Document de travail de l'OFCE

PEER HETEROGENEITY, PARENTAL BACKGROUND AND TRACKING: EVIDENCE FROM PISA 2006

OFCE

N° 2010-23

OCTOBER 2010

Michele Raitano
University of Rome
Francesco Vona
OFCE/DRIC

OFCE - Centre de recherche en économie de Sciences Po
69, quai d'Orsay - 75340 Paris Cedex 07 - Tél/ 01 44 18 54 00 - Fax/ 01 45 56 06 15
www.ofce.sciences-po.fr

Peer Heterogeneity, Parental Background and Tracking: Evidence from PISA 2006

Michele Raitano[*] and Francesco Vona[§],

Abstract

The empirical literature using large international students' assessments tends to neglect the role of school composition variables in order not to incur in a misidentification of peer effects. However, this leads to an error of higher logical type since the learning environment crucially depends on peers' family background and on peer heterogeneity. In this paper, using PISA 2006, we show how peer heterogeneity is a key determinant of student attainment and of opportunity equalization. Interestingly, the effect of school compositional variables differs depending on the country tracking policy: peer heterogeneity reduces efficiency in comprehensive systems whereas it has a non-linear impact in early-tracking ones. In turn, linear peer effects are larger in early-tracking systems. Besides, higher heterogeneity tends to equalize student differences related to family background. Results do not change in school- and student-level regressions suggesting that the impact of heterogeneity is correctly identified. Results are also robust when we add school-level dummies and several controls correlated with the school choice to alleviate the selectivity bias of linear peer effects.

JEL: I21, I28, J24

Key words: peer heterogeneity, peer effects, schooling tracking, educational production function, equality of opportunities.

[*] University of Rome La Sapienza, Department of Public Economics and CRISS. michele.raitano@uniroma1.it (corresponding author).

[§] OFCE, Sophia-Antipolis, (France), and University of Rome La Sapienza, Department of Public Economics (Italy).

1. Introduction

The quality of the educational system is commonly recognized to have a remarkable impact on growth and on the equalization of student outcomes. Large international assessment programs constitute a valid tool for analysing how differences in educational policies translate into different student outcomes, circumventing problems of skill comparability. Existing studies using these surveys attempt to reconcile the observed lack of correlation between resources invested and educational outcomes accounting for the institutional features of the educational process, such as the ones associated to the degree of autonomy of the school and of the accountability (e.g. Woessmann et al. 2010). Much less attention has been devoted to analyse the effect of the school (or class[1]) composition by background and/or abilities —i.e. the so-called "peer effect".

By fostering or hindering skill formation, peer effects and social interactions have provided to be a fundamental source of both efficiency—if peer effects are non-linear (Benabou 1996)—and intergenerational inequality (Durlauf 2004). Moreover, especially up to the secondary level of education, the influence of the social background at school has been shown to be critical for the development of cognitive skills and therefore of labour market success (Carneiro and Heckman 2003; Fernandez and Rogerson 1996; Hanushek and Woessmannn 2010). In spite of the well-recognized importance of these effects, the limitation to the cross-sectional dimension and the lack of initial information on student ability makes it difficult to identify linear "peer effects" using these surveys. In particular, the fact that the assignment of students to classes and schools of different quality is endogenous severely distorts the estimation of peer effects[2].

However, neglecting to account for the characteristics of the social interactions and, more in general, of the "external environment" at school can also raise serious biases in the estimates; at least as large as the ones that would emerge from not considering endogeneity issues in the

[1] Note that peer effects at the class and the school level capture two different ways in which social interactions affect student outcomes: whereas the former is more correlated with direct effects on the learning environment, the second includes a broader set of interactions.

[2] For instance, since the school composition is endogenous both to educational policies, such as the admission procedures and the age of tracking, and to the characteristics of the neighbourhood of residence, the choice of the school would be strongly influenced by unobservable individual, parental and urban characteristics. An important caveat is required here. If the selection problem is not perfectly solved and hence identification of the peer is not transparent, the estimated coefficient on peer variable turns out measures both the school and the community peer effects (Toma and Zimmer 2000). The distinction of the two effects is crucial for targeting policies at the national level, whereas in international comparisons the exact estimates of the production function externalities generated at school, net of the community externality due the social interaction out-of-school, is far less important as urban and schooling policies are intrinsically indistinguishable. Another well-known problem is the one of reflexivity, namely each student outcome is affected by the average mean of the other student and at the same time affect the outcome of all other students, (Maskin 1993). Due to the reflexivity problem, the data requirements for unbiased estimates of peer are almost impossible to meet. For policy purpose what is relevant is to quantify the peer effect, not to identify the source of it. Hence, many studies ignore the reflexivity problem.

estimates of peer effects. So far, except for few recent cases[3], the mainstream empirical strategy in studies using large international assessment programs attempted to minimize the first-type of bias associated to an improper identification of peer effects. This, however, brings to a second-type of error, *of higher logical level*, associated to a misspecification of the true educational production function.

The "error" made by not explicitly including "class- or school-compositional variables" is of great policy concern when the characteristics of the school environment are strongly linked with other policies that are found to affect the student choice. Of particular interest is the interaction between school compositional variables and tracking policies as long as the latter influence the school choice of students from different backgrounds (e.g. Dustmann 2004; Checchi and Flabbi 2005). In turn, even in comprehensive systems –i.e. Anglo-Saxon and Scandinavian ones where school tracking is absent – there might be a strong tendency to self-select students by ability and background through several other factors such as residential segregation, admission procedures, use of private sector and within-school ability tracking (Waldinger 2006). Unfortunately, large international student assessments do not contain information of student and school residential locations, hence good instruments for the school composition are not available especially for this type of schooling system. Information on admission procedures and within-school sorting by ability—included in the PISA dataset used in this work—allow to partially attenuate the estimation bias of the "school-composition effect".

Using PISA-2006 dataset, the aim of this paper is to fill the existing gap in the literature on international comparison of educational systems by analysing how heterogeneity in the school environment, proxied by the standard deviation of the student backgrounds, affects both school efficiency and equity. In particular, starting from a widely accepted specification of the schooling production function (see Hanushek 1986, 2003; Fuchs and Woessmann 2007), the paper seeks to investigate whether the impact of school-composition variables changes in system with different tracking policies. To partially address the selectivity bias, as a first step we carry on regressions at the school-level in order to reduce unobservable student variability. Secondly, we move to student-level regressions (i.e. controlling for individual and parental characteristics), where we include a school dummy for each quantile of the country distribution of the average parental background in order to attenuate the selectivity bias in estimating peer effects and the impact of peer heterogeneity. What we argue is that the comparison of individual- and school-level estimates represents a reliable

[3]Vanderberghe (2002), Rangvid (2007), Schneeweis and Winter-Ebmer (2007), Entrof and Lauk (2006), Ammermueller and Pischke (2006). For the PIRLS survey, a reliable identification strategy of the peer effects is available as long as within-school variation can be exploited due to the detailed class-level information of the peer variables (see Ammermueller and Pischke 2006).

way to identify the effect of peer heterogeneity on outcomes. Finally, the impact of peer heterogeneity and tracking policies on the equality of opportunity is assessed taking into account confounding factors such as the duration of pre-primary education and the share of private schools (see Schuetz et al. 2008).

Our empirical analysis strongly confirms that the impact of school-composition variables is strong, very significant and the single most important determinant of student performance. Unlike previous studies[4], often focussing on a single country, heterogeneity has a significant impact on student outcomes, but the patterns followed by countries with different age of tracking widely differ. On the one hand, peer heterogeneity by background reduces attainments in comprehensive schooling systems even if this result is largely driven by pupils attending vocationally-oriented programs. On the other hand, in early tracking systems, there exists an optimal degree of heterogeneity that maximizes attainments. Consistently with the theoretical literature (e.g. Brunello et al. 2007), linear peer effects are found to be stronger under the early-tracking regime. Besides, as expected, higher peer heterogeneity reduces the socio-economic gradient both in early and late tracking systems.

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 some preliminary evidence supporting our way of measuring background. In section 4 (resp. 5), we present the results of school- (resp. student-) level regressions. Section 6 analyses the effect of heterogeneity on the socio-economic gradient, whereas section 7 concludes.

2. Related literature and Empirical Strategy

The observed weak correlation between the educational inputs and student outcomes represented the main puzzle for the literature attempting to explain the determinants of educational quality (Hanushek 2003). A simple principal-agent approach to educational production emphasizes the importance of the institutional design of the educational sector to explain this puzzle (Bishop and Woessmann 2004). In this framework, institutions enhancing school competition, autonomy and accountability are expected to increase the pressure towards higher standards, to enable the full exploitation of local knowledge regarding students' characteristics and to reduce the risk of opportunistic behaviour that would emerge in absence of appropriate monitoring practices (Woessmannn et al. 2010). Recent empirical studies using international assessment programs supported this view of educational production and highlighted possible complementarities among

[4] See Hanushek et al. (2003), Zimmer and Toma (2000), Rangvid (2007), Schneeweis and Winter-Ebmer (2007).

different institutions; in particular, between the ones accountability practices and the degree of school autonomy (Woessmann 2003; Fuchs and Woessmann 2007; Woessmannn et al. 2010).

Whereas at the empirical level these institutions seem to explain part of the missing correlation between resources and educational quality, background variables still represent the ones with the larger explanatory power in all works using international surveys (e.g. Fuchs and Woessmann 2007). Hence, the puzzle of the missing resource-quality link can be explained from a theoretical perspective that explicitly includes school-composition variables as inputs of educational production (De Bartolome 1992, Benabou 1996, Fernandez and Rogerson 1996). The learning environment can not, in fact, be reduced to a vector of school characteristics as long as the abilities and the home background of school- and class-mates determine the learning standards for the class as a whole together with the out-of-school social context.

From a theoretical standpoint, the influence of class heterogeneity on student outcomes is ambiguous as forces going in opposite directions tend to offset each other. 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 (Laezar 2001), or positive knowledge spillovers from good students to average and/or bad ones (e.g. Durlauf 2004). Which one tends to prevail depends on the shape of the educational production function. More precisely, if school-composition effects 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 matching process of students to schools might depend in a substantial way from factors beyond the sphere of educational policies. For instance, the “rich” can successfully isolate themselves by approving residential restrictions to school admission. More in general, the assignment of students to schools that maximizes aggregate human capital is very unlikely to emerge as a market outcome because several structural constraints shape schooling choices: admission procedures, physical distance, early tracking policies, within-school ability tracking, etc. (e.g. de Bartolome 1992).

Recent theoretical and empirical contributions underlie the role of early tracking policies in affecting schooling decisions of individuals from different backgrounds (Epple and Romano 2002; Dustnmann 2004; Brunello et al. 2007). Because parental influence matters more at the beginning of the student life, an earlier streaming increases the probability that students from worse backgrounds end up in the vocational streaming, which offers less promising learning perspectives in terms of teacher quality, resources invested and course content. On the other hand, students from disadvantaged socio-economic backgrounds that decide to be enrolled in a gymnasium are likely to be more motivated and/or particularly able. Finally, educational systems with an early tracking age often puts vocational and specific training at the centre of their development strategy (Hall and Soskice 2001; Krueger and Kumar 2004), hence vocational schools might attract also students with background above the average. With these premises in mind, one would expect that heterogeneity in unobservable student characteristics within the school ends up being substantially lower in early tracking systems with respect to comprehensive ones. Thus, the effect of peers' heterogeneity on student outcomes should vary in systems with different tracking policies (Brunello et al. 2007) and empirical assessments are required to quantify this difference. To the best of our knowledge, existing empirical works do not assess whether the impact of peer heterogeneity on student attainments varies in schooling system with different tracking policies. This represents a main contribution of the present work.

