SCHOOL COMPOSITION EFFECTS IN SPAIN

Antonio Di Paolo

Universitat Autònoma de Barcelona & IEB

Abstract:
Drawing on the PISA 2006 data set, this study examines the impact of school's socio-economic composition on the science test scores of Spanish students enrolled in compulsory secondary education. We define school's composition in terms of the average parental human capital of students at that same school. These contextual peer effects are estimated using a semi-parametric methodology, which enables spillovers to affect all the parameters in the educational production function. We also deal with the potential problem of student self-selection into specific schools by using an artificial sorting mechanism, which we believe to be independent of a student's unobserved abilities. The results indicate that the association between a school's socio-economic composition and test score results is clearly positive and significantly higher when computed using a semi-parametric approach. However, we find that the endogenous sorting of students into schools also plays a fundamental role, given that spillovers are significantly reduced when this selection process is eliminated from our measure of school's composition effects. Specifically, the estimations suggest that contextual peer effects are moderately positive only in those schools where the socio-economic composition is comparatively high. In addition, we find some evidence of asymmetry as to how the external effects and the sorting process actually operate, apparently affecting males and females and high and low performance students differently.
Keywords: Educational Attainments, Peer Effects, PISA, Spain
JEL Classifications: I20, I21, I29

Corresponding author: Antonio Di Paolo; e-mail: antonio.dipaolo@uab.cat, tel.: +34935813415; fax: +34935812292. Departament d'Economia Aplicada, Universitat Autònoma de Barcelona (UAB); Campus de Bellaterra, Edifici B 08193 Bellaterra (Cerdanyola), Spain. Institut d'Economia de Barcelona, Universitat de Barcelona.

1. Introduction

This paper investigates school composition effects on Spain's lower-secondary schools (*Educación Secundaria Obligatoria, ESO*), using PISA data from 2006, and as such is the first study to examine this question explicitly in the Spanish case. Quantifying the impact of the socio-economic composition of Spain's schools is especially important in this country because, where there is an excess of demand — as is common in large cities — the admission process in public and public-funded (*concertada*) private schools is very closely related to zoning laws and school district policies. Therefore, given that school admission criteria assign greatest weight to the proximity of the student's home to the school, these educational policies are inextricably linked to the effects of the schools' socio-economic mix. In fact, they may result in the direct transfer of the existing socio-economic residential segregation into the schools (Hoxby 2000, Gorard et al. 2003). Moreover, school composition effects might have gained additional relevance in Spain as a result of the significant increase in the number of immigrant students from less affluent social backgrounds in recent years and the subsequent interaction with existing zoning laws — i.e. less advantaged immigrant families tend to reside within ethnic enclaves and, as a consequence, their children inevitably tend to concentrate in schools characterised by a low socio-economic composition.

Empirically, there are many channels via which the features of an individual's schoolmates or classmates — namely, the peer effects — might influence individual attainment. In the general framework proposed by Manski (1993, 2000), the overall effect of the peer group on individual outcomes primarily involves elements of social interaction that include both endogenous and contextual (or exogenous) effects. The former are the direct effects that peer behaviour or outcomes can have on individual outcomes; that is, students may well learn more because their school/classmates learn more. The latter are the impact that certain exogenous characteristics of the peer group can have on a student's achievement — i.e. individual performance depends on the socio-economic composition of his/her group. In addition, the extent of peer effects might be confounded by the presence of shared environmental/school elements or individual characteristics (e.g. cognitive and non-cognitive skills) that go unobserved by the econometrician; the so-called correlated effects.

