on the kind of course attended (upper/lower secondary and vocational program) and two dummies on the migration status (foreign language at home and immigrant), plus several proxies of family background (i.e. the single components of the escs index): the highest parental education (in years) and occupational status, the OECD variable summarising in a quantitative index the family "home possessions" (OECD 2009) and dummies capturing the "number of books at home".
2. *School Resources and Institutions*: class size, school location, two quantitative indexes concerning school responsibility for resources' allocation ("respres") and for curriculum and internal assessment ("respcurr")[11], two dummies for private and public schools, two dummies for the admission procedures, signalling if residence or abilities are a priority or a prerequisite for being enrolled in that school. Note that variables on admission procedures seem particularly well suited in order to reduce the omitted variable bias due to a non-random assignment of students to schools.
3. *School Compositional Variables* (all calculated net of the individual one): the mean age and the shares of females, of "foreigners" (immigrants and people speaking a foreign language at home), of students attending vocational programs and our main peer variables based on three statistics on the schoolmates family background (proxied by the escs index), i.e. the mean, the square of the mean and the standard deviation.
4. *Country-level controls:* the share of students subjected to external evaluation and/or standard test in science[12], the age of tracking between different kinds of programs (general or vocational), the duration of and the pupils' share enrolled in pre-primary education, GDP per capita and the expenditures in secondary education. In some robust checks, country level controls are replaced by country fixed effects.

In robustness checks a fifth block of control variables is added:

5. *School Additional Controls*: pseudo school fixed effects (i.e. dummies for the quintiles of the country-specific distribution of the average parental escs which the school belongs to), controls about the degree of school competition and accountability practices internal to the school, two dummies capturing if there is an ability grouping internal to the school for some subjects or for all subjects (see table A1 for a detailed description of these additional controls).

11. In all regressions shown in this paper, we include the 'respres' and 'respcurr' indexes (see table A1 and OECD 2009) instead of the dummies about the single components of school autonomy and responsibility about resources and curricula, due to the several missing values characterizing each dummy. Replacing these dummies with the two OECD indexes, which by construction have much less missing values, does not alter our results.
12. The share of students subject to standardized external examination has been shown to be an important determinant of students' attainments (Fuchs and Woessmann 2007).

4.1 Student-level regressions

In student level regressions, three additional methodological adjustments are in order for improving the reliability of our estimates. First, missing values on certain individual characteristics are not randomly distributed, but appear related to family background and ability. As a result, dropping students with missing information on some variables could engender a sample selection bias. In order to copy with this issue, we impute individual missing values regarding family background (escs, pared, hisei and homepos variables, see table A1) and some individual characteristics (immigrant and foreign language) according to the usual methodology followed in the literature (Woessmann 2004). Second, the “school-level clustering-robust” linear regression method is always used in student-level regressions to estimate standard errors that recognize the schools as the basic unit of sampling in the survey (Woessmann 2004). Third, in order to obtain representative coefficient estimates from the stratified survey data, students’ sample weights are used in all estimations of students’ scores.

In the following tables we present estimated coefficients, standard errors and p-values only for peer variables, because the sign and the significance of the other variables included is consistent with main findings of the existing literature emphasizing the importance of school and country institutional features over school resources (see Fuchs and Woessmann 2007, Woessmann et al. 2009, Hanushek and Woessmann 2010).

In table 4, we present models that differ for the functional specification of peer effects chosen. The model ST-0 is the standard model considered by the literature on the students’ scores production function (Fuchs and Woessmann 2007), whereas the following models include school compositional and peer variables. The comparison between model ST-0 and the subsequent ones clearly confirms that the inclusion of schoolmates characteristics substantially increases the fitness of the estimations: the R^2 grows from 0.307 to values around 0.340[13]. This suggests that a penalty is associated both with the choice of including peer variables and probably incurring in an identification problem, on the one hand, and with the one of not including peer variables and of reducing the explanatory power of the regression, on the other.

As shown by model ST-1 of table 4, an improvement in the average escs of the schoolmates is associated with a better student’s achievement. The size of this effect is also significant and somehow larger than the one found by other studies (Ammermueller and Pischke 2009). In particular, a 1-standard deviation increase in the average escs leads to a 25 points increase in the student’s scores, that is a quarter of the standard deviation in PISA scores, normalized to 100[14].

In model ST-2, we add also the escs standard deviation to capture the influence of peer heterogeneity. While the sign and significance of the linear peer effect do not change, peer heterogeneity displays a negative but small and weakly significant association with achievements: a change in 1 standard deviation leads to a reduction in the average score of 1.35 points. This result is in line with the ones of the previous literature finding a small, but often insignificant, negative effect of heterogeneity on students’ performances (Hanushek et al. 2003, Rangvid 2007).

13. A R2 around 34 % is in line with the one found by the two studies using a large set of controls to reduce the omitted variable bias in the estimation of peer effects on PISA scores (Rangvid 2007; Schneeweis and Winter-Ebmer 2007).
14. One standard deviation of schoolmates’ average escs and of the standard deviation of escs amount, respectively, to 0.49 and 0.14.

The model ST-3 shows that the small negative influence of heterogeneity on scores does not hinge upon the particular shape of the schooling production function, which appears concave-shaped. The quadratic term is in fact highly significant and negative. These results remain unchanged if we include both the quadratic term of the mean escs and the standard deviation (model ST-4). Overall, mixing students negatively affects student's outcome, but keeping relatively more homogeneous schools around the average is better than having relatively more homogeneous schools spread over the entire support of the escs distribution.

In model ST-5 we consider the interplay between schoolmates composition, school tracking policies and students' performances. Hence, we include also the interactions of the mean and the standard deviation of the schoolmates' escs with early tracking dummy. We prefer to leave aside the interaction between the linear peer squared and the tracking dummy as it could render less transparent the sign of the interaction between peer heterogeneity and the tracking dummy. As expected, both interaction terms display a positive and highly significant estimated coefficient. In line with the theoretical literature (Brunello et al. 2007) and the empirical findings by Entorf and Lauer (2006), linear peer effects are higher in the early tracking system. The novel result is that more heterogeneous schools seem to exert an opposite impact in the two groups of countries: they improve students' performances in early-tracking countries, whereas they worsen achievements in late tracking ones.