Not only student outcomes, but also the distribution of educational opportunities depends on peer heterogeneity and early tracking policies. Theoretical works predict that highly segregated schools and an earlier streaming age both widen educational opportunities (e.g. Brunello et al. 2007). The latter effect is well documented in the empirical literature (e.g. Ammermueller 2005; Hanushek and Woessmann 2006; Schuetz et el. 2008), although more recent works using difference-in-difference estimation at the individual level, or more reliable measures of tracking, seem to discard the hypothesis that an earlier tracking age increases the inequality of opportunities (Waldinger 2006; Brunello and Checchi 2007). The former effect is less analysed using a direct measure of peer heterogeneity as we do here. However, a positive and significant peer effect mechanically leads to levelling opportunities. Still, it is not clear whether the levelling of educational outcomes is stronger for low ability students, as it appeared in earlier works (Zimmer and Toma 2000; Vandenberghe 2002; Hanushek et al. 2003; Rangvid 2007; Schneeweiss and Winter-Ebmer 2007), or from high to average ability students, as more recent researches for the UK have demonstrated (Gibbons and Telhaj 2008; Lavy et al. 2009).

This paper is related to the literature on peer effects using large international assessment surveys. The more rigorous attempt to identify peer effects in this literature is the paper Ammermueller and Pischke (2006). Similarly to Hoxby (2000), Hanushek et al. (2003), McEvan (2003), they use within-school variations in the class composition to solve the identification problem associated to a non-random student assignment. Under the assumption that the within-school allocation of student and resources is random, all the distortions due to non-random assignment are associated to variations between-schools. Hence, class variations in the peer composition within the school enable to disentangle the pure peer effect from an endogenous school selectivity bias. Using the PIRLS dataset, they found modestly large peer effects even when controlling for measurement errors. Moreover, accounting for the selection bias only slightly reduces the peer effect obtained in a standard OLS specification.

Unfortunately, a similar identification strategy is not available in the PISA survey that does not provide detailed classroom information. Conversely, for the scope of this paper, the main advantage of the PISA dataset is that it allows to uncover cross-country variations in tracking policies and offers information on several policies affecting the school composition, i.e. admission procedures, ability grouping.

In the few papers attempting to assess peer effects using PISA, the identification strategy has been based upon the claim that the omitted variable bias is the most important source of selectivity problems. Therefore, the selectivity problem is reduced by having a large set of controls, both at school- and at individual-level, which are likely to affect the assignment of students to schools (Rangvid 2007; Schneeweiss and Winter-Ebmer 2007). For instance, Rangvid (2007) measures peer quality with the average education of the mother and uses variables of parental care, encouragement and time spend with their children to reduce the omitted variable bias. By using quantile regressions techniques, she conditions the effect of peers to the ability distribution and found stronger peer effects for low ability students in Denmark. In turn, a higher heterogeneity has an insignificant impact on student achievement along the entire distribution of test scores. Schneeweiss and Winter-Ebmer (2007) found a similar differential impact of peer quality along the ability distribution for Austria, whereas background heterogeneity appears to have a slightly negative and significant effect on outcomes. In a paper with a logic similar to our but with a different focus, Entrof and Lauer (2006) attempts to distinguish the peer effect of immigrants and natives in different tracking systems and found stronger peer effects in countries with an earlier tracking. Another possibility, followed by Fertig (2003), lies in instrumenting peer heterogeneity with proxies of the caring behaviour of parents at home and of admission procedures. For the U.S., he found a strongly negative effect of heterogeneity, measured with the coefficient of variation in the achievement of schoolmates, on performance in reading. However, both the measure of schooling heterogeneity—the coefficient of variation—and the instruments chosen appear rather weak. In particular, the coefficient of variation is such that a lower mean in the test leads to an increase in the coefficient of variation. As a result, since a lower mean of the schoolmates negatively affects the final outcome, this impact appears the mechanical consequence of a higher heterogeneity rather than the one of a lower mean[5]. An opposite, negative and significant effect of class heterogeneity—measured with

[5]Moreover, the instruments used become not anymore valid if one includes other contextual variables such as the share of parents working in the peer group. More in general, it seems difficult to find convincing instruments in cross-sectional regressions without having information on the characteristics of the neighbourhood of residence.

the standard deviation of a composite index of family background—on student attainments is found by Vandenberghe (2002) in a cross-country study using the TIMMS dataset. However, he introduces several non-linear terms in the class composition variables that makes the effect of heterogeneity difficult to isolate.

Using the international student assessment PISA 2006, the purpose of this paper is to estimate the impact of school-composition variables, in particular of peer heterogeneity, on efficiency and equity, and how this impact varies in systems with different tracking policies; hence we are implicitly testing the validity of a linear and of a "pooled" specification of the peer influence. In order to minimize biases in the estimates of school-composition variables, the core of our empirical strategy is to compare school- and student-level regressions adding several controls that are correlated with the sorting of individuals to schools of different quality, i.e. admission procedures, dummies for school competition, ability grouping. What we claim is that the impact of heterogeneity should be correctly identified if once moving from school to individual estimates, the sign, the size and the significance of the coefficient of peer heterogeneity remain substantially unchanged. The further advantage of school-level regressions is to attenuate the bias associated to unobservable individual characteristics; at least under the plausible assumption that the mean of these unobservable characteristics – i.e. the average individual selected by the school – are fully captured by compositional variables (e.g. share of immigrants) and certain schooling characteristics (e.g. school type or admission procedure). This advantage should be balanced against the cost that in school-level regressions linear peer effects are indistinguishable from the average background effect.

In student-level regressions, we perform several robust checks with the purpose of improving the reliability of the estimated impact of heterogeneity and, at the same time, to adopt updated empirical strategies to reduce the bias in the estimation of linear peer effects. In particular, as in Ammermueller and Pischke (2006), we use school level dummies in order to attenuate the selectivity bias. Recall that this strategy is valid under the assumption that the correlation between school and individual unobservable characteristics is mainly dependent on school characteristics. According to Ammermueller and Pischke (2006), this necessary condition for identification is less likely to be satisfied in secondary schools where within-school sorting by ability matters, especially in certain countries. Here, controlling for within-school ability tracking enables us to mitigate this "confounding effect" (while, due to data limitations, we are not able to identify classroom peer effects)[6].

[6] PISA surveys do not allow to estimate classroom peer effects.

To be more precise, consider the following specification of the schooling production function that is basically the one proposed by Fuchs and Woessmann (2007) extended to include school composition variables:

$$A_{isc} = \alpha + \beta X_i + \chi X_s + \delta X_c + \gamma BACK_i + \upsilon \overline{X}_s + \mu \overline{BACK}_s + \sigma Var_s(BACK_i) + u_s + u_c + u_i + u_{is} + u_{isc} \quad \text{(eq.1)},$$

where, for sake of space, we do not write down imputation dummies for missing variables (see section 5). The student achievement A in school s and country c is the resultant of a vector of individual (X_i), school (X_s), country (X_c) controls plus individual background ($BACK_i$) and school compositional factors in terms of students' background ($\overline{BACK}_s$ and $Var_s(BACK_i)$) and other characteristics ($\overline{X}_s$), such as gender, immigrant status, etc. The error is decomposed here in a country effect u_c, an individual effect u_i, a school effect u_s and a correlated school-individual effect u_{is} plus the standard independent error term u_{isc}. The school-individual interaction and the individual effect are the ones that are likely to be correlated with both the school composition variables and to the student achievement. This is because individuals are selected by schools upon certain unobservable variables and procedures. By schools, eq.1 becomes:

$$\overline{A}_{sc} = \alpha + (\upsilon + \beta)\overline{X}_s + \chi X_s + \delta X_c + (\gamma + \mu)\overline{BACK}_s + \sigma Var_s(BACK) + u_c + \overline{u}_s + u_{isc} \quad \text{(eq.2)}.$$

Under the plausible assumption that the average unobservable individual characteristics boils down onto the school-composition variables and school characteristics, the impact of heterogeneity is correctly identified using “school-level clustering-robust” linear regressions. At the individual level, instead, the way of reducing the selectivity problem rests on the assumption that individual and correlated school-individual effects are fully captured by observable and unobservable schooling characteristics. Unobservable schooling characteristics are proxied including both other socio-demographic school-composition variables (e.g. the share of immigrants and of females), and a dummy equal 1 for the quantile of the country-specific distribution of the average parental background at which the school belongs to. Hence we estimate the following function:

$$A_{isc} = \alpha_{sc} + \beta X_i + \chi X_s + \delta X_c + \gamma BACK_i + \upsilon \overline{X}_s + \mu \overline{BACK}_s + \sigma Var_s(BACK_i) + u_c + u_s + u_{isc} \quad \text{(eq.1')},$$

where $\overline{X}_s$ and $\overline{BACK}_s$ are now the schoolmates composition net of individual and α_{sc} is the school quantile fixed effect. Equation 1' leads to unbiased estimates of both the linear and the heterogeneous peer effect if the correlated and the idiosyncratic individual term are fully absorbed in the new covariates. Throughout the paper, the fact that equation 1' is often estimated separated by the type of tracking policy (see below) and always using standard errors clustered by school further reduces the endogeneity bias especially in countries that track students earlier—where the within-school variation in the unobservable individual characteristics is expected to be lower.

Concerning the estimation of the effect of heterogeneity and tracking on equity, we follow the specification of Schuetz et al. (2008) and Brunello and Checchi (2007) where a full set of interaction dummies between a measure of background and several factors that might affect the size of the socio-economic gradient are included. Among these factors, we add heterogeneity in background in a reduced-form model where school characteristics are excluded since, differently from estimation of the standard production function, the impact of student background on performance should be depurated by any effect that might act through families' differential access to schools of different quality (Schuetz et al. 2008). In section 6 we estimate the following relationship:

$$\begin{aligned} A_{isc} &= \alpha + \beta X_i + \gamma BACK_i + \eta(X_c * BACK_i) + \ldots \\ &+ \vartheta(T_c * BACK_i) + \varphi(BACK_i * Var_s(BACK_i)) + \lambda(BACK_i * \overline{BACK}_s)) + u_{isc} \qquad \text{(eq.3).} \end{aligned}$$

The first interactions are between the factors – i.e. duration of pre-primary school, share of public schools, student/teacher ratio – that might disturb the relationship between background and or variables of interest—i.e. tracking and school compositional variables—which are captured by the other interaction terms. Next section briefly describes the PISA dataset and provides a descriptive glance of the different impact of heterogeneity in countries with different tracking policies.

3. Descriptive Statistics and Preliminary Analysis.

In the empirical analysis, we use the 2006 PISA survey that so far has not been used yet to assess the impact of school-composition variables on student outcomes. PISA's target population is 15-year-old students in each country, regardless of the grade they currently attend. Differently from other internationally comparable surveys such as PIRLS and TIMMS programs, the PISA dataset presents the additional desirable feature of being more oriented on problem solving capacities (know-how) rather than on curricula skill (know-what). PISA dataset contains detailed information on student's home background, school resources and a wide range of institutional variables capturing the degree of school autonomy, accountability practices and variables affecting the student choice (Woessmann et al. 2010). Individual controls such as sex, age, grade, etc. are also available, whereas policy variables at the national level are usually integrated by other dataset (Oecd, UNESCO, etc.). Since PISA 2006 is focussed on science, we consider only the outcome in science as the dependent variable.