Obtaining separate estimates of endogenous and contextual effects is fraught with empirical complications[1] and, moreover, is highly data-demanding. Thus, this study concerns itself solely with contextual effects, which has been a fairly common approach in the empirical literature to date. More specifically, this paper uses a broad measure of the socio-economic composition of schools, based on the average parental educational background (defined as the highest educational level completed by either one of the two parents) for each school. Its main contribution to the existing literature consists in the implementation of a semi-parametric methodology that allows school-contextual effects to influence all parameters in the educational production function (as such, adapting the original proposal made by Raymond & Roig 2010). Intuitively, taking as a reference the most disadvantaged schools in terms of their socio-economic composition (i.e. schools in the lowest quintile of average parental background), the paper shows that moving to better endowed schools might generate a level shift as well as various potential gradient shifts. Indeed, the measure of contextual peer effects proposed here should capture the global impact of school composition on the educational production function. In addition, the flexible strategy adopted ensures that such school composition effects are non-linear, since they are separately computed for each successive quintile of the schools' average parental education.

Finally, the paper deals with the most common problem encountered in such studies, namely the self-selection of students into different schools (i.e. a specific type of correlated

[1] Specifically, what is commonly referred to as reflection problems, which involve the simultaneous determination of achievement for all students within a peer group (i.e. a simultaneity bias problem).

effect). Specifically, the presence of a sorting mechanism that allocates those students that are better endowed of unobserved characteristics into schools with higher average parental schooling might bias our measure of peer effects. Therefore, an alternative sorting mechanism is provided that can be assumed to be unrelated to an individual student's unobserved characteristics. Such reordering is based on the predicted linear score, obtained from an ordered probit model that estimates the probability of membership in each quintile of the schools' average parental education. This artificial sorting is then used to reduce selection bias in the definition of reference and non-reference groups. Thus, the study is able to provide a measure of a school's composition effects that is significantly less affected by correlated effects.

With these purposes in mind, the rest of the paper is organized as follows: section 2 contains a brief review of selected papers examining peer effects, focusing on the various estimation strategies adopted to eliminate correlated effects. Section 3 describes the empirical methodology that is used in this study and Section 4 is dedicated to a description of the data. Section 5 contains the empirical results, as well as a robustness check and an analysis of the potential asymmetries of school composition effects. Section 6 concludes.

2. Selected Contributions

Previous studies of peer effects on scholastic achievement present quite mixed findings and, to date, there is no unified evidence as to the existence or to the actual form that these effects might take. This line of research has sought to capture these potential spillovers at several points in the educational process (from primary to tertiary stages), and by considering different peer features (actual or lagged peers' test scores, ethnic and socio-economic composition of the peer group, etc.). This points to the fact that the resulting spillovers will be either positive or negative (or even zero), while dependent at all times on the nature of the peer variables, and as such the final net effects become an empirical question. Furthermore, governed primarily by data availability, the definition of these peer groups has been markedly heterogeneous, ranging from school, school-by-grade and classroom to other social peers such as roommates or friends. As a result, the findings tend to be highly case-specific and not always strictly comparable. In general, this lack of explicit comparability is attributable to *i*) the specific characteristics of the sample used, and *ii*) the (subsequent) econometric technique adopted in identifying peer effects other than the correlated effects.

Interestingly, some studies are based on special samples in which students are assigned randomly into peer groups, thereby possibly eliminating the bias attributable to correlated effects. More specifically, such quasi-experimental studies exploit the randomized trials generated by the chance matching of students with first-year roommates in college accommodation (see Sacerdote 2001, Zimmerman 2003, Foster 2006 and Brunello et al. 2010 among others), or class assignment on the basis of surname during first year university courses (De Paola & Scoppa 2010).