Table 5 shows that this differential impact of heterogeneity in early and late tracking countries is confirmed by robustness checks analyzing the model ST-5 for math and read. Looking at model ST-5a in the same table, also the inclusion of additional school variables and of pseudo school fixed effects does not alter our main findings, in spite of the significant reduction in the sample size due to several missing values on these additional variables. Finally, it is worth noticing that the differential impact of peer variables depending on school tracking mainly concerns peer heterogeneity. In fact, in the tougher specification with the full set of school controls, the interaction between the early tracking dummy and the linear peer effect is not anymore significant.

Table 6A shows that our results remain unchanged when running models ST-1 and ST-2 separately for the two educational systems. First, linear peer effects are higher in early tracking countries: one standard deviation explains 28.5 and 22.9 PISA points in early and late tracking systems, respectively. Second, the coefficient of the peer heterogeneity term is significant with opposite signs in the two systems: early-tracking (+) and comprehensive (-). The magnitude of this effect remains modest as one standard deviation in the escs standard deviation leads, respectively, to a 2.1 increase and a 2.5 decrease in PISA scores. Finally, the R^2 is much higher in regressions concerning early tracking countries, perhaps reflecting the lower within-school unobservable noise.

In many schools, students can choose between vocational and non-vocational education. Moreover, the fraction of students enrolled in vocational programs substantially differs in the two groups of countries. To capture these important aspects of the two systems, we interact peer variables with the dummy about the kind of school program. Model ST-3 shows that this interaction term is statistically insignificant in early tracking countries, but it is highly significant in the comprehensive school system. In the late tracking system, the linear peer effect appears higher for students attending vocational courses, whereas higher heterogeneity reduces students' performances. As a result, the negative relationship between peer heterogeneity and performances in late tracking countries appears mainly driven by schools offering vocational programs, which are often of lower quality.

Adding to the base model pseudo-school fixed effects and additional controls does not alter results, but the significance of the heterogeneity coefficients in early tracking countries reduces at cut-off significance level of 82% (model ST-2a).

Finally, in both systems the relationship between schoolmates' family background and students' scores is concave-shaped and of similar size (model ST-3 in table 6A). This result might be driven by the fact that, in the comprehensive system, there are three outlier countries (Spain, Portugal and Poland; see table 3) with a very low average escs. If we exclude these three countries from the late tracking group, the linear peer substantially increases and the square of the average peer turns out being statistically insignificant (table 6B). This latter result reinforces our claims that school tracking heavily influence the shape of the peer-achievement relationship and that the scope for mixing students is much higher in the early tracking system.

4.2 School-level regressions

As discussed in section 2, school-level regressions – i.e. the mean score at the school level is the dependent variable – could allow to improve the identification of the peer heterogeneity. Recall that, in this case, only the standard deviation of peer's background can be identified, but perhaps in a better way because averaging individual unobservables reduce the influence of outliers.

Table 7 shows school-level estimates which are carried out always including school additional controls. In the pooled SC-1 model, a negative and significant impact of students' heterogeneity is confirmed[15]. Moreover, the differential impact of heterogeneity in the two systems is confirmed, both in the specification with the interaction terms and in separated regressions (models SC-2). Now, the size of the estimated coefficients of peer heterogeneity appears larger in comparison with student-level estimates: a change in one standard deviation of heterogeneity leads to a decrease of 4.7 school mean PISA points in late tracking countries and to an increase of 2.3 points in early tracking ones. As expected, because a large part of the performance variation across students has to be attributed to individual unobserved variables (e.g. their innate ability or learning motivation), when averaging variables at the school level the R^2 significantly increases and again it is still much higher in the group of early tracking countries.

4.3 Quantile regressions on students' scores

A final interesting exercise consists in running quantile regressions to condition the estimates of peer effects to an implicit distribution of student abilities. This exercise is particularly important here as the impact of early-tracking on sorting could mainly operate through unobservable student abilities. Estimating

15. A change in one standard deviation of the average escs index turns out to explain 41 out of the 100 points of the standard deviation in the student attainments. This is not surprisingly as long as, in school-level regressions, the average escs identifies both the peer and the individual background effect, which is usually the larger explanatory factor of student outcomes (e.g. Hanushek and Woessmann 2010).

model ST-2 of table 4 for each decile of the distribution of students' achievements, quantile regressions show that the linear peer effect is positive all along the conditional test score distribution and remarkably higher in early tracking countries (table 8 and figure 1). Moreover, similarly to previous studies (Schneeweis and Winter-Ebmer 2007, Rangvid 2007), this effect is slightly larger in lower deciles while the main differences between the two groups of countries pertain the effect of peer heterogeneity (table 8 and figure 2). In the early tracking group, its sign is significantly positive and slightly U-shaped along the entire test score distribution. In late tracking group, the effect of peer heterogeneity is significantly negative and slightly lower in upper deciles.

In sum, quantile regressions reinforce the previous finding in terms of a significant efficiency-enhancing effect of mixing students in early tracking systems, where individuals are probably more homogeneous in their unobservable features. Conversely, the picture in comprehensive systems is nuanced: on one hand, stronger peer effects at the bottom of the ability distribution would lead to support policies aimed at increasing background heterogeneity; on the other hand, a too high heterogeneity might turn out offsetting the efficiency-enhancing effect of mixing students.

5. Concluding remarks

This paper analyses the impact peer heterogeneity on students' performances in different tracking regimes. Unlike previous studies often focussing on a single country, the cross-country dimension of our analysis consents us to exploit the variability in school tracking policies to investigate how peer effects interact with this important aspect of the educational system. Our descriptive analysis reveals profound differences between the early tracking and the comprehensive system along several policies affecting students' sorting and lends further support to our main research focus.