Some variables used in the econometric analysis of next sections are indexes built by PISA experts in order to summarize various school or individual characteristics related to each other. For instance, the degree of autonomy in managing resources at the school level is captured either by a vector of dummies (autonomy in within-school allocation, in hiring and firing teacher, etc.) that are highly dependent to each other or by a synthetic indicator built upon these dummies (see table A1 for details). The same holds for indexes of school resources and background[7]. Of particular interest for our work is the variable of background built by the Oecd—called the Economic Social Cultural Status “escs”—which combines information provided by widely used measures of parental background: highest parental years of education, the highest occupational level quantified with the index of occupational status (Ganzeboom et al. 1992), the number of books at home and the resources available at home to study, i.e. “homepos”. The “escs” variable is chosen here as our baseline measure of background since it encompasses in a synthetic way the various and multidimensional aspects shaping the impact of family characteristics on the student’s attainment.

Table 1 displays by countries the descriptive statistics of the main variables on which we focus on to explain student performance in science: age and grade of first tracking, mean results in science and mean and standard deviation of family background indexes, etc. (for a full description of the variables used see table A1 in the appendix). Table 2 shows the degree of correlation between the composite “escs” index and each of its components, which follows by the construction. However, a much lower correlation with the variable books-at-home suggests to include such variable together with either “escs” index (or its components) as controls of individual background in the empirical specification.

A key issue for our paper is to find a reliable measure of peer effects and of heterogeneity in backgrounds at the school level. As standard in the literature, the average level of the “escs”, net of the individual one, is our favourite measure of the linear peer effect, whereas – similarly to Rangvid (2007) – the standard deviation of the “escs” account for peer heterogeneity. Moreover, in order to partially account for cross-country differences in the allocation process of students of different background to schools of different quality, we normalize the standard deviation at the school level

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

with the one at the country level[8]. In fact, a high heterogeneity at the school level can be due to a high heterogeneity in the country rather than to a random sorting of students to schools, hence the desired level of heterogeneity at the school level is bounded by the overall background heterogeneity at the country level.

A first look to the data is useful to ground on more solid bases the following empirical analysis. Simple scatter plots adjusted for school weights highlight a pronounced non-linearity in the relationship between peer heterogeneity and performance resulting as the balance of the positive and negative externalities triggered by the interaction of individuals from different backgrounds (figure 1). However, as it appears clear from figure 2 and table 3, this relationship is largely driven by countries tracking students earlier. In turn, in comprehensive systems, higher heterogeneity negatively influences outcomes along the entire school distribution (figure 2). Since background variables are highly correlated with test outcomes, figure 3 suggests that this pattern is somehow driven by the one between the mean and the variance of the "escs" at the school level. A further comparison of the two school systems in table 4 (e.g. rows 4-8) shows that differences in the school composition by background are not as large as one would expect. What substantially differs between the two groups of countries is the quota of persons doing vocational programs (significantly higher in the early tracking system), the sorting within school by ability (significantly higher in comprehensive systems) and the admission procedures (relatively more based on student records and residence in comprehensive systems). Besides, the two systems seem to have a different degree of school dispersion in terms of unobservable individual characteristics: between-school variance in the average attainments is much larger in the early tracking with respect to the comprehensive systems (table 4, lines 1-2). This evidence seems to suggest that differences in tracking age mainly translate into differences in sorting by unobservable.

By and large, the descriptive analysis presented here confirms that different patterns emerge between countries with or without an early student tracking. This evidence motivates the inclusion of a non-linear specification of the effect of heterogeneity on outcomes. In the pooled specification, this is obtained with an interaction dummy between heterogeneity and early tracking. In the separated regressions, the square of the standard deviation is also included to account for non-linear effects of heterogeneity. Next sections present the results.

4. School-level regressions

[8] 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.

The sample of countries used in this paper consists in OECD ones but France where school variables have not been recorded in the PISA 2006 survey. As in Woessmann et al. (2010), Mexico and Turkey are excluded because they have an average "escs" that is a full standard deviation below the OECD average. Also following many studies using PISA surveys, we excluded from the sample those very few students enrolled in grades lower than 8 or higher than 11. Finally, as we intend to analyze the effect of social interactions at school, we restricted the sample to students attending schools for which PISA 2006 provided data for at least 15 students, i.e. we dropped schools with less than 15 interviewed students[9]. Our final sample includes 202.817 students clustered in 6.728 schools.

As discussed in section 2, the first stage of our analysis focuses on regressions at the schooling level in order to reduce unobservable student variability. Several control variables identified at school level are included. The first type of controls are compositional variables that proxy certain basic features of the demographic, social and cultural environment at school: the mean students' age and the share of females, of immigrants, of students speaking a foreign language at home and of students enrolled in a vocational programme (see table A1). In turn, as stated before, the average student index "escs" and its standard deviation at the school-level (normalized by the country standard deviation) are our measures of linear peer and of school-mates heterogeneity respectively.

As further control variables we added a set of variables concerning school resources and institutions, class size, school location and country-level controls (see table A1). Among country controls, we included institutional variables that are provided to be important determinants of student attainments (Fuchs and Woessmann 2007); in particular, the share of students subjected to external evaluation and/or standard test in science, the age of tracking between different kinds of programmes (general or vocational, OECD educational dataset) and the quota of pupils attending pre-primary education (UNESCO educational dataset).

Among school-level institutional variables, we included quantitative PISA indexes concerning school responsibility for allocating resources and for curriculum and assessment[10], school-type dummies (built crossing information on school management and the main source of founding) and two dummies for the admission procedures followed by the school (i.e. signalling if residence or students' ability are a high priority or a prerequisite for being enrolled in that school). The latter

[9] Note that the same sample restriction for similar purposes has been applied in Rangvid (2007).

[10] Among controls about school institutions, in all regressions shown in this paper, we included 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.

dummies seem particularly well suited in order to reduce the selectivity bias due to a non-random assignment of students to schools.

In table 5, we show the results of school-level regressions on science performance for the pooled sample of countries. For sake of space, in what follows we present estimated coefficients for the variables of interest, i.e. average and standard dev. of “escs”. Results available upon request show that, in both regressions at school- and at student-level (see section 5), other variables display the expected signs and significance consistently with the empirical literature on students’ performances using international assessment programmes (see Fuchs and Woessmann 2007; Hanushek and Woessmann 2010), in particular school resources seem to exert a lower influence than school and country institutional features.

Model SC-1 in table 5 highlights the large positive effect played by the average parental background; in fact, a change in one standard deviation of the “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 parental background effect, which has been found to be the larger explanatory factor of student outcomes (e.g. Hanushek and Woessmann 2010). Unlike linear peer effects, the impact of heterogeneity is correctly identified in school-level regressions under the plausible assumption that the average unobservable individual characteristics boils down onto the school composition variables and school characteristics. Background heterogeneity exerts a negative and significant impact on the average performance, even if the magnitude of this impact is rather small: a one standard deviation increase in the degree of heterogeneity leads to a 1.8 point decrease in the average science score (see model SC-2, table 5). This result in favour of segregation appears nuanced when we allow for non-linear effects of heterogeneity. The inclusion of the “escs” variance, so as suggested by the preliminary analysis in section 3, makes the relationship between heterogeneity and performance inversely U-shaped, being now positive and significant the linear term while the coefficient of the quadratic term is negative and significant (see model SC-3 table 5).

In models so far discussed we included, as controls of the link between school composition and performances, variables recording resources and institutional aspects at the school-level. However, educational inputs is likely to be related to student background; hence estimates of background variables can be plagued by endogeneity since pupils from better families attain schools with more resources and better institutions. Since this source of endogeneity stems from a more or less distributed allocation of resources and institutions within the country, aggregating school-level resources and institutional characteristics at the country-level allows circumventing these endogeneity problems, then providing unbiased estimates (see Woessmann 2003). Accordingly, the robustness of model SC-3 can be checked replacing school resources and institutional variables with their country average[11], which largely confirms previous results (see SC-4, table 5). Interestingly, with respect to model SC-3 the estimated joint impact of background and peer increases by only 1.9 points of a full standard deviation in the PISA score suggesting that the distortion induced by this source of endogeneity is negligible.

Consistently with the focus of the paper and with the preliminary analysis of section 3, the next step is to consider the joint influence of tracking and peer heterogeneity on achievements. Recall that tracking can occur within-school or between different types of schools. The former is based on ability grouping and prevails in Anglo-Saxon countries; the latter implies the streaming into completely different segments of the education process, generally offering general or vocational programmes such as in Germany and in many central European countries (Brunello and Checchi, 2007). Here we mainly focus on schooling tracking[12] to split countries according to the age when students have to choose between programs[13].

A first way to differentiate the effect of heterogeneity by tracking systems is to introduce an interaction term between peer heterogeneity and a dummy classifying OECD countries into early or late tracking ones. In this case, the negative impact and the significance of the "escs" standard deviation increases, but at the same time the fact that the interaction term between the "escs" standard deviation and the early tracking dummy is positive and significant highlights the differential impact of heterogeneity on student outcomes in countries tracking students before the age of 13 (model SC-5, table 5).

As a next step in order to better assess differences between the early-tracking and the comprehensive system, we run school-level regressions separated by the two groups of countries (table 6). Replicating model SC-2 for each group of countries it is possible to better disentangle the large difference in the impact of peer heterogeneity between early tracking and comprehensive systems. On the one hand, in countries tracking students after the age of 13, heterogeneity exerts a negative influence on student outcomes. On the other hand, the sign reverts in early tracking countries, but the positive effect is significant only at the cut-off level of 85%. Note that the opposite influence of peer heterogeneity in different tracking systems is confirmed even when we

[11] Woessmann et al. (2010) show the robustness of considering country averages based on PISA instead of data provided by other data sources. In model SC-4 also the share of students enrolled in vocational courses is considered as a country average.

[12] However, to account for within-school tracking, in model SC-7 in table 6, we will also control for information on ability grouping within the school.

[13] Literature provides several measures of tracking systems: 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. In model SC-5, in line with the Hanushek and Woessmann (2006), we consider as early-trackers countries where students have to choose before they are 13 years old.

split countries following the method proposed by Waldinger (2006), as shown in table 6 by model SC-2A. Finally, the size of the impact of heterogeneity on student outcomes increases when separated regressions are carried on with the impact of a one standard deviation increase ranging from +2.0 (resp. -4.1) to +2.4 (resp. -4.3) change in students' attainments in the early tracking (resp. comprehensive) system.

When including also non-linear effect of peer heterogeneity, differences between the two groups widen. In comprehensive systems both the linear and the quadratic term become not significant, whereas in early-tracking ones both terms appear highly significant, showing an inverted U-shaped relationship between background heterogeneity and the average performance (model SC-3, table 6). Moreover, this relationship remains robust either to the inclusion of country fixed effects (model SC-6, table 6) or – although at a much lower significance level – when country averages instead of school-level resources and institutional variables are considered (model SC-4, table 6)[14]. It is worth noticing that, calibrating with the coefficients estimated in table 2, the optimal degree of heterogeneity that maximizes attainments in early-tracking systems is located near to the median level of the "escs" standard deviation.