Several other papers, which focus principally on primary and secondary schools, adopt fixed effect frameworks in order to control for any potential bias in peer effect estimates. For example, McEwan (2003) controls for both school and family fixed effects when estimating peer contextual effects at the classroom level, finding a positive and slightly concave effect of the classroom mean of the mothers' education. Hanushek et al. (2003) exploits a unique panel dataset covering three successive cohorts of students; hence, they are able to control for individual, school and school-by-grade fixed effects in a value-added specification of the educational production function. They report a positive effect of mean peer achievement on improvements recorded in test scores, which remains almost constant over the test score distribution. They also found no important effect of the average family income of the peers and test score heterogeneity in the peer group. Lavy et al. (2008) exploit cohort-to-cohort and within-school changes in the proportion of low achievers (i.e. their proxy of peer ability) to identify endogenous peer effects and the mechanisms via which they impact on an individual's own achievement. They report a clear negative impact of the proportion of low achievers in the classroom, which tends to be more pronounced for students of low socio-economic backgrounds[2]. Ammermuller & Pischke (2009) consider the contextual effects in primary schools for several European countries (using PIRLS data). They assume that contextual peer effects at the classroom level are captured by the average number of books at home. These peer effects are identified by exploiting variations across the classrooms within the same grade for the same cohort of students (once established that these classes had been formed in what was a largely random manner). Their results indicate that, in general, contextual peer effects do exist; however, they also point out that simple OLS estimations might be equally affected by selection bias as well as by measurement error in the peer variable, which tend to operate in opposite directions.

The present study is most closely related to those undertaken by Fertig (2003), Schneeweis & Winter-Ebmer (2007) and Rangvid (2007), which also draw on PISA data. Specifically, Fertig (2003) investigates the effect of reading achievement heterogeneity in US schools, which is identified through Instrumental Variables (IV)[3] — namely, dummies for private and selective schools and a set of variables related to the prevalence of parental caring behaviour in each school. His results indicate that attending a heterogeneous school in terms of student achievement undermines individual performance; however, the negative effect he reports

[2] Moreover, their results also suggest that the negative impact of the proportion of low achievers mainly operates via the disruptive influence it has on teachers' pedagogical practices, interaction with other students and classroom disorder.

[3] Other papers in which the identification of peer effects relies on IV strategies include those by Feinstein & Symons (1999) and Robertson & Symons (2003), where the instruments consist of location variables and teacher assessment of a student's previous ability combined with region of birth dummies, respectively.

appears to be excessive when estimated using IV (which raises the question about the validity of the instruments used).

The paper by Schneeweis & Winter-Ebmer (2007) explores the effect of socio-economic composition at the school-by-grade level in Austria. The authors present evidence obtained, on the one hand, from OLS estimations based on an extensive set of individual and school controls and, on the other hand, from the application of school fixed effects. They argue that, when accounting for school type — given the marked track system in Austrian lower and upper secondary schools — school fixed effects reduce the selection bias in the estimation of peer effects. Their results highlight a significant asymmetry in the peer effects on reading[4], which seems to have a more beneficial effect in the case of students of a low socio-economic background. Moreover, they also adopt a quantile regression strategy, which reveals that students in the lower part of the ability distribution are more positively affected by the socio-economic composition of their peer group.

Finally, Rangvid (2007) analyses the effect of the socio-economic composition of a school in terms of the three PISA subjects (reading, maths and science) drawing on Danish data, which are complemented with administrative registers to overcome the potential problems caused by the limited sample of students within each school[5]. Given the comprehensive nature of the Danish secondary school education system, the author cannot rely on the school-fixed effect estimation as was the case in Schneeweis & Winter-Ebmer (2007); indeed, she cannot assume that individuals (and their families) who are placed in a given school of a certain track share similar unobserved characteristics[6]. Her identification strategy is instead based on controlling for a large set of individual, family and school variables, without explicitly considering the role of selection on unobservable features. The results in this study suggest a clear positive effect of attending a school with a higher socio-economic composition in the middle of the test score distribution, whereas no significant effect is found for the socio-economic heterogeneity at the school level. Moreover, the quantile estimation reveals that school composition effects tend to be higher for low-ability students on the reading test score, but the author finds a U-shaped effect for science, which means that low and high ability students benefit equally from a better socio-economic school composition.