We confirm previous studies showing that linear peer effects are heavier in the early tracking system. More interestingly, we show that the impact of peer heterogeneity changes sign depending on the school tracking policies. In the early-tracking system, peer heterogeneity has a positive impact on students' outcomes and the peer-achievement relationship is concave-shaped. In the comprehensive system, we do not find evidence of non-linear peer effects while higher peer heterogeneity negatively affects student achievements. However, the latter result appears driven by pupils attending vocationally-oriented programs. Our findings remain robust to several robust exercises, such as school-level regressions, the inclusion of several controls correlated with students' sorting and of pseudo school fixed effects.

All these findings point, as a possible explanation, to a different way in which the tracking system affects the sorting of students by unobservable characteristics. In the early tracking system, better students might put more efforts to signal their higher abilities and motivations sooner in order to enter the academic track. If this is the case, the unobservable degree of heterogeneity should be lower in early tracking countries and, hence, the scope for mixing students should take also this ability dimension into account. However, our analysis is not able to fully distinguish the effect of school tracking on sorting from the one related to heterogeneity in school quality, which is likely to be higher in the early tracking system. In fact, differences in school quality are only partially accounted for in our analysis using the share of students enrolled in vocational programs.

A final caveat is required to use these results for policy purposes. The significant impact of peer heterogeneity on students' performances is rather small both in the comprehensive and in the early-tracking system, hence favouring student mobility and the mixing of background might have a cost well-above the benefits in terms of efficiency. Also, the large variation in the factors affecting the students' sorting and selection, both within- and between-country, would require further analyses to obtain more limpid policy implications regarding the scope of policies aimed at mixing students of different backgrounds.

References

Ammermueller A. (2005), "Educational opportunities and the role of institutions", *ZEW Discussion Paper*, n. 05-44.

Ammermueller A., Pischke J. (2009), "Peer Effects in European Primary Schools: Evidence from the Progress in International Reading Literacy Study", *Journal of Labor Economics*, vol. 27: 315-348.

Benabou R. (1996), "Equity and effectiveness in human capital investment: the local connection", *Review of Economic Studies* 63: 237-264.

Brunello G., Checchi D. (2007), "Does school tracking affect equality of opportunity? New international evidence", *Economic Policy* 22: 781-862.

Brunello G., Ariga K., Giannini M. (2007), "The optimal timing of school tracking", in Peterson P., Woessmann L. (eds.), *Schools and the Equal Opportunity Problem*, Cambridge MA, MIT Press.

Checchi D., Flabbi L. (2007). "Intergenerational mobility and schooling decisions in Italy and Germany", *IZA Discussion Paper*, n. 2876.

De Bartolome C. (1990), "Equilibrium and Inefficiency in A Community Model with Peer Group Effects", *Journal of Political Economy* 98: 110-133.

Durlauf S. (2004), "Neighborhood effects", in Henderson J., Thisse J. (eds.), *Handbook of Regional and Urban Economics*, NewYork, Elsevier.

Dustmann C. (2004), "Parental background, secondary school track choice, and wages", *Oxford Economic Papers* 56: 209-230.

Entorf H., Lauk M. (2006). "Peer effects, social multipliers and migrants at school: An international comparison", *IZA Discussion Paper*, n. 2182.

Epple R., Newlon E., Romano R. (2002), "Ability tracking, school competition and the distribution of educational benefits", *Journal of Public Economics* 83: 1-48.

Fernandez R., Rogerson R. (1996), "Income Distribution, Communities and the Quality of Public Education", *Quarterly Journal of Economics* 111: 135-164.

Fertig M. (2003), "Educational Production, Endogenous Peer Group Formation and Class Composition: Evidence from the PISA 2000 Study", *IZA Discussion Paper*, n. 714.

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

Ganzeboom H., De Graaf P., Treiman D. (1992), "A standard international socio-economic index of occupational status", *Social Science Research* 21: 1-56.

Hanushek E. (1986)."The economics of schooling: production and efficiency in public schools", *Journal of Economic Literature* 24: 1141-1177.

Hanushek E. (2003), "The failure of Input-Based schooling policies", *Economic Journal* 113: F64-F98.

Hanushek E., Kain J., Markman J., Rivkin S. (2003), "Does peer ability affect student achievement?", *Journal of Applied Econometrics* 18: 527-544.

Hanushek E., Woessmann L. (2006), "Does Educational Tracking Affect Educational Inequality and Performance? Differences-in-Differences Evidence across Countries", *Economic Journal* 116: C63-C76.

Hanushek E., Woessman L. (2010), "The economics of international differences in educational achievement", *NBER Working Paper*, n. 15949.

Hoxby C. (2000), "Peer Effects in the Classroom: Learning from Gender and Race Variation", *NBER Working Paper*, n. 7867.

Lazear E. (2001), "Educational Production", *Quarterly Journal of Economics*116: 777-803.

Manski C. (1993), "Identification of endogenous social effects: The reflection problem", *Review of Economic Studies* 60: 531-542.

McEwan P. (2003), "Peer effects on student achievement: Evidence from Chile", *Economics of Education Review* 22: 131-141.

Nechyba T., (2006), "Income and Peer Quality Sorting in Public and Private Schools", in Hanushek E., Welch F. (eds.), *Handbook of the Economics of Education,* Vol. 2, New York, Elsevier.

OECD - Organisation for Economic Cooperation and Development (2009), *PISA 2006 Technical Report*, Paris.

Rangvid-Schindler B. (2007), "School composition effects in Denmark: quantile regression evidence from PISA 2000", *Empirical Economics* 32: 359-386.

Ravallion M., Heil M., Jalan J. (2000), "Carbon Emissions and Income Inequality", *Oxford Economic Papers* 52: 651-669.

Schneeweis N., Winter-Ebmer R. (2007), “Peer effects in Austrian schools”, *Empirical Economics* 32: 387-409.

Vandenberghe V. (2002), “Evaluating the magnitude and the stakes of peer effects analysing science and math achievement across OECD”, *Applied Economics* 34: 1283-90.

Waldinger F. (2006), “Does tracking affect the importance of family background on students’ test scores?”, University of Warwick, *mimeo*.

Woessmann L. (2004), “How equal are educational opportunities? Family background and students achievement in Europe and the United States”, *CESifo Working Paper*, n. 1162.