Finally, in order to account for across countries differences in ability tracking within the school (a widely used policy particularly in Anglo-Saxon countries), we run another model (SC-7, table 6) with additional school-level dummies capturing the procedures followed within the school for grouping students by ability and also, following Woessmann et al. (2010), accountability practices internal to the school (see table A1). The inclusion of these additional controls, which in principle should distort the impact of heterogeneity, reinforces our results in so far as the positive influence of peer heterogeneity in early tracking systems becomes stronger and more significant.

As stated in previous sections, the main advantage of school-level regressions presented so far is that they allow to attenuate the bias associated to unobservable individual characteristics. This advantage should be balanced against the cost that in school-level regressions linear peer effects are indistinguishable from the background effect. With the aim of identifying also this effect, we now move to student-level estimations.

5. Student-level regressions

Pooled student-level regressions lead to a substantial increase in the number of observations and hence allow controlling for several additional factors. First of all, when running regressions using

[14] These differences between early and late tracking countries emerge also when proxies of peer heterogeneity based on different parental background variables are computed (e.g. highest parental occupational status and educational attainment). Detailed results are available upon request by the authors.

students as the unity of observations individual characteristics (age, sex, grade etc…) are included (see table A1). Second, the multifaceted and complex mechanisms that drive the transmission of parental characteristics to children can be considered by unpacking the individual background effect into the several 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), dummies capturing the "number of books at home". Thirdly, compared to school-level regressions, "peer composition" variables are net of the individual ones and are based on six students' characteristics: sex, age, immigrant and "foreign language" status, type of school programme (general or vocational) and the "escs" index. Finally, in an extended model, we also include additional controls proxying the effort devoted in studying science (see table A1).

Student-level regressions might lead to biased estimates in so far as missing values on certain individual characteristics are not randomly distributed, but turn out to be related to background and ability. As a result, dropping students with missing information for 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). Thereafter, we regress each variable subjected to the imputation procedure on basic controls available for nearly all students (age, sex, grade, dummies "vocational" and "isced 3", two country-level controls – GDP and expenditure on education per capita – and the number of books at home) and replace missing values with predicted ones. Once having replaced missing values with imputed ones, in all student-level regressions carried on we correct for the measurement error that could arise due to the imputation procedure by allowing the observations with missing data on each variable to have their own intercepts and slopes (Woessmann 2004)[15]. As an additional methodological caveat, 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).

Table 7 shows OLS estimations for the pooled sample of OECD countries[16]. With respect to school-level estimates, the impact of heterogeneity is also negative but at a significant level around the cut-off level of 15% (ST-1, table 7), whereas it is not significant at all when non-linear effects are included (ST-2 and ST-3, table 7). The size of the heterogeneity effect only slightly decreases

[15] In particular, we include a dummy that takes the value 1 for an imputed data and 0 for observations with original data and an interaction term between this imputation dummy and the respective variable subjected to the imputation procedure

[16] To obtain representative coefficient estimates from the stratified survey data – as in section 4, where regressions were ran using schools' sample weights provided in the PISA dataset, students' sample weights are used in all estimations of sections 5-6.

from around 1.9 to around 1.3 points of a full standard deviation in the test scores. In turn, linear peer effects are significant and very large with a change in one standard deviation of the "escs" accounting for more than a 20% change in the standard deviation of the science test (ST-1, table7). Whereas the first result is somehow in line with the one of the previous literature finding small (but insignificant!) effects of heterogeneity on student performance (e.g. Hanushek et al. 2003; Rangvid 2007), the estimated impact of the linear peer effect is larger than the bulk ones founded in the literature (see Ammermueller and Pischke 2006). However, when we adopt a more precise identification strategy to isolate the linear peer effect (see §2 and ST-0, table 7; i.e. including school fixed effects and school- and student-level additional controls and excluding heterogeneity terms), the estimated effect decreases up to around 18%, closer to the impact found by other studies using PISA surveys (e.g. Rangvid 2007; Schneeweis and Winter-Ebmer 2007).

Note that the R^2 reduces compared to the very high level (around 60%) shown in school-level regressions. This is expected since a large part of the performance variation across students has to be attributed to unobserved variables (e.g. their innate ability or learning motivation). However, its level, around 34%, is in line with the one of the two studies using a large set of controls to reduce the omitted variable bias in the estimation of peer effects (Rangvid 2007; Schneeweis and Winter-Ebmer 2007).

As in school-level analysis, the picture substantially changes when the interaction between the heterogeneity and the tracking system is added (model ST-4, table 7). Again, this interaction is positive and significant suggesting that in early-tracking countries heterogeneity can foster students' performances, even once controlling for other school composition aspects. School-level results are also confirmed in separate regressions with a higher heterogeneity being significant with opposite signs in systems with early-tracking (+) and comprehensive (-) schools (ST-1, table8). Looking at table 7, results remain robust to different classifications of the countries by tracking (ST-1A) and to the inclusion of country fixed effects (ST-6). Moreover, the difference between the two tracking systems is further more evident when the quadratic heterogeneity term is also included (model ST-5): with respect to school regressions both the inverted U-shaped relationship – again increasing up to median level of the "escs" standard deviation – for early-tracking countries and the insignificance of the polynomial function for comprehensive ones are confirmed at the student-level. It has to be emphasized that moving from school- to student-level regressions estimated signs and sizes of the heterogeneity terms remains the same, hence the impact of heterogeneity should be correctly identified. Finally, consistently with the theoretical literature (e.g. Brunello et al. 2007) and with Entrof and Lauer (2006) – but with a focus on the effect of immigrant peers – separate regressions display a larger (linear) peer effect in early tracking systems (table 8).

Interestingly, in comprehensive school systems, the interaction term between heterogeneity and the country-level share of students enrolled in vocational programs displays a highly negative and significant coefficient suggesting that the negative impact of heterogeneity on student performance is largely driven by schools offering vocational programs (model ST-5, table 8). In contrast, in early-tracking ones, the relationship between heterogeneity and performances is not driven by the share of students enrolled in vocational programmes. All in all, this finding has a strong policy implication in so far as, also in comprehensive systems, the impact of higher background heterogeneity appears to be negative only in a minority of schools oriented towards training[17].

Adding further school and student controls (i.e. admission procedures, school accountability and proxies of individual efforts, see table A1) corrects for the omitted variable bias in the estimation of school composition variables. In this case, the significance of the two opposite effects of the "escs standard deviation" slightly decreases but still emerges, whereas, as expected, the impact of the linear peer effect is mitigated (model ST-7, table 9).

However, correcting for the omitted variable bias might not be sufficient to attenuate the selectivity bias in the estimation of peer effects if unobservable schooling characteristics are still present (see section 3). In order to attempt a better identification of the linear peer effect, following Ammermueller and Pischke (2006)[18] and according to the empirical strategy described in section 2, we add school-level fixed effects, identified, for each country, by the quantile of the average parental "escs" distribution to which the school belongs to (models ST-8 – ST-9, table 9). The linear peer effect reduces in size but only in countries tracking earlier, whereas its size remains unchanged in countries with comprehensive school. This implies that the identification strategy of linear peer effect suggested by Ammermueller and Pischke (2006) is particularly suitable for early-tracking systems where the early selection process might create more homogeneous but "less observable" school types.

Concerning the impact of heterogeneity, in comprehensive systems a strong difference emerges comparing models ST-1, ST-7 and ST-8 (tables 8 and 9); indeed, when variables about students' time of work and school sorting are added, the significance of the negative heterogeneity effect strongly reduces and it disappears when school fixed effects are included too. In turn, in the most complete model (ST-9) where both types of additional controls and the quadratic heterogeneity term are included, the inverted U-shaped relationship between heterogeneity and student's competences in science is confirmed for early-tracking countries.

[17] However, the share of students enrolled in vocational programs is zero in several countries considered, hence this result is difficult to interpret. Using the share of students enrolled in schools that mainly offer training, a higher share of students attending schools which offer training also leads to a significantly negative effect of heterogeneity in comprehensive systems. Further details are available upon request by the authors.

[18] Also Schneeweiss and Winter-Ebmer (2007) include a school fixed effect in their analysis of peer effects in Austria.

So far, using OLS techniques, we have focused on average peer and heterogeneity effects. This standard methodology may neglect how school composition affects achievements differently at different points of the conditional test score distribution and hence might lead to misleading policy implications. For instance, while the effect of peer heterogeneity may not be significant for average test scores, it is useful to know whether this effect is also not significant for all quantiles of the conditional test score distribution (Rangvid 2007).

In order to answer this question, we use quantile regressions to estimate model ST-1 separated for early and late tracking countries. Quantile regressions confirm that the linear peer effect is positive all along the conditional test score distribution and remarkably higher in early tracking countries (figure 4 and table 10). Moreover, consistently with previous studies limited to Austria (Schneeweiss and Winter-Ebmer 2007) and Denmark (Rangvid 2007), it is slightly larger in lower deciles. As expected, main differences between the two groups of countries emerges with respect to the influence of peer heterogeneity (figure 5 and table 10). In all deciles, the "escs" standard deviation is always statistical significant at the 99% level. However, its sign is largely positive and slightly U-shaped along the entire test score distribution in early tracking systems, while it remains always negative in comprehensive systems where the size of the negative effect is only slightly lower in upper deciles.

In sum, quantile regressions reinforce the previous finding in terms of a small 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 the one hand, stronger peer effects at the bottom of the ability distribution would lead to support policies aimed at increasing background heterogeneity[19]; on the other hand, a too high heterogeneity turns out to offset the efficiency-enhancing effect of mixing background. For policy purposes, the effect of school-composition variables on efficiency should be seen together with the one on equity; this is the objective of next section.

6. Peer heterogeneity and equality of opportunity

In this section, our focus moves to the effect of peer heterogeneity on equalizing the attainments of students from different backgrounds; in particular, we want to analyse the extent to which the

[19] However, a caveat is required here. The effect of regrouping students by background should be balanced against the associated regrouping of students by ability. It might be that the regrouping would bring about peer effects due to interactions of individuals of different abilities that offset or amplify the ones due to the interactions of individuals from different backgrounds. Since we can not disentangle ability peer effect from background ones, policy implications are less clear cut.

theoretical prediction that a heterogeneous school environment tends to level opportunities of pupils from different backgrounds (e.g. Benabou 1996) is empirically warranted. A way to answer this question empirically consists in assessing whether the family background effect, i.e. the link between individual performances and family background, is linked to peers' average and heterogeneity.

The existing literature (Woessmann 2007; Schuetz et al. 2008; Brunello and Checchi 2007) interprets the coefficient of a synthetic variable of parental background as a proper measure of inequality of educational opportunities using a parsimonious, reduced-form specification of the determinants of student performance (see section 2). Following this literature, we run regressions where only individual characteristics are included among the control variables, whereas the family background is summarized in a single variable – the student's parental escs index[20]. In particular, in order to analyse the differential impact of school compositional variables and tracking on family background, we interact the individual family background effect with the early-tracking dummy, the average and the standard deviation of school parental escs respectively. Also consistently with the existing literature, we interact the individual escs with possible confounding factors in order to isolate the pure effect of heterogeneity and tracking on background. These confounding factors are four country-level features: the duration of pre-primary school, the share of public schools, the average students/teachers ratio and the per-capita spending in education[21]. Besides, following Schuetz et al. (2008), we run two different sets of regressions, respectively including or excluding country fixed effects.