[4] By contrast, they also suggest that the apparent peer effects in mathematics, as estimated by OLS, are due only to selection effects, given that their fixed-effect estimates are not statistically significant. Additionally, in this case, peer group heterogeneity seems to play a very limited role in explaining test score attainment.

[5] As argued by Micklewright et al. (2010), the limited student sampling made by PISA can result in a measurement error in the estimation of peer effects. This would bias the effect of school composition towards zero. Unfortunately, such administrative data are not available for public use in the Spanish case; therefore, it should be borne in mind that the estimates reported in this study represent a lower boundary of the true impact of school-average parental education.

[6] Notice that since the LOGSE reform of 1990, the Spanish secondary education system has been compulsory and comprehensive until the age of sixteen, which (as in the case of Denmark) makes the school-fixed effect framework unfeasible for controlling endogenous peer group selection. See section 3 for details as to how such a problem is addressed in this paper.

3. Empirical Framework

The estimation strategy proposed in this paper represents a step forward in terms of the measurement of peer effects. Indeed, the main innovation with respect to previous studies consists, as briefly commented in the introduction, in the idea that the spillovers produced by an improvement in a school's socio-economic composition may affect not only the intercept, but all the parameters of the educational production function. This original proposal has been taken (and adapted) from the paper by Raymond & Roig (2010), in which they estimate the externality produced by the average human capital of workers in the same firm. In keeping with this externality, this paper takes as its starting point the standard educational production function,

$$T_{i,s} = \alpha_s + \beta' X_i + \delta' Z_s + \varepsilon_{i,s} \qquad (1)$$

where test score $T_{i,s}$ of student i in school s depends on a set of individual and family characteristics (X_i) as well as on a set of school characteristics (Z_s), plus a composite error term ($\varepsilon_{i,s}$). Usually, exogenous peer effects are simply estimated by considering that the intercept term (α_s) is not fixed, but instead dependent on an average characteristic of the peer group — i.e. in this case, the average parental education of students at that school s ($\overline{PE^s}$). This means that the intercept term in (1) can be rewritten as,

$$\alpha_s = \alpha + \mu \cdot \left(\overline{PE^s}\right) \qquad (1a)$$

which indicates that a unit increase in the average parental education in the school modifies the mean test score by μ points, through a shift in the intercept term. We could also adopt a non-linear specification, where the impact of the school's composition of parental human capital is allowed to vary for each successive quintile of school-average parental education ($Q_j\left(\overline{PE^s}\right), j = 1,..,5$). In this case, the intercept term in (1) can be expressed as,

$$\alpha_s = \alpha_1 + \sum_{j=2}^{5} \alpha_j \cdot Q_j\left(\overline{PE^s}\right) \qquad (1b)$$

where the contextual peer effects are now α_j ($j=1,..,5$) and are allowed to be different for each quintile of average parental schooling. Even in this case, the impact of the peer group's characteristics is only produced by a level effect, which operate through a modification of the educational production function intercept; in fact, once the expression (1b) is substituted into equation (1) we obtain,

$$T_{i,s} = \alpha_1 + \sum_{j=2}^{5} \alpha_j \cdot Q_j\left(\overline{PE^s}\right) + \beta' X_i + \delta' Z_s + \varepsilon_{i,s} \,. \qquad (2)$$

This corresponds to the standard equation used in the peer effects literature, except for the non-linear specification of the contextual peer effects.

Equation (2) clearly specifies that the standard approach constrains school composition spillovers so as to affect only the intercept term and no other parameter in the educational production function (even allowing for a non-linear effect). However, there is no theoretical reason to believe that the contextual peer effects consist only of a simple level effect. For example, an improvement in the socio-economic composition of the peer group might modify the gradient of the effect of a student's family background and home environment on his/her test score. Additionally, belonging to a "good" peer group in terms of average parental human capital might relax the relationship between other school characteristics and an individual's achievements.