Woessmann L., Luedemann E., Schuetz G., West M., (2009). *School Accountability, Autonomy and Choice around the World*, Cheltenham, Edward Elgar Publishing.

Zimmer R., Toma E. (2000), “Peer Effects in Private and Public Schools Across Countries”, *Journal of Policy Analysis and Management* 19: 75-92.

Table 1: Distribution of science scores in OECD countries[1] in PISA 2006

	Mean	Standard deviation	Standard deviation of school means	Mean of school standard deviations	Between school variance in science scores (%)	Within school variance in science scores (%)	Number of sampled students	Number of sampled schools
Austria	518.4	92.1	74.3	62.5	57.0	43.0	4,648	155
Germany	523.1	94.9	78.9	61.0	59.9	40.1	4,570	203
Czech Rep.	520.5	91.9	62.7	65.3	57.8	42.2	5,516	200
Hungary	514. 1	79.7	57.9	53.8	70.4	29.6	4,245	136
Slovak Rep.	493.7	88.5	57.2	67.2	42.4	57.6	4,366	151
Belgium	512.5	96.2	78.1	63.6	52.3	47.7	8,724	258
Netherlands	525.8	92.2	79.5	51.9	58.9	41.1	4,835	182
Switzerland	515.1	95.9	59.2	72.6	34.2	65.8	11,181	402
Early tracking countries[2]	***520.0***	***93.5***	***73.6***	***61.9***	***51.3***	***48.7***	***48,085***	***1,687***
Luxembourg	486.3	94.1	54.6	76.5	29.2	70.8	4,532	29
Italy	479.0	90.9	68.6	63.4	52.1	47.9	21,012	696
Korea	522. 9	86.4	58.8	67.9	35.3	64.7	5,120	149
Greece	481.3	84.8	64.2	65.0	51.7	48.3	4,651	151
Ireland	508.6	91.2	47.4	81.5	17.0	83.0	4,557	163
Japan	532.0	96.4	77.1	68.8	48.5	51.5	5,910	180
Portugal	488.7	77.9	44.2	63.3	31.9	68.1	4,587	158
Australia	527.3	96.6	44.7	87.2	17.9	82.1	14,046	341
Canada	536.3	90.0	49.1	80.7	18.4	81.6	21,516	762
Denmark	495.0	89.3	38.3	81.1	15.4	84.6	4,363	190
Finland	564.0	81.8	23.9	78.8	5.8	94.2	4,674	151
Iceland	489.7	93.2	29.0	89.3	9.0	91.0	3,359	77
Norway	485.6	92.4	37.8	85.1	9.9	90.1	4,561	182
New Zealand	529.3	102.8	47.9	92.6	15.9	84.1	4,510	161
Poland	497.7	85.6	36.3	79.2	13.6	86.4	5,345	174
Spain	488.9	87.5	41.1	77.0	13.9	86.1	19,286	652
Sweden	502.9	90.7	34.4	84.7	12.0	88.0	4,318	172
UK	514.6	103.6	58.2	87.9	18.9	81.1	12,822	490
US	491.0	101.8	45.2	83.6	23.3	76.7	5,563	163
Late tracking countries[2]	***503.9***	***98.4***	***54.5***	***78.7***	***24.7***	***75.3***	***154,732***	***5,041***
OECD[1]	***506.4***	***97.9***	***60.1***	***74.9***	***31.0***	***69.0***	***202,817***	***6,728***

[1] Mexico, Turkey and France are not included. [2]Weighted averages by sample sizes. Source: elaborations on PISA 2006 data.

Table 2: Distribution of institutional variables in OECD countries[1] in PISA 2006

	Age of first track	Share of public schools	Share of schools not tracking by ability	Share of students enrolled in vocational programs	Share of schools considering students' records a priority for admission	Share of schools considering students' residence a priorityfor admission
Austria	10	0.87	0.60	0.44	0.69	0.22
Germany	10	0.89	0.59	0.00	0.40	0.65
Czech Rep.	11	0.91	0.35	0.43	0.45	0.20
Hungary	11	0.81	0.31	0.62	0.70	0.01
Slovak Rep.	11	0.88	0.25	0.46	0.50	0.17
Belgium	12	0.37	0.56	0.47	0.26	0.02
Netherlands	12	0.32	0.19	0.30	0.66	0.10
Switzerland	12	0.93	0.24	0.07	0.54	0.82
Early tracking countries[2]		***0.79***	***0.47***	***0.20***	***0.47***	***0.43***
Luxembourg	13	0.85	0.27	0.13	0.41	0.42
Italy	14	0.92	0.54	0.57	0.07	0.11
Korea	14	0.55	0.12	0.24	0.60	0.22
Greece	15	0.95	0.89	0.15	0.05	0.71
Ireland	15	0.40	0.02	0.02	0.02	0.42
Japan	15	0.74	0.44	0.24	0.87	0.20
Portugal	15	0.89	0.48	0.14	0.06	0.56
Australia	16	0.00	0.05	0.10	0.09	0.42
Canada	16	0.85	0.08	0.00	0.11	0.78
Denmark	16	0.63	0.16	0.00	0.03	0.55
Finland	16	0.96	0.49	0.00	0.04	0.75
Iceland	16	0.96	0.15	0.00	0.01	0.94
Norway	16	0.96	0.58	0.00	0.00	0.78
New Zealand	16	0.93	0.03	0.00	0.10	0.50
Poland	16	0.97	0.53	0.00	0.13	0.83
Spain	16	0.53	0.29	0.00	0.03	0.68
Sweden	16	0.88	0.24	0.00	0.01	0.58
UK	16	0.74	0.00	0.00	0.10	0.61
US	16	0.87	0.12	0.00	0.08	0.81
Late tracking countries[2]		***0.82***	***0.23***	***0.09***	***0.22***	***0.60***
OECD[1]		***0.82***	***0.26***	***0.11***	***0.26***	***0.57***

[1] Mexico, Turkey and France are not included. [2] Weighted averages by sample sizes. Source: elaborations on PISA 2006 data and Brunello and Checchi (2007) for age of first track.