Without including country fixed effects (table 11), the usual result that an earlier tracking widen the opportunity gap between student from different background is strongly confirmed even if all the caveat due to the incorrect identification of the true effect of tracking in cross-sections should be kept in mind here (Hanushek and Woessmann 2006; Ammermueller 2005; Waldinger 2006). More to the point, the "escs" coefficient is twice as large in early tracking with respect to comprehensive systems in separated regressions, whereas the interaction term between tracking and "escs" is large, positive and significant in pooled ones. However, it is worth to emphasize that such significance disappears when interactions between "escs" and school composition variables, i.e. the degree of heterogeneity, are also added. This finding adds new insights to the growing literature on tracking and equality of opportunities (see Brunello and Checchi 2007) since the effect of tracking appears as spurious and largely driven by school compositional variables.

[20] Results presented in tables 6 and 7 are robust to the use of different background variables instead or the escs (e.g. parental highest occupational status or educational attainment). Detailed results are available upon request by authors.

[21] The interaction with the share of students enrolled in pre-primary school (a further potential confounding factor highlighted by Schuetz et al. 2008 and Brunello and Checchi 2007) has not been included since we did not found reliable data for Korea and Ireland.

Looking to the effect of school compositional variables per se, in all the empirical specifications considered (table 11) a higher average background significantly reinforces the peer impact on the individual attainment, while the opposite happens regarding the impact of schoolmates heterogeneity: i.e. a higher heterogeneity offsets the impact of family background. Interestingly, this reduction is higher in early tracking countries where the socio-economic gradient is much higher.

When country fixed effects are included (table 12), school compositional variables keep the same sign and high significance, but the size of the interaction between the “escs” and the degree of heterogeneity becomes similar in the two tracking systems. More puzzling is the inversion in the size of the family background effect that turns out to be higher in comprehensive schooling systems.

All in all, school compositional variables affect equity in the way expected by the theory (e.g. Benabou 1996). In turn, including these variables makes the negative impact of early tracking on opportunity equalization less limpid suggesting that school compositional variables should be included in future, more detailed analyses. Finally, the effect of mixing student by background appears socially desirable both in terms of equity and efficiency in early tracking systems.

7. Concluding remarks

The main effort of this paper has been devoted to study the impact of peer heterogeneity in different tracking regimes. It has been shown that peer heterogeneity does have an impact on both efficiency and equity. Whereas a higher heterogeneity leads to a substantial levelling of the educational opportunities in both systems, the impact of heterogeneity on efficiency is opposite in schooling systems with different school tracking policies. In early-tracking systems, peer heterogeneity has a positive but non-linear impact on student outcomes. In comprehensive ones, instead, heterogeneity negatively affects student outcomes but this result is largely driven by pupils attending vocationally-oriented programs. This result holds both in school- and in individual-level regression leading us to conclude that the effect of heterogeneity is correctly identified. In turn, the linear impact of peers is far larger in early tracking systems and seems correctly identified either by adding controls correlated with the school selection process or by using school-level fixed effects.

All these findings point, as a possible explanation, to a different way in which the tracking age affects the sorting of students by unobservable characteristics. For instance, in order to avoid the vocational streaming, better students might put more efforts to signal their higher abilities and motivations sooner in early tracking systems. If this is the case, the unobservable degree of heterogeneity should be lower in early tracking systems and, hence, policies attempting to enhance the opportunities of disadvantaged students should intervene before tracking occurs. Further empirical researches should investigate more carefully the effect of early tracking and school admittance policies on student sorting by both ability and background.

A final caveat is required to use these results for policy purposes. The significant impact of peer heterogeneity on student performance is rather small both in comprehensive and in early-tracking systems, 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 selection of student by schools of different quality, 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

Ammermüller A., Pischke J.S. (2006). "Peer effects in European primary schools: Evidence from PIRLS", *NBER Working Paper*, n. 12180.

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

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

Bishop J.H. (1997), "The Effect of National Standards and Curriculum-Based Exams on Achievement", *American Economic Review*, 87, 260–264.

Bishop J.H., Woessmann L. (2004), "Institutional Effects in a Simple Model of Educational Production", *Education Economics*, 12, 17–38.

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 P. Peterson and L. Wößmann (eds.), *Schools and the Equal Opportunity Problem*, Cambridge MA, MIT Press.

Carneiro P., Heckman J. (2003), Human capital policy", in Heckman J., Krueger A. (eds.), *Inequality in America: what role for human capital policies?*, Cambridge MA, MIT Press.

Checchi D., Flabbi L. (2005). "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*, 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.

Fernandez R., Rogerson R. (1997) "Keeping People Out: Income Distribution, Zoning, and the Quality of Public Education", *International Economic Review*, 38, 23-42.

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.

Gibbons S., Telhaj S. (2005), "Peer effects and pupil attainment: Evidence from secondary school transition", *London School of Economics*, mimeo.

Hall P., Soskice D. (2001), eds., *Varieties of capitalism*, Oxford University Press.

Hanushek, E.A. (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., Woessmann L. (2007), " The Role of Education Quality for Economic Growth", *World Bank Policy Research Working Paper*, n. 4122.

Hanushek E., Woessmann L. (2008), "The Role of Cognitive Skills in Economic Development", *Journal of Economic Literature*, 46, 607-668.

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

Hoxby C. (2000), "The effects of class size on student achievement: New evidence from population variation", *Quarterly Journal of Economics*, 115, 1239-1285.

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

Krueger, D., Kumar K. (2004), "U.S.-Europe Differences in Technology-Driven Growth: Quantifying the Role of Education", *Journal of Monetary Economics*, 51, 161-190.

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.

OECD - Organisation for Economic Co-operation 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, 359-388.

Schneeweiss N., Winter-Ebmer R. (2007), "Peer effects in Austrian schools", *Empirical Economics*, 3, 387-409.

Schuetz G., Ursprung H., Woessmann L. (2008), "Education Policy and Equality of Opportunity", *Kyklos*, 61, 279-308.

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. (2003), "Schooling resources, educational institutions, and student performance: The international evidence", *Oxford Bulletin of Economics and Statistics*, 65, 117-170.

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. (2007), "Fundamental Determinants of School Efficiency and Equity: German States as a Microcosm for OECD Countries", *IZA Discussion Paper*, n. 2880.

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

Tab. 1: Descriptive statistics of PISA 2006 selected variables

		Age Track	Grade Track	Mean Score Science	Mean Parent. Edu. (Pared)	Mean ESCS	Pared Std. Dev. (mean by school)	ESCS Std. Dev. (mean by school)	Share Immigr.	Share Foreign. Lang.	Share of Public	Share of no ability track	Share of stud Vocat.	Share admitted by stud. record	Share admitted by residence
Early Tracking	Austria	10	4	518.4	13.8	0.23	2.05	0.69	0.12	0.10	0.87	0.60	0.44	0.69	0.22
	Czech Rep.	11	5	520.5	13.5	0.07	1.86	0.66	0.02	0.02	0.91	0.35	0.43	0.45	0.20
	Germany	10	4	523.1	14.2	0.32	2.86	0.78	0.17	0.13	0.89	0.59	0.00	0.40	0.65
	Hungary	11	4	514. 1	12.8	-0.01	2.13	0.71	0.02	0.01	0.81	0.31	0.62	0.70	0.01
	Slovak Rep.	11	4	493.7	13.3	-0.10	2.19	0.74	0.00	0.14	0.88	0.25	0.46	0.50	0.17
	Belgium	12	6	512.5	13.8	0.18	2.67	0.79	0.13	0.19	0.37	0.56	0.47	0.26	0.02
	Netherlands	12	6	525.8	13.7	0.25	2.61	0.78	0.11	0.07	0.32	0.19	0.30	0.66	0.10
	Luxembourg	13	6	486.3	13.1	0.10	3.91	0.96	0.37	0.91	0.85	0.27	0.13	0.41	0.42
	Switzerland	12	6	515.1	13.4	0.10	2.94	0.79	0.23	0.19	0.93	0.24	0.07	0.54	0.82
Late Tracking	Italy	14	8	479.0	12.5	-0.05	3.06	0.82	0.04	0.13	0.92	0.54	0.57	0.07	0.11
	Korea	14	9	522. 9	13.2	-0.01	2.30	0.70	0.00	0.00	0.55	0.12	0.24	0.60	0.22
	Greece	15	9	481.3	13.4	-0.09	2.92	0.80	0.05	0.03	0.95	0.89	0.15	0.05	0.71
	Ireland	15	6	508.6	12.9	-0.01	2.21	0.74	0.08	0.06	0.40	0.02	0.02	0.02	0.42
	Japan	15	9	532.0	14.0	-0.01	1.72	0.60	0.00	0.00	0.74	0.44	0.24	0.87	0.20
	Portugal	15	9	488.7	9.9	-0.52	4.19	1.01	0.05	0.02	0.89	0.48	0.14	0.06	0.56
	Australia	16	10	527.3	13.2	0.21	1.83	0.68	0.22	0.09	0.00	0.05	0.10	0.09	0.42
	Canada	16	8	536.3	14.7	0.37	2.28	0.71	0.23	0.15	0.85	0.08	0.00	0.11	0.78
	Denmark	16	9	495.0	14.0	0.30	2.46	0.82	0.08	0.07	0.63	0.16	0.00	0.03	0.55
	Finland	16	9	564.0	14.4	0.25	2.34	0.75	0.02	0.02	0.96	0.49	0.00	0.04	0.75
	Icelend	16	10	489.7	15.1	0.82	2.74	0.81	0.03	0.03	0.96	0.15	0.00	0.01	0.94
	Norway	16	10	485.6	13.8	0.43	1.73	0.68	0.08	0.08	0.96	0.58	0.00	0.00	0.78
	New Zealand	16	6	529.3	12.8	0.09	2.10	0.72	0.21	0.09	0.93	0.03	0.00	0.10	0.50
	Poland	16	9	497.7	12.2	-0.31	1.75	0.75	0.00	0.01	0.97	0.53	0.00	0.13	0.83
	Spain	16	10	488.9	11.1	-0.31	3.53	0.88	0.07	0.16	0.53	0.29	0.00	0.03	0.68
	Sweden	16	9	502.9	13.8	0.23	2.11	0.70	0.11	0.09	0.88	0.24	0.00	0.01	0.58
	UK	16	12	514.6	13.7	0.19	1.97	0.69	0.10	0.06	0.74	0.00	0.00	0.10	0.61
	US	16	12	491.0	13.6	0.15	1.96	0.75	0.17	0.12	0.87	0.12	0.00	0.08	0.81

Source: elaborations on PISA 2006 data.

Tab. 2: Correlation matrix between Background Measures

	escs	pared	hisei	homepos	books at home
escs	1				
pared	0.769	1			
hisei	0.796	0.461	1		
homepos	0.708	0.321	0.339	1	
books at home[1]	0.499	0.323	0.327	0.524	1

[1] The variable books at home has been linearized. Source: elaborations on PISA 2006 data.

Tab. 3: Correlation between escs mean and. escs std. dev.

	Escs standard deviation	Escs standard deviation related to country escs S.D.
Early tracking	-0.005	-0.024
Comprehensive	-0.222	-0.151
All countries	-0.171	-0.119

[1] The variable books at home has been linearized. Source: elaborations on PISA 2006 data.