In order to capture any potential shape effect of school composition, we consider a reference group, which consists of all the students who belong to the least-advantaged schools in terms of average parental educational background. In the present application, the least-advantaged schools are defined as those schools that appear in the first quintile of the average parental education[7] (i.e. $Q_j\left(\overline{PE^s}\right) = Q_1\left(\overline{PE^s}\right)$). Therefore, the educational production function is separately estimated for the reference category, as in equation (3):

$$\left(T_{i,s} \mid Q_1\left(\overline{PE^s}\right)\right) = \hat{\alpha}_1 + {}^{Q_1}\hat{\beta}' X_i + {}^{Q_1}\hat{\delta}' Z_s + {}^{Q_1}\hat{\varepsilon}_{i,s} = {}^{Q_1}\hat{\psi}' R_{i,s} + {}^{Q_1}\hat{\varepsilon}_{i,s} \quad \text{if } Q_j\left(\overline{PE^s}\right) = Q_1\left(\overline{PE^s}\right)$$

. (3)

From the obtained parameter estimates (${}^{Q_1}\hat{\psi}$), we then proceed to forecast the test score for all the individuals who do not belong to the reference group, that is,

$$\left(\hat{T}_{i,s} \mid Q_j\left(\overline{PE^s}\right); {}^{Q_1}\hat{\psi}\right) = \hat{\alpha}_1 + {}^{Q_1}\hat{\beta}' X_i + {}^{Q_1}\hat{\delta}' Z_s \qquad \forall i \in Q_j\left(\overline{PE^s}\right), j > 1. \qquad (4)$$

Finally, for each successive quintile of the school-average parental education, the measure of school composition spillovers presented here consists of the average difference between the actual and the forecasted test score within each quintile:

[7] As noted by Raymond & Roig (2010), the definition of the reference group is always subject to some degree of arbitrariness; in their case, they define the reference group as those productive establishments in which the average workers' human capital is equal to or less than eight years of schooling. This definition follows the logic that eight years of education corresponds to the compulsory length of education under the institutional framework that was then valid for individuals in their sample; moreover, it should represent those firms that chiefly employ unskilled workers. In our case, we consider it better to define the reference group in an endogenous way — i.e. dividing the sample into quintiles and taking the first one as the reference group. This definition allows us *i*) to consider schools as being more heterogeneous units than firms, and *ii*) to maintain a sufficient number of observations in the reference and non-reference groups.

$$IEX_j = \frac{\sum_{i=1}^{N_j} T_{i,s} - \left(\hat{T}_{i,s} \mid Q_j\left(\overline{PE^s}\right); {}^{Q_1}\hat{\psi}\right)}{N_j} \qquad \forall i \in Q_j\left(\overline{PE^s}\right), j > 1 \tag{5}$$

In other words, this measure of contextual peer effects consists of counterfactual evidence, which is based on the *ceteris paribus* within-quintile mean differential between the observed and the predicted test score, where the latter is obtained by using the parameters estimated for students in the least-advantaged schools. More intuitively, this methodology represents a semi-parametric approach to capture the *ceteris paribus* change in the test score, produced by moving a representative student from the first quintile to successive quintiles of the school-average parental education. Note that this measure of the effects of school composition captures in a semi-parametric way the change in each parameter making up the whole educational production function (both level and shape effects), produced by incrementing the average parental schooling from the first to the higher quintiles. In this way, we are able to provide more compelling and complete evidence about the effect of school composition on individual test scores, obtained without constraining these potential spillovers of the peer group's socio-economic status to operate only through a shift in the intercept term.

3.1 School Composition and Selection Bias

This semi-parametric methodology is not, however, exempt from the most relevant empirical problem in the estimation of contextual peer effects, represented by the self-selection of students into schools and peer groups. In this paper we seek to reduce the bias produced by the sorting mechanism that allocates those students with a greater (lesser) endowment of unobserved abilities into better (worse) peer groups, which may bias our measure of school composition effects. Indeed, were this to be the case, the test score forecast for non-reference group students from eq. (3) would present a downward bias, pointing to an overestimation of the effects of school composition. In other words, even if we tried to account for selection on observable variables by conditioning for a large set of individual and school controls (similar to Rangvid 2007, see section 4), we would not be particularly confident about the conditional zero mean of the error term in the test score equation estimated for the reference group (eq. 3).