Table 3: Distribution of escs and foreign students in OECD countries[1] in PISA 2006

	Country mean escs	Country escs standard deviation	Standard deviation of the mean escs by schools	Mean of the standard deviations of escs by schools	Standard deviation of the standard deviations of escs by schools	Country share of immigrant students	Country share of students mainly speaking a foreign language
Austria	0.23	0.83	0.46	0.69	0.13	0.12	0.10
Germany	0.32	0.92	0.48	0.79	0.15	0.17	0.13
Czech Rep.	0.07	0.75	0.37	0.66	0.09	0.02	0.02
Hungary	-0.01	0.90	0.55	0.71	0.11	0.02	0.01
Slovak Rep.	-0.10	0.88	0.48	0.73	0.13	0.00	0.14
Belgium	0.18	0.91	0.48	0.77	0.14	0.13	0.19
Netherlands	0.25	0.89	0.46	0.77	0.13	0.11	0.07
Switzerland	0.10	0.88	0.40	0.79	0.14	0.23	0.19
Early tracking countries[2]	***0.23***	***0.90***	***0.49***	***0.76***	***0.15***	***0.13***	***0.12***
Luxembourg	0.10	1.10	0.53	0.96	0.13	0.37	0.91
Italy	-0.05	0.97	0.50	0.84	0.13	0.04	0.13
Korea	-0.01	0.81	0.42	0.69	0.10	0.00	0.00
Greece	-0.09	0.95	0.52	0.80	0.13	0.05	0.03
Ireland	-0.01	0.85	0.41	0.75	0.11	0.08	0.06
Japan	-0.01	0.69	0.35	0.60	0.08	0.00	0.00
Portugal	-0.52	1.28	0.73	1.05	0.19	0.05	0.02
Australia	0.21	0.77	0.39	0.67	0.09	0.22	0.09
Canada	0.37	0.80	0.37	0.71	0.12	0.23	0.15
Denmark	0.30	0.89	0.36	0.82	0.14	0.08	0.07
Finland	0.25	0.79	0.28	0.74	0.11	0.02	0.02
Iceland	0.82	0.86	0.30	0.81	0.10	0.03	0.03
Norway	0.43	0.75	0.29	0.69	0.14	0.08	0.08
New Zealand	0.09	0.82	0.37	0.73	0.12	0.21	0.09
Poland	-0.31	0.86	0.41	0.76	0.11	0.00	0.01
Spain	-0.31	1.07	0.60	0.89	0.16	0.07	0.16
Sweden	0.23	0.78	0.31	0.72	0.14	0.11	0.09
UK	0.19	0.81	0.39	0.71	0.11	0.10	0.06
US	0.15	0.91	0.48	0.77	0.13	0.17	0.12
Late tracking countries[2]	***0.07***	***0.89***	***0.48***	***0.74***	***0.14***	***0.11***	***0.08***
OECD[1]	***0.09***	***0.89***	***0.49***	***0.74***	***0.14***	***0.11***	***0.09***

[1] Mexico, Turkey and France are not included. [2]Weighted averages by sample sizes. Source: elaborations on PISA 2006 data.

Table 4: Students performances in science in OECD countries[1]. OLS regressions[2, 3]

	ST-0	ST-1	ST-2	ST-3	ST-4	ST-5
		50.04	49.72	66.84	67.18	44.48
Average Escs		2.58	2.56	5.88	6.00	2.78
		0.000	*0.000*	*0.000*	0.000	*0.000*
			-9.47		-11.83	-18.57
Escs standard deviation			6.54		6.23	6.68
			0.148		0.057	*0.005*
				-20.15	-21.04	
Average Escs squared				6.48	6.66	
				0.002	0.002	
						26.00
Early track* AverageEscs						3.66
						0.000
						35.79
Early track* Escs standard deviation						5.68
						0.000
Groups of Control Variables[4]						
Individual characteristics	√	√	√	√	√	√
Family background	√	√	√	√	√	√
Peer composition		√	√	√	√	√
School location and size	√	√	√	√	√	√
School resources	√	√	√	√	√	√
Institutions	√	√	√	√	√	√
Country level controls	√	√	√	√	√	√
Number of observations	174,921	174,921	174,921	174,921	174,921	174,921
F	183.69	178.95	175.8	182.88	179.95	182.76
Prob.>F	0.000	0.000	0.000	0.000	0.000	0.000
R^2	0.307	0.340	0.340	0.341	0.341	0.343

[1] Mexico, Turkey and France are not included. [2] Regressions are run using students sample weights provided in PISA database. [3] For each variable, the first row refers to the estimated coefficient, the second to the robust standard error – adjusted for clustering at the school level - and the third to the P value. [4] Imputation dummies for missing data are included in all regressions. Source: elaborations on PISA 2006 data.

Table 5: Students performances OECD countries[1]: robustness checks. OLS regressions[2, 3]

	ST-5 Math	ST-5 Reading	ST-5a
Average Escs	44.06	53.89	40.38
	3.05	2.65	7.71
	0.000	*0.000*	*0.000*
Escs standard deviation	-19.78	-22.12	-17.47
	7.64	6.35	7.34
	0.010	*0.000*	*0.017*
Early track* Average Escs	32.07	24.62	9.75
	4.16	4.01	12.21
	0.000	*0.000*	*0.425*
Early track* Escs standard deviation	50.55	13.07	26.15
	6.00	6.33	12.55
	0.000	*0.039*	*0.037*
Groups of Control Variables[4]			
Individual characteristics	√	√	√
Family background	√	√	√
Peer composition	√	√	√
School location and size	√	√	√
School resources	√	√	√
Institutions	√	√	√
Country level controls	√	√	
Country fixed effect			√
Pseudo-School fixed effects[5]			√
School additional controls			√
Number of observations	174,921	170,713	153,483
F	203.74	197.48	175.58
Prob.>F	0.000	0.000	0.000
R^2	0.3913	0.3844	0.3633

[1] Mexico, Turkey and France are not included. Data on scores in reading are not provided for US.[2] Regressions are run using students sample weights provided in PISA database. [3] For each variable, the first row refers to the estimated coefficient, the second to the robust standard error – adjusted for clustering at the school level - and the third to the P value. [4] Imputation dummies for missing data are included in all regressions. [5] Pseudo-School fixed effects are identified, for each country, according to the quintile of escs average of every school. Source: elaborations on PISA 2006 data.