Tab. 4: Descriptive statistics of PISA 2006 selected school-level variables

	Early Tracking	Early2: grade track<6	Comprehensive
Mean SCIE	502.6	502.9	498.8
Std. Dev. SCIE	73.6	73.3	54.5
Std. Dev. Escs average	0.49	0.49	0.51
Std. Dev. Pared average	1.3	1.3	1.5
Average escs std. dev.	0.75	0.75	0.75
Average pared std. dev.	2.6	2.6	2.2
Share of students attending vocational	0.21 (0.39)	0.20 (0.39)	0.1 (0.30)
Share of immigrants	0.14 (0.19)	0.14 (0.19)	0.07 (0.14)
Share of students speaking foreign languages	0.13 (0.17)	0.13 (0.17)	0.07 (0.14)
Share of school no sorting students by ability	0.44 (0.50)	0.43 (0.49)	0.29 (0.45)
Share of schools that admit according to students records	0.43 (0.49)	0.42 (0.49)	0.19 (0.39)
Share of schools that admit according to residence	0.46 (0.50)	0.45 (0.50)	0.53 (0.50)

Source: elaborations on PISA 2006 data.

Tab. 5: School average performances in science in OECD countries[1]. OLS regressions[2, 3].

	SC-1	SC-2	SC-3	SC-4	SC-5
Escs average	81.25 (2.74) *0.000*	81.07 (2.76) *0.000*	81.04 (2.73) *0.000*	83.61 (2.44) *0.000*	80.40 (2.66) *0.000*
Escs standard deviation		-13.38 (7.15) *0.061*	102.39 (57.47) *0.075*	106.39 (50.20) *0.034*	-26.76 (7.00) *0.000*
Escs standard deviation^2			-66.85 (31.04) *0.031*	-68.25 (27.55) *0.013*	
Early track* Escs standard deviation					39.05 (15.07) *0.010*
Groups of Control Variables					
School Location and Class Size	yes	yes	yes	yes	yes
School Composition	yes	yes	yes	yes[4]	yes
School Resources	yes	yes	yes	country average	yes
School Institutions	yes	yes	yes	country average	yes
Country level controls	yes	yes	yes	yes	yes
Number of observations	5,831	5,831	5,831	6,482	5,831
F	87.6	85.7	84.0	85.0	83.7
Prob.>F	0.000	0.000	0.000	0.000	0.000
R^2	0.6235	0.6245	0.6259	0.6195	0.6297

[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. [4]The share of students enrolled in vocational programme is considered as country average. Source: elaborations on PISA 2006 data.

Tab. 6: School average performances in science in OECD countries[1] by early and no early tracking countries[2]. OLS regressions[3, 4].

	SC-2		**SC-2A**[5]		**SC-3**		**SC-4**		**SC-6**		**SC-7**	
	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	No early track	Early track
Escs average	74.75	81.17	75.62	84.36	74.74	80.47	77.16	95.27	72.09	80.35	76.30	78.18
	(3.21)	(4.29)	(3.25)	(4.22)	(3.20)	(4.30)	(2.84)	(3.56)	(3.40)	(4.33)	(3.36)	(5.01)
	0.000	*0.000*	*0.000*	*0.000*	*0.000*	*0.000*	*0.000*	*0.000*	*0.000*	*0.000*	*0.000*	*0.000*
Escs standard Deviation	-31.22	11.70	-31.00	12.63	-39.63	136.28	-11.52	99.54	-37.04	135.78	-32.64	15.34
	(6.71)	(8.15)	(6.77)	(8.15)	(56.40)	(57.55)	(54.93)	(60.00)	(56.85)	(57.47)	(6.89)	(8.89)
	0.000	*0.151*	*0.000*	*0.121*	*0.482*	*0.018*	*0.834*	*0.097*	*0.515*	*0.018*	*0.000*	*0.084*
Escs standard deviation^2					4.80	-72.88	-9.95	-53.24	1.08	-72.81		
					(31.10)	(32.67)	(30.34)	(34.57)	(31.42)	(32.59)		
					0.877	*0.026*	*0.743*	*0.124*	*0.973*	*0.026*		
Groups of Control Variables												
Sc. Loc & Size	yes	yes	yes	yes	yes	yes	Yes	yes	yes	yes	yes	yes
Sc. Comp.	yes	yes	yes	yes	yes	yes	yes[6]	yes[6]	yes	yes	yes	yes
Sc. Resources	yes	yes	yes	yes	yes	yes	cnt average	cnt average	yes	yes	yes	yes
Sc. Institutions	yes	yes	yes	yes	yes	yes	cnt average	cnt average	yes	yes	yes	yes
Country level controls	yes	yes	yes	yes	yes	yes	Yes	yes	yes	yes	yes	yes
Sc. additional controls											yes	yes
Country F.E.									yes	yes		
Number of obs.	4,319	1,512	4,153	1,678	4,319	1,512	4,835	1,647	4,319	1,512	3,846	1,252
F	54.5	98.0	54.8	97.4	54.1	94.7	70.4	122.6	65.9	94.4	42.8	78.4
Prob.>F	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
R^2	0.5766	0.8046	0.5795	0.7957	0.5766	0.8059	0.5695	0.7863	0.6079	0.806	0.5804	0.8121

[1] Mexico, Turkey and France are not included. [2] Early track countries are considered those countries where the age of first school tracking is before 13. [3] Regressions are run using school sample weights provided in PISA database. [4] 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. [5] In model SC-2A the split between early and no early track countries is the one proposed by Waldinger (2006), coded in variable early_track2 (see tab. A1). [6]The share of students enrolled in vocational programme is considered as country average. Source: elaborations on PISA 2006 data.

Tab. 7: Students performances in science in OECD countries[1]. OLS regressions[2, 3].

	ST-1	ST-2	ST-3	ST-4	ST-0
Peer Escs average	49.57	49.63	48.95	50.10	38.67
	(2.54)	(2.50)	(2.71)	(2.54)	(6.64)
	0.000	*0.000*	*0.000*	*0.000*	*0.000*
Escs standard Deviation	-9.53	53.35	23.97	-16.62	
	(6.54)	(55.23)	(50.01)	(8.08)	
	0.145	*0.334*	*0.632*	*0.040*	
Escs standard Deviation^2		-36.02	-21.22		
		(30.82)	(28.06)		
		0.243	*0.450*		
Early track* Escs standard deviation				24.68	
				(12.53)	
				0.049	
Groups of Control Variables[4]					
Individual Characteristics	yes	yes	yes	yes	yes
Family background	yes	yes	yes	yes	yes
Peer composition	yes	yes	yes	yes	yes
School Location and size	yes	yes	yes	yes	yes
School Resources	yes	yes	country average	yes	yes
School Institutions	yes	yes	country average	yes	yes
Country level controls	yes	yes	yes	yes	yes
School fixed effects[5]					yes
School additional controls					yes
Students additional controls					yes
Number of observations	174,921	174,921	193,467	177,795	126,949
F	175.5	173.9	208.1	174.1	212.6
Prob.>F	0.000	0.000	0.000	0.000	0.000
R^2	0.3399	0.3400	0.3477	0.3393	0.4183

[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] School fixed effects are identified, for each country, according to the quintile of escs average of every school. Source: elaborations on PISA 2006 data.

Tab. 8: Students performances in science in OECD countries[1] by early and no early tracking countries[2]. OLS regressions[3, 4].

	ST-1		ST-1A[5]		ST-2		ST-5		ST-6	
	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
Peer Escs average	44.96 (2.92) *0.000*	60.36 (3.71) *0.000*	45.28 (2.94) *0.000*	60.12 (3.51) *0.000*	45.00 (2.89) *0.000*	60.07 (3.70) *0.000*	45.35 (2.87) *0.000*	60.44 (3.71) *0.000*	41.66 (3.13) *0.000*	60.09 (3.71) *0.000*
Escs standard Deviation	-17.68 (7.57) *0.020*	13.58 (8.57) *0.113*	-16.79 (7.64) *0.028*	13.25 (8.27) *0.109*	7.51 (71.23) *0.916*	110.59 (52.95) *0.037*	-10.16 (8.05) *0.207*	14.66 (9.80) *0.135*	-22.84 (7.59) *0.003*	14.40 (8.65) *0.096*
Escs Standard Deviation^2					-14.36 (39.37) *0.715*	-56.32 (29.48) *0.056*				
Vocational*Escs standard deviation							-68.25 (20.55) *0.001*	-6.68 (15.24) *0.661*		
Groups of Control Variables[4]										
Individual characteristics	yes	yes	yes	yes	yes	yes	yes	yes	yes	yes
Family Background	yes	yes	yes	yes	yes	yes	yes	yes	yes	yes
Peer composition	yes	yes	yes	yes	yes	yes	yes	yes	yes	yes
School Location and Size	yes	yes	yes	yes	yes	yes	yes	yes	yes	yes
School Resources	yes	yes	yes	yes	yes	yes	yes	yes	yes	yes
School Institutions	yes	yes	yes	yes	yes	yes	yes	yes	yes	yes
Country level controls	yes	yes	yes	yes	yes	yes	yes	yes	yes	yes
Country fixed effects									yes	yes
Number of observations	132,104	42,817	124,046	50,875	132,104	42,817	132,104	42,817	132,104	42,817
F	121.9	149.9	119.8	149.2	121.3	146.9	122.6	148.0	133.6	148.2
Prob.>F	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
R^2	0.314	0.5455	0.3143	0.5323	0.314	0.5459	0.3147	0.5455	0.3245	0.5459

[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] In model ST-1A the split between early and no early track countries is the one proposed by Waldinger (2006), coded in variable early_track2 (see tab. A1). Source: elaborations on PISA 2006 data.

Tab. 9: Students performances in science in OECD countries[1] by early and no early tracking countries[2]. OLS regressions[3, 4], including additional school and student controls and quintile of school escs average by country fixed effects

	ST-7		**ST-8**		**ST-9**	
	No early track	Early track	No early track	Early track	No early track	Early track
Peer Escs average	38.13 (3.11) *0.000*	53.76 (3.71) *0.000*	37.70 (7.59) *0.000*	48.05 (8.37) *0.000*	37.40 (7.48) *0.000*	48.08 (8.37) *0.000*
Escs standard deviation	-10.48 (7.26) *0.149*	12.00 (8.18) *0.143*	-6.73 (7.16) *0.348*	11.33 (8.49) *0.182*	-40.83 (49.49) *0.409*	127.95 (51.49) *0.013*
Escs Standard Deviation^2					19.23 (27.27) *0.481*	-67.50 (28.19) *0.017*
Groups of Control Variables[4]						
Individual characteristics	yes	yes	yes	yes	yes	yes
Family Background	yes	yes	yes	yes	yes	yes
Peer composition	yes	yes	yes	yes	yes	yes
Sc. Loc & Size	yes	yes	yes	yes	yes	yes
Sc. Resources	yes	yes	yes	yes	yes	yes
Sc. Institutions	yes	yes	yes	yes	yes	yes
Country level controls	yes	yes	yes	yes	yes	yes
School fixed effects[5]			yes	yes	yes	yes
School additional controls	yes	yes	yes	yes	yes	yes
Students additional controls	yes	yes	yes	yes	yes	yes
Number of observations	94,656	32,293	94,656	32,293	94,656	32,293
F	113.0	119.9	93.2	108.3	93.0	106.6
Prob.>F	0.000	0.000	0.000	0.000	0.000	0.000
R^2	0.3797	0.5778	0.3981	0.5824	0.3982	0.5831

[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] School fixed effects are identified, for each country, according to the quintile of escs average of every school. Source: elaborations on PISA 2006 data.