In line once more with Raymond & Roig (2010), rather than using a classification of reference and non-reference groups based on actual school-average parental education, students were allocated to reference and non-reference groups on the basis of their predicted linear score obtained from an ordered probit model, which estimates the probability of membership in each of the five quintiles of the average parental education at the school level. Specifically, we computed the predicted linear score that represents a proxy of the (latent) parental human capital in each school, obtained from the following equation:

$$\overline{PE_i^s}^* = \gamma' W_i + \mu_i \Rightarrow \widehat{\overline{PE_i^s}}^* = \tilde{\gamma}' W_i \qquad (6)$$

The explanatory variables that are specifically included in the vector (W_i) in eq. (6) comprise a set of dummies for school availability (one, and more than one, school available) and the student's age on arrival in Spain (for immigrants), as well as region and municipality size — control variables that also appear in the test score equation (to capture unobserved school characteristics that are common within region and municipalities of similar dimensions). Subsequently, the observations are sorted according to the quintiles of the predicted linear score ($\tilde{\gamma} W_i$); this proxy of the schoolmates' parental human capital would be correlated to the school-average parental education, but at the same time it can be considered as independent of a student's unobserved abilities. Therefore, we take as our reference group those students in the first quintile of the predicted parental schooling and we estimate the test score equation for them as

$$\left(T_{i,s} \mid Q_1^*\left(\tilde{\gamma}' W\right)\right) = \widehat{\alpha}_1 + {}^{Q_1^*}\widehat{\beta}' X_i + {}^{Q_1^*}\widehat{\delta}' Z_s + {}^{Q_1^*}\widehat{\varepsilon}_{i,s} = {}^{Q_1^*}\widehat{\psi}' R_{i,s} + {}^{Q_1^*}\widehat{\varepsilon}_{i,s} \quad \text{if } Q_j^*\left(\tilde{\gamma}' W\right) = Q_1^*\left(\tilde{\gamma}' W\right)$$

. (7)

It is then possible to re-compute the index of school composition spillovers in the same fashion as above, but now without such a marked effect of the self-selection of students into peer groups:

$$IEX_j^* = \frac{\sum_{i=1}^{N_j^*} T_{i,s} - \left(\tilde{T}_{i,s} \mid Q_j^*\left(\tilde{\gamma}' W_i\right); {}^{Q_1^*}\widehat{\psi}\right)}{N_j^*} \qquad \forall i \in Q_j^*\left(\tilde{\gamma}' W_i\right) = Q_j\left(\widehat{\overline{PE_i^s}}^*\right), j > 1. \qquad (8)$$

Similarly to the IV estimation, we exploit the between student variability of school availability and, in the case of immigrants, of arrival age, within municipalities of the same dimension within the same region. Again, in line with IV, a valid exclusion restriction is needed to rule out endogenous student sorting. We consider that once controlling for parental education, socio-economic status and many other family characteristics in the test score equation (see the next section for details), we can reasonably assume that the only channel through which school availability and age on arrival might affect a student's test score is via the effect of school selection (i.e. they are independent of unobserved student characteristics). If this is true, eq. (7) is correctly estimated[8] and the measure of socio-economic school composition obtained from (8) is now "clean" thanks to the potential endogenous selection of students into schools.

[8] Notice that the composite error term in eq. (1) may assume the general form $\varepsilon_{i,s} = \eta_i + \nu_s + \varsigma_{i,s}$, which means that apart from individual unobserved ability (η_i), unobserved school characteristics (ν_s) may also cause some bias in the results. However, we are not able to deal explicitly with this problem using the PISA database. We are, therefore, forced to assume that the correlated school effects are zero once conditioned by a school's characteristics, at least in the case of the reference group. It should be borne in mind that, should this assumption prove invalid, the results presented in what follows may still be affected by the presence of some unobserved correlated school effect.