Table 6A: Students performances in science in OECD countries[1] by early and no early tracking countries[2]. OLS regressions[3, 4]

	ST-1		ST-2		ST-2 vocat		ST-2a		ST-3	
	No early track	Early track	No early track	Early track	No early track	Early track	No early track	Early track	No early track	Early track
Average Escs	45.80	61.83	45.05	61.56	43.97	61.32	41.97	48.83	64.16	78.41
	2.95	3.86	2.93	3.85	2.97	4.14	7.53	8.91	6.73	7.54
	0.000	*0.000*	*0.000*	*0.000*	*0.000*	*0.000*	*0.000*	*0.000*	*0.000*	*0.000*
Escs standard deviation			-17.65	13.73	-10.70	14.28	-19.00	12.61		
			7.57	8.57	7.99	9.81	7.27	9.58		
			0.020	*0.109*	*0.180*	*0.146*	*0.009*	*0.188*		
Vocational * Average Escs					27.23	2.16				
					7.81	6.20				
					0.000	*0.728*				
Vocational * Escs standard deviation					-73.96	-3.77				
					20.70	15.45				
					0.000	*0.807*				
Average Escs squared									-22.35	-18.60
									7.44	7.60
									0.003	*0.014*
Groups of Control Variables[4]										
Individual characteristics	√	√	√	√	√	√	√	√	√	√
Family background	√	√	√	√	√	√	√	√	√	√
Peer composition	√	√	√	√	√	√	√	√	√	√
School location and size	√	√	√	√	√	√	√	√	√	√
School resources	√	√	√	√	√	√	√	√	√	√
Institutions	√	√	√	√	√	√	√	√	√	√
Country level controls	√	√	√	√	√	√	√	√	√	√
Pseudo-School fixed effects[5]							√	√		
School additional controls							√	√		
Number of observations	132,104	42,817	132,104	42,817	132,104	42,817	117,808	35,675	132,104	42,817
F	120.88	147.40	120.55	149.84	120.02	145.77	79.56	154.57	125.46	152.23
Prob.>F	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
R^2	0.313	0.546	0.314	0.546	0.315	0.546	0.338	0.549	0.315	0.546

[1] Mexico, Turkey and France are not included. [2] Regressions are run using students sample weights provided in PISA database. [3] For each variable, the first row refers to the estimated coefficient, the second to the robust standard error – adjusted for clustering at the school level - and the third to the P value. [4] Imputation dummies for missing data are included in all regressions. [5] Pseudo-School fixed effects are identified, for each country, according to the quintile of escs average of every school. Source: elaborations on PISA 2006 data.

Table 6B: Students performances in science by early and no early tracking countries, excluding countries outliers in mean escs[1,2]. OLS regressions[3,4]

	ST-1		**ST-2**		**ST-2 vocat**		**ST-2a**		**ST-3**	
	No early track	Early track	No early track	Early track	No early track	Early track	No early track	Early track	No early track	Early track
Average Escs	54.19	61.83	53.09	61.56	52.26	61.32	49.32	48.83	62.45	78.41
	3.74	3.86	3.67	3.85	3.75	4.14	9.21	8.91	8.74	7.54
	0.000	*0.000*	*0.000*	*0.000*	*0.000*	*0.000*	*0.000*	*0.000*	*0.000*	*0.000*
Escs standard deviation			-15.86	13.73	-7.06	14.28	-19.90	12.61		
			8.84	8.57	9.53	9.81	8.42	9.58		
			0.073	0.109	0.459	0.146	0.018	0.188		
Vocational * Average Escs					24.46	2.16				
					8.74	6.20				
					0.005	*0.728*				
Vocational * Escs standard deviation					-77.78	-3.77				
					21.81	15.45				
					0.000	*0.807*				
Average Escs squared									-10.66	-18.60
									11.00	7.60
									0.332	*0.014*
Groups of Control Variables[4]										
Individual characteristics	√	√	√	√	√	√	√	√	√	√
Family background	√	√	√	√	√	√	√	√	√	√
Peer composition	√	√	√	√	√	√	√	√	√	√
School location and size	√	√	√	√	√	√	√	√	√	√
School resources	√	√	√	√	√	√	√	√	√	√
Institutions	√	√	√	√	√	√	√	√	√	√
Country level controls	√	√	√	√	√	√	√	√	√	√
Pseudo-School fixed effects[5]							√	√		
School additional controls							√	√		
Number of observations	105,003	42,817	105,003	42,817	105,003	42,817	93,423	35,675	105,003	42,817
F	87.61	147.40	87.39	149.84	87.59	145.77	66.73	154.57	89.14	152.23
Prob.>F	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
R^2	0.319	0.546	0.319	0.546	0.321	0.546	0.339	0.549	0.319	0.546

[1] Countries with a mean escs lower than -0.3 – Portugal, Spain and Poland – are excluded together with Mexico, Turkey and France. [2] Regressions are run using students sample weights provided in PISA database. [3] For each variable, the first row refers to the estimated coefficient, the second to the robust standard error – adjusted for clustering at the school level - and the third to the P value. [4] Imputation dummies for missing data are included in all regressions. [5] Pseudo-School fixed effects are identified, for each country, according to the quintile of escs average of every school. Source: elaborations on PISA 2006 data.