Tab. 10: Students performances in science in OECD countries[1] by early and no early tracking countries. Estimated coefficients by model ST-1 of "Peer escs average" and "Escs standard deviation". Quantile regressions[2].

	Peer Escs average		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.

Tab. 11: School composition, educational policies and inequality of opportunity: interactions with student level family background effects. OLS regressions. No country fixed effects.

	Pooled		No early track countries	Early track countries
Escs	26.32***	50.64***	40.13***	105.26***
escs*early track	2.87*	1.18		
escs*peer escs average		9.93***	10.10***	11.09***
escs*escs standard deviation		-21.09***	-17.72***	-29.40***
Controls				
Individual characteristics	Yes	Yes	Yes	Yes
Interactions with counfounding factors				
Escs*Dur_preprimary	-2.50***	-1.94***	-2.27***	-8.80***
Escs*Public_cnt	1.14	0.14	1.96	16.50***
Escs*Stratio_cnt	1.29***	0.99***	1.46***	-1.16**
Escs*Educ_spending	0.00	0.00*	0.00	0.00***
Country Fixed effects	No	No	No	No
Number of observations	202,804	202,804	154,719	48,085
F	447.5	393.4	322.2	245.9
Prob.>F	0.000	0.000	0.000	0.000
R^2	0.1839	0.1879	0.1746	0.3168

[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; ** 95% significance level. * 90% significance level. Source: elaborations on PISA 2006 data.

Tab. 12: School composition, educational policies and inequality of opportunity: interactions with student level family background effects. OLS regressions. Country fixed effects.

	Pooled		No early track countries	Early track countries
Escs	25.57***	55.42***	44.02***	94.90***
escs*early track	-3.91***	-5.91***		
escs*peer escs average		11.66***	11.73***	10.99***
escs*escs standard deviation		-26.66***	-25.33***	-28.23***
Controls				
Individual characteristics	Yes	Yes	Yes	Yes
Interactions with counfounding factors				
Escs*Dur_preprimary	-2.56***	-1.93***	-1.85***	-5.42**
Escs*Public_cnt	4.67**	3.64*	7.69***	12.11***
Escs*Stratio_cnt	1.18***	0.85***	1.51***	-1.24**
Escs*Educ_spending	0.00	0.00**	0.00***	0.00***
Country Fixed effects	Yes	Yes	Yes	Yes
Number of observations	202,804	202,804	154,719	48,085
F	232.7	222.6	237.5	220.5
Prob.>F	0.000	0.000	0.000	0.000
R^2	0.2328	0.2385	0.2217	0.3361

[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; ** 95% significance level. * 90% significance level. Source: elaborations on PISA 2006 data.

Fig. 1: Educational achievements and "escs" standard deviation. Source: elaborations on PISA 2006 data.

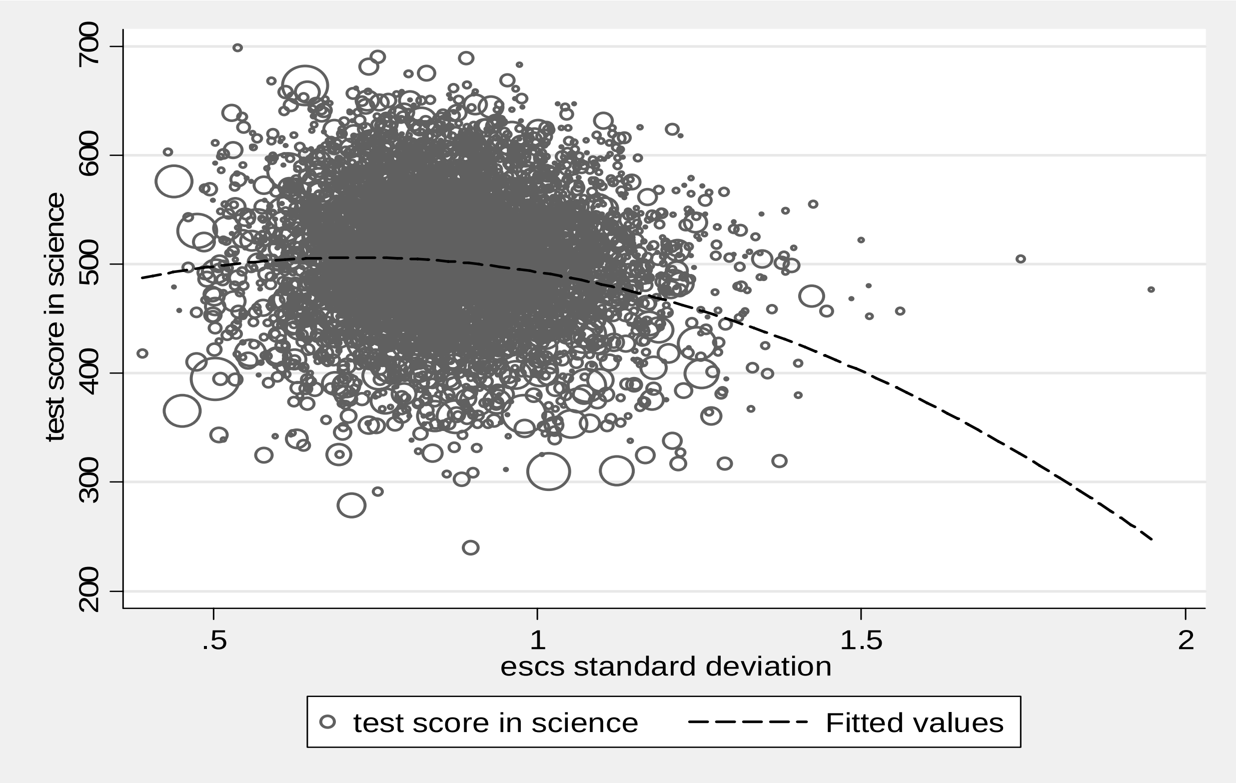


Fig. 2: Educational achievements and “escs” standard deviation by tracking policies. Source: elaborations on PISA 2006 data.

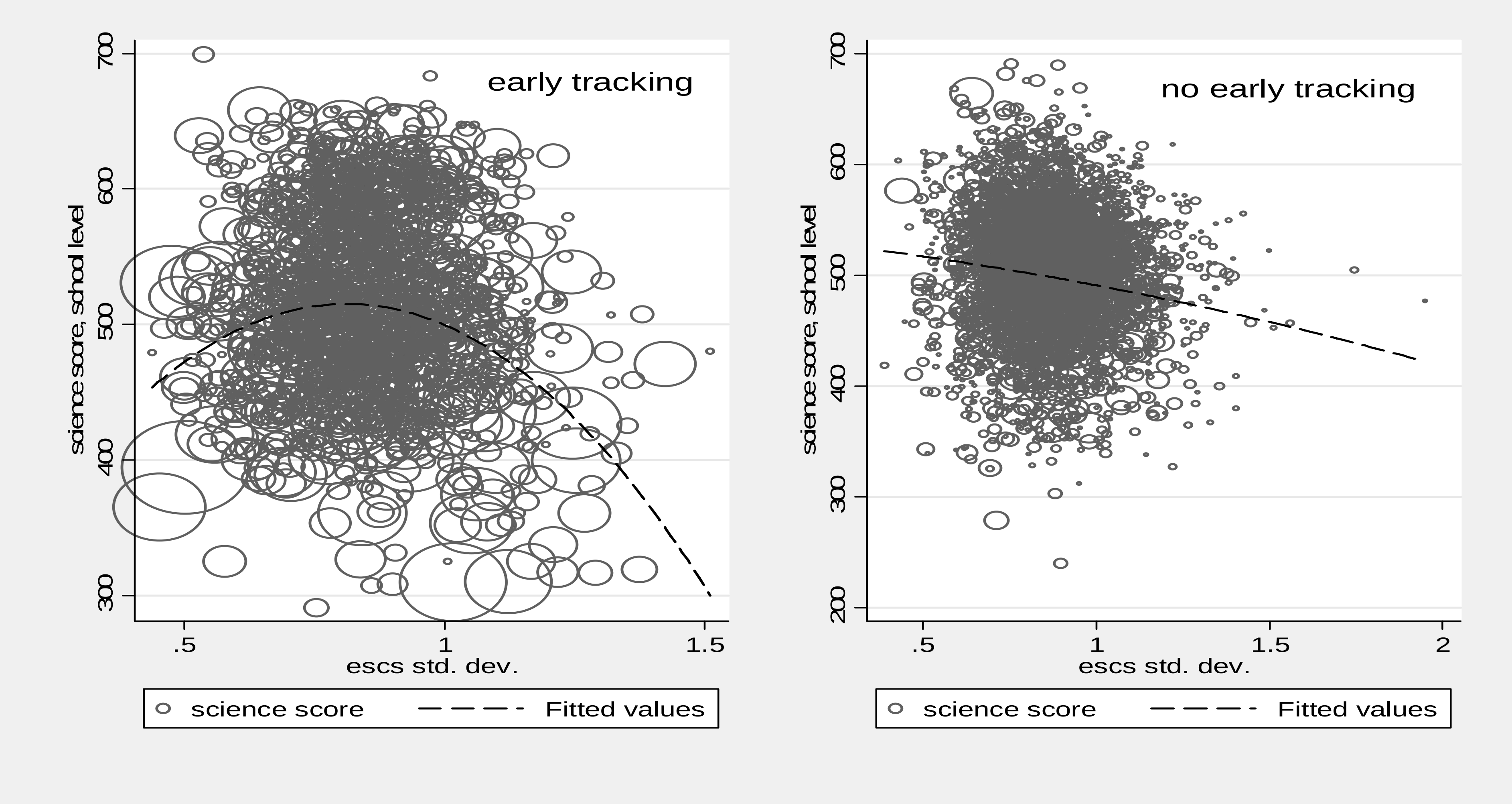


Fig.3: Relationship between “escs average” and “escs standard deviation”. Source: elaborations on PISA 2006 data.