Nevertheless, we recognize that the choice of the exclusion restrictions (school availability and age on arrival) is not free of criticism. It is quite obvious that both variables might have an effect on the probability of being in a given quintile of the school-average parental education. What is not so immediately obvious is the belief that, having controlled for a large set of family characteristics, these variables are completely orthogonal to a student's unobserved characteristics. In order to ensure a greater degree of reliability for our results, in sub-section 5.1 we provide an intuitive falsification test for the validity of the exclusion restrictions used here, which is basically aimed at showing that these variables are not likely to contribute to school composition spillovers (but they do explain the likelihood of membership in reference and non-reference groups).

4. Data Description

As discussed above, the empirical analysis is based on Spanish data from the 2006 Program for International Student Assessment (PISA), undertaken by the OECD (see OECD 2009 for details). PISA focuses on the acquisition of skills in reading, mathematics and science among a target population of students aged 15 to 16. The 2006 assessment was specifically concerned with the testing of science skills and as such is the only skill considered in this study[9]. In the specific case of Spain, the students interviewed were drawn from a cohort of individuals born in 1990 and enrolled in lower-secondary schools (*Educación Secundaria Obligatoria, ESO*) during the survey year. As outlined earlier, Spanish lower-secondary education is completely comprehensive and compulsory until the age of 16. Normally, 15-year-old pupils will be enrolled in the 4th grade of lower-secondary education; however, the sample contains students from lower grades as well (3rd, 2nd and 1st grade), representing those who have repeated one or more grades. The original Spanish sample comprised 19,604 students enrolled at 686 different schools.

The PISA survey has several statistical peculiarities that must be taken into account in the estimation phase. First of all, the skills assessment was carried out using five Plausible Values for each field, which are then normalized to obtain a global average of 500 and a standard deviation of 100. This technique, derived from Item Response Theory, allows students' (latent) skills to be represented consistently when the number of submitted items is too small to represent true individual ability. Moreover, the structure of the final sample must also be taken into account, given that it is the product of a complex two-stage stratification procedure used to ensure that the entire population is represented. Specifically, the first step consists in the

[9] Despite this, the 2006 survey also contains information about reading and mathematics skills. Attention is limited here to the science domain for reasons of space. The results for the other two skills are qualitatively similar, and are available upon request from the author.

stratified selection of schools with 15- to 16-year-olds enrolled in their classes, with sampling probabilities that are proportional to the number of eligible students enrolled; in the second step, a given number of students are randomly selected within each sampled school (up to 35). In order to take into account the specific statistical properties of the PISA sample, all the statistics and estimations that we present in this study have been carried out with the STATA routine “pv”, specifically designed for PISA and similar surveys (Macdonald 2008, Lauzon 2004).

The PISA survey contains, apart from the plausible values of the test score, an extensive (but often not exhaustive) battery of questions about a student’s and his/her family’s characteristics, as well as several other school characteristics. The empirical analysis has been conditioned to the information drawn from a large subset of relevant questions so as to limit the role of the unobservable variables (following Rangvid 2007 and the OLS specification of Schneeweis & Winter-Ebmer 2007, given the available variables). The whole set of control variables are reported in the complete version of this work (see Di Paolo 2010), together with the exact definition of each variable, its mean and standard deviation. In summary, the conditioning variables can be divided into individual controls (sex, grade attended, age, migration status and the language spoken at home), family controls (paternal and maternal education, family socio-economic status, maternal working situation, number of books at home and educational resources), school controls (prevalence of immigrants, girls and part-time teachers, lack of qualified teachers, school autonomy, student/teacher ratio, school size, school ownership, streaming