Table 7: School average performances in science in OECD countries[1]. OLS regressions[2, 3]

	SC-1	SC-2	SC-2a	SC-2: No early track	SC-2: Early track
	81.68	73.64	70.83	76.30	78.18
Average Escs	2.93	3.17	3.36	3.36	5.01
	0.000	*0.000*	*0.000*	*0.000*	*0.000*
	-13.36	-29.68	-36.48	-32.64	15.34
Escs standard deviation	7.49	6.50	6.97	6.89	8.89
	0.075	*0.000*	*0.000*	*0.000*	*0.084*
		27.11	21.13		
Early track * Average Escs		5.18	5.12		
		0.000	*0.000*		
		40.98	52.82		
Early track* Escs standard deviation		8.78	14.07		
		0.000	*0.000*		
Groups of Control Variables					
School location and size	√	√	√	√	√
School composition	√	√	√	√	√
School resources	√	√	√	√	√
Institutions	√	√	√	√	√
School additional controls	√	√	√	√	√
Country level controls	√	√		√	√
Country fixed effects			√		
Number of observations	5,098	5,098	5,098	3,846	1,252
F	65.95	69.24	68.49	42.84	78.42
Prob.>F	0.000	0.000	0.000	0.000	0.000
R^2	0.628	0.640	0.669	0.580	0.812

[1] Mexico, Turkey and France are not included. [2] Regressions are run using school sample weights provided in PISA database. [3] For each variable, the first row refers to the estimated coefficient, the second to the robust standard error and the third to the P value. Source: elaborations on PISA 2006 data.

Table 8: Students performances in science in OECD countries[1] by early and no early tracking countries. Estimated coefficients by model ST-2 of "Averageescs" and "Escs standard deviation". Quantile regressions[2]

	Average Escs		Escs standard deviation	
Percentile	No early track countries	Early track countries	No early track countries	Early track countries
10	46.38***	63.75***	-22.38***	16.18***
20	45.36***	61.40***	-22.60***	15.38***
30	45.89***	60.94***	-18.40***	8.39***
40	46.39***	58.28***	-23.56***	12.06***
50	45.29***	60.57***	-24.54***	11.75***
60	44.12***	63.20***	-21.35***	7.44***
70	45.01***	60.97***	-15.79***	12.52***
80	42.11***	59.29***	-16.43***	11.27***
90	41.08***	57.26***	-12.42***	22.05***

[1] Mexico, Turkey and France are not included. [2] Regressions are run using students sample weights provided in PISA database.Robust standard error – adjusted for clustering at the school level have been computed. Imputation dummies for missing data are included in all regressions. *** 99% significance level. Source: elaborations on PISA 2006 data.

Fig. 1: Quantile regressions on students performances in science in OECD countries. Estimated coefficients of "Average Escs". Source: elaborations on PISA 2006 data

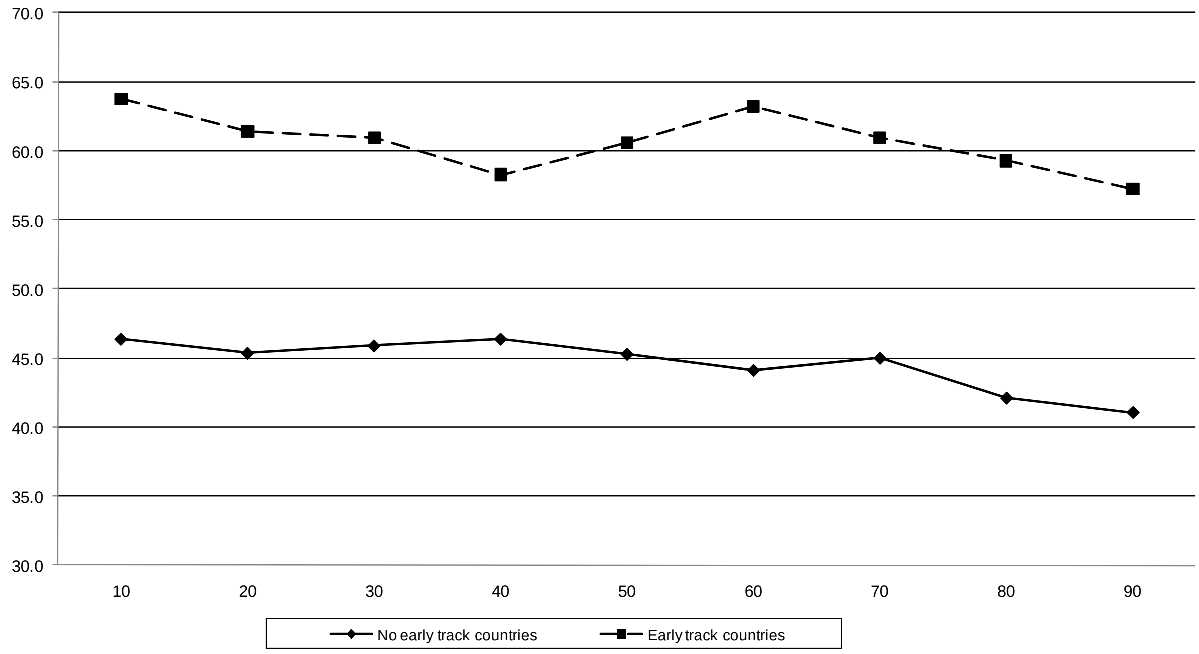

Fig. 2: Quantile regressions on students performances in science in OECD countries. Estimated coefficients of "Escs standard deviation". Source: elaborations on PISA 2006 data