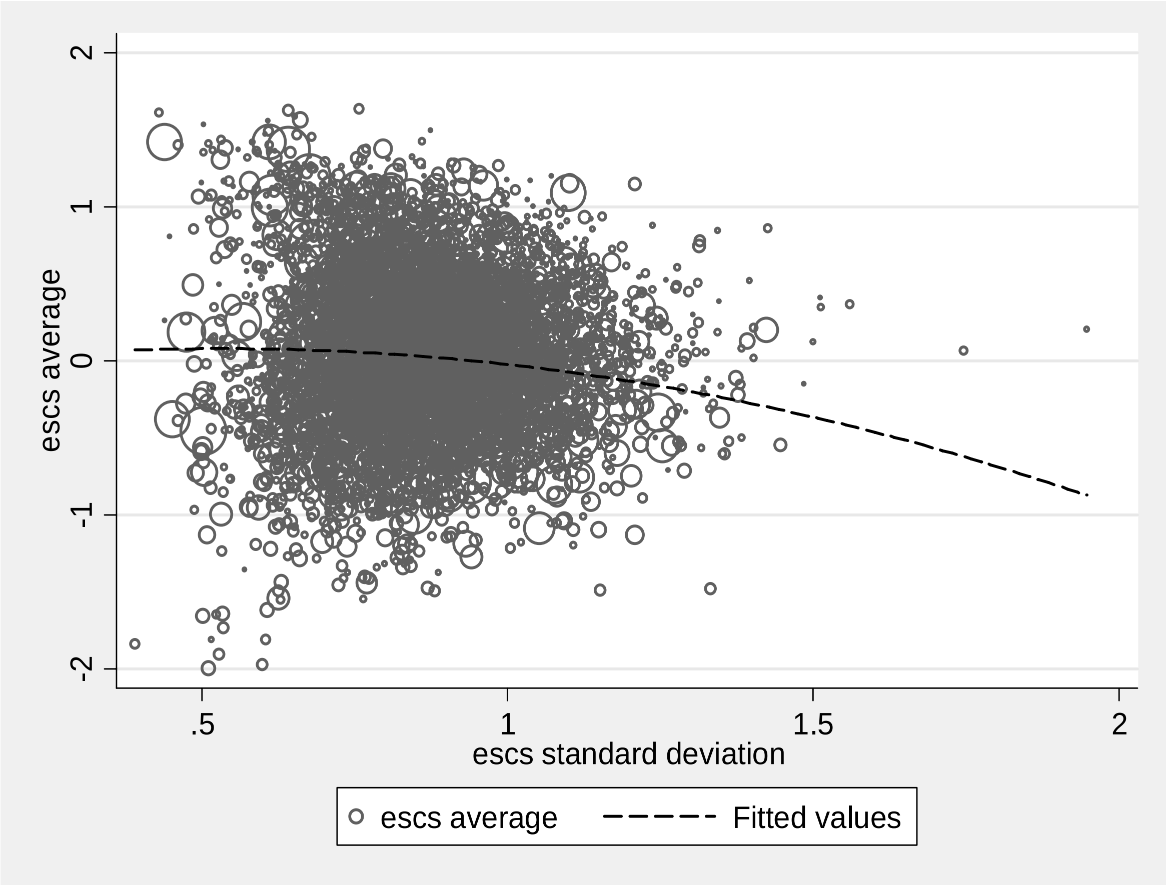


Fig. 4: Quantile regressions on students performances in science in OECD countries. Estimated coefficients of "Peer Escs average". Source: elaborations on PISA 2006 data.

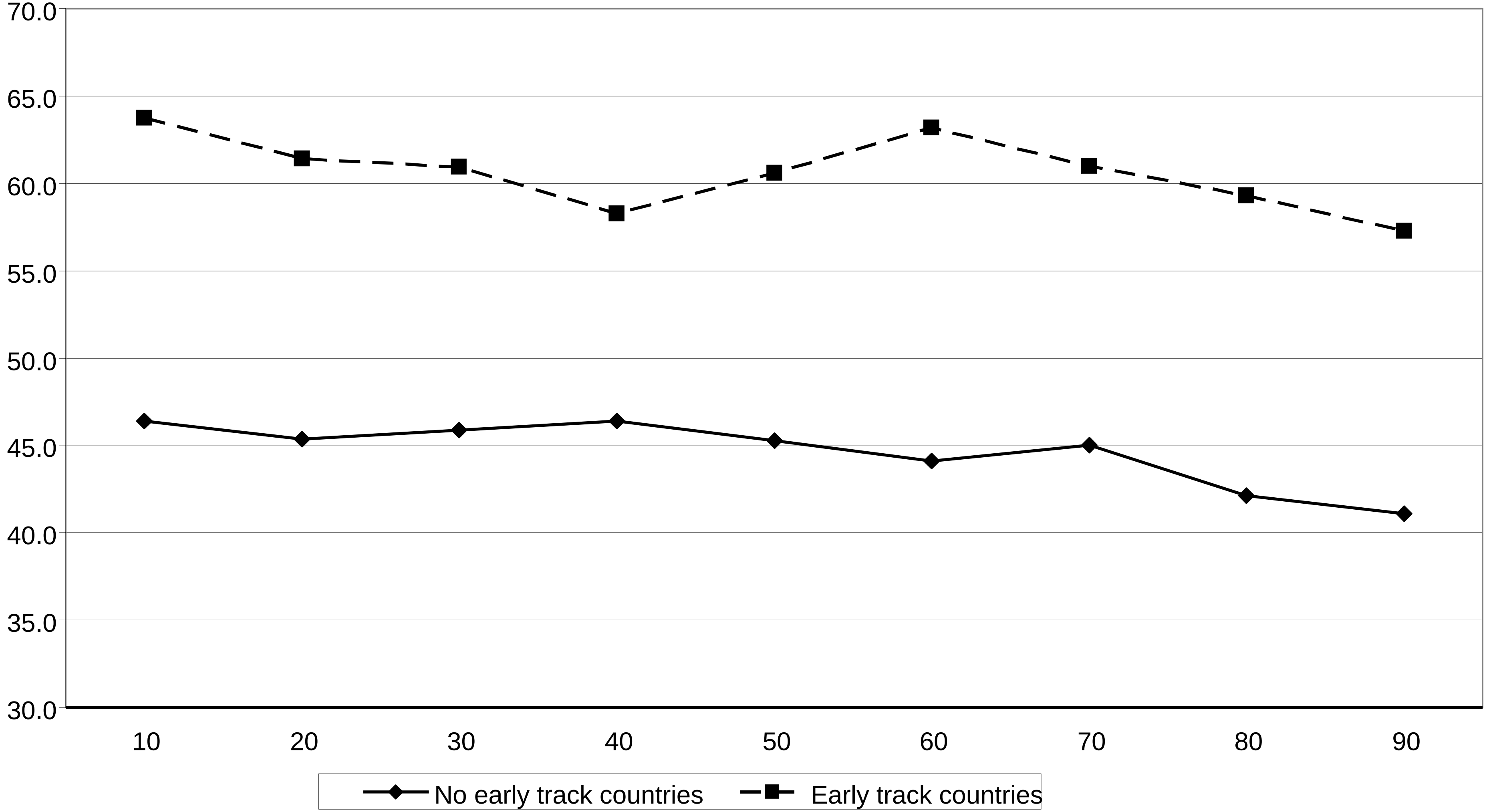
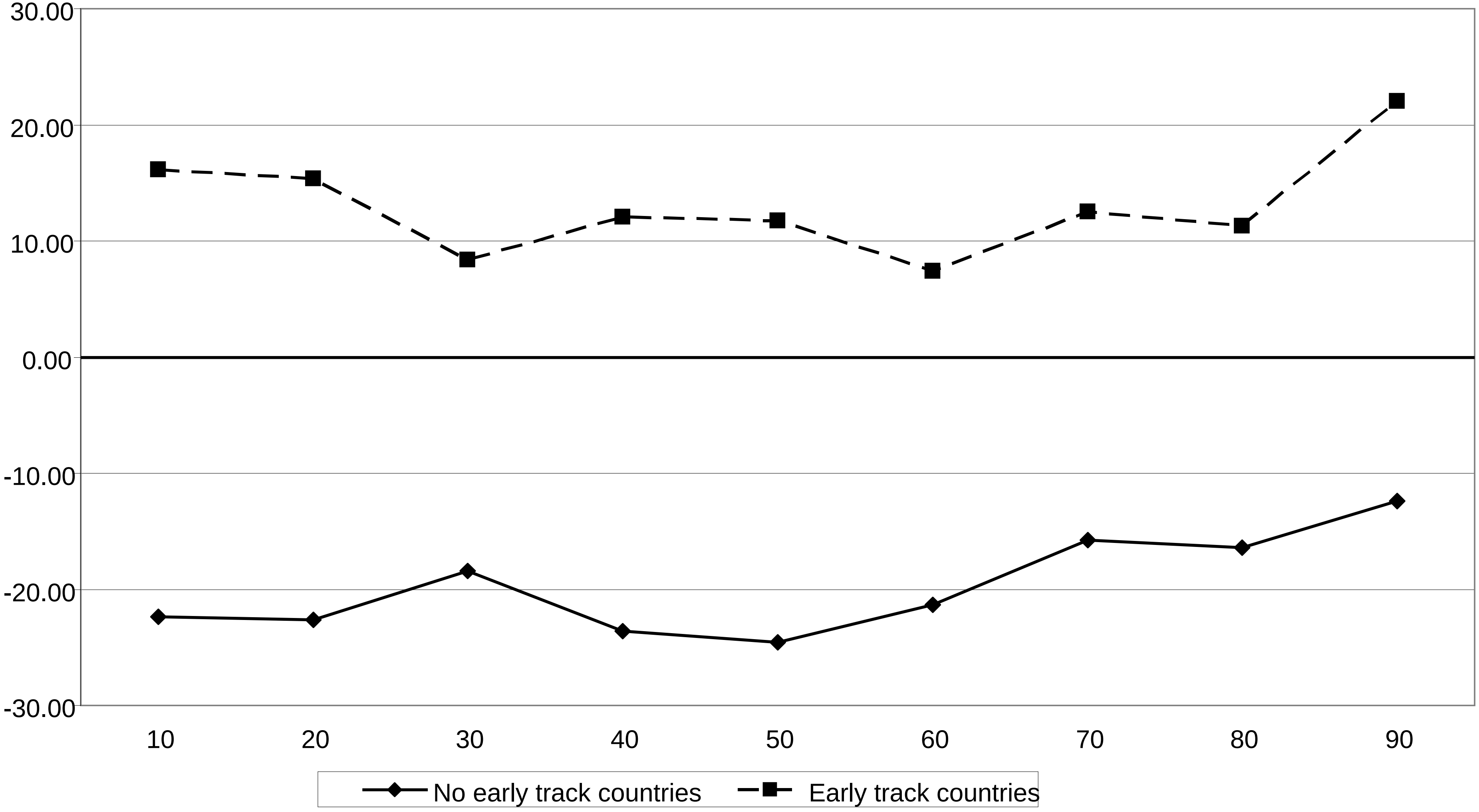


Fig. 5: Quantile regressions on students performances in science in OECD countries. Estimated coefficients of "Escs standard deviation". Source: elaborations on PISA 2006 data.



Tab. A1: Control variables used in regressions

School level controls	
School 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)
School 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)
School 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 showing the single components of school autonomy and responsibility about resource and curricula (the modalities of the scq11 school 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
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 programme.
Escs average	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
Escs variance	Square of the escs standard deviation
N.B. in regressions at student level these variables (apart from escs standard deviation and variance) are considered net of the individual responses.	
<u>Country level controls</u>	
Gdp per capita	
Spending in education per capita	
Age of first track	
Early_track	Dummy variable: 1 if school track occurs before age 13, 0 otherwise
Early_track2	Dummy variable: 1 if the decision about which school track to attend happens before grade 7, 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
<u>Student level controls</u>	
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 programme, 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
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"

Student Additional Controls	
Time science at school	Two dummies about the number of hours spent studying science at school per week: 2-4 and more than 4 (less than 2 is the omitted modality)
Time science at home	Two dummies about the number of hours spent studying science at home per week: 2-4 and more than 4 (less than 2 is the omitted modality)
Type of out-of-school-time lessons	Two dummies about the number of hours spent studying science in out of school lessons per week: 2-4 and more than 4 (less than 2 is the omitted modality)
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