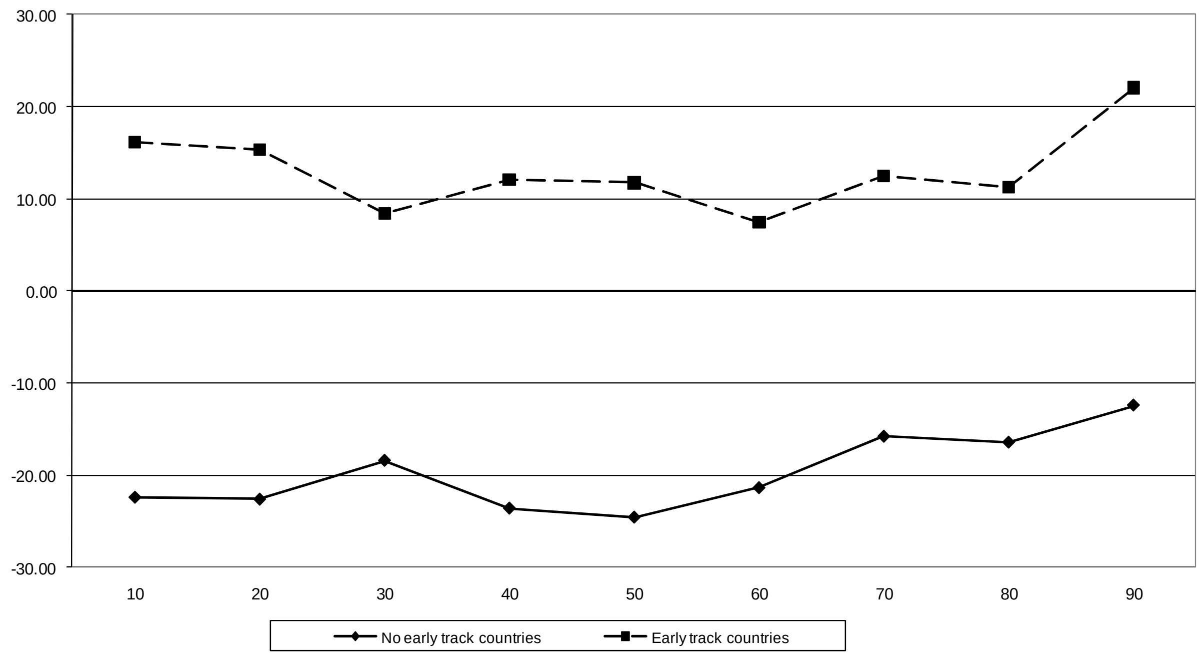

Table A1: Control variables used in regressions

School level controls	
Location and Class Sizes	
School Location	4 dummies: village, small town, town, city (large city is the omitted modality)
Class sizes	5 dummies: less than 15, 16-20, 26-30, 31-35, more than 35 (21-25 is the omitted modality)
Resources	
Ratcomp	Ratio of computers to school size
Compweb	Proportion of computers connected to web
Stratio	Student-Teacher ratio
Scmatedu	Quantitative index provided in PISA 2006 dataset about "Quality of educational resources"
Tcshort	Quantitative index provided in PISA 2006 dataset about "Teacher shortage" (on a negative scale)
Institutions	
Respres	Quantitative index provided in PISA 2006 dataset about "Responsibility for resource allocation index"
Respcurr	Quantitative index provided in PISA 2006 dataset about "Responsibility for curriculum & assessment"
School type	3 dummies: public, private dependent, private independent ("missing school type" is the omitted modality, due to the several missing values of the school type variable)
Residence	Dummy variable showing if residence is a prerequisite or a high priority for being admitted to the school
Student record	Dummy variable showing if previous academic records (or a specific test) are a prerequisite or a high priority for being admitted to the school
N.B. the dummies on the single components of school autonomy and responsibility about resource and curricula (the modalities of the scq11 Pisa questionnaire) have not been included because of the several missing values. Replacing these variables with the respres and respcurr indexes (with much less missing values) does not alter regression results.	
School Additional Controls	
Sorting by ability	Two dummies from the 3 modalities of the ability group variable showing, respectively, if students are grouped according to their abilities within schools for all subjects of for some subjects
School competition	Two dummies capturing the degree of school competition in the area.
Principal evaluation	Dummy variable showing if achievements are used in the evaluation of the principal's performance
Teacher evaluation	Dummy variable showing if achievements are used in the evaluation of teachers' performance
Allocation evaluation	Dummy variable showing if achievements are used in decisions about instructional resource allocation to the school
Over time evaluation	Dummy variable showing if achievements are tracked over time by an administrative authority

Tab. A1: Control variables used in regressions (suite)

School composition	
Average age	Average age of interviewed students
Share of females	Share of females among interviewed students
Share of immigrants	Share of immigrants among interviewed students
Share of "foreign languages"	Share of interviewed students which speak a foreign language at home.
Share of vocational students	Share of interviewed students enrolled in a vocational program.
Average Escs	Average level of the escs index of interviewed students
Escs standard deviation	Standard deviation (corrected for the country escs standard deviation) of the escs index of interviewed students
Average Escs squared	Square of the average level of the escs index of interviewed students
N.B. in regressions at student level these variables (apart from escs standard deviation) are considered net of the individual responses.	
Country level controls	
Gdp per capita	
Spending in education per capita	
Share of students enrolled to pre-primary school	
Age of first track	
Early track	Dummy variable: 1 if school track occurs before age 13, 0 otherwise
Duration of pre-primary schools	In years
External exam	Share of students subjected to an external evaluation in science
Standard test	Share of students subjected to standard evaluation tests in science
Student level controls	
Individual characteristics	
Age	
Sex	
Grade	Students below grade 8 and beyond grade 11 are excluded from the sample; hence, grade is captured by 3 dummies
Vocational	Dummy variable: 1 if the student is enrolled in a vocational program, 0 otherwise
Isced 3	Dummy variable: 1 if the student is enrolled in an upper secondary, 0 otherwise
Immigrant	Dummy variable: 1 if the student was not born in the country of test, 0 otherwise
Foreign Language	Dummy variable: 1 if the student speaks a foreign language at home, 0 otherwise

Table A1: Control variables used in regressions (suite)

Family background	
Hisei	Quantitative index provided in PISA 2006 dataset showing "the highest parental occupational status"
Pared	Quantitative index provided in PISA 2006 dataset showing (in years) "the highest parental educational level"
Homepos	Quantitative index provided in PISA 2006 dataset about "home possessions"
Books at home	Five dummies on number of books: 11-25, 26-100, 101-200, 201-500, more than 500 (less than 11 is the omitted modality)
Escs	Quantitative index provided in PISA 2006 dataset showing the "Family economic, social and cultural status"
Imputation dummies	
Intercept dummies	One dummy for each imputed variable (concerning escs, pared, hisei, homepos, immigrant, foreign language) showing if the value has been imputed
Slope dummies	One dummy for each imputed variable (concerning escs, pared, hisei, homepos, immigrant, foreign language) showing the interaction between the intercept imputation dummy and the value of the imputed variable

Source: information provided by PISA 2006, OECD and UNESCO datasets, Brunello and Checchi (2007), Woessmann et al. (2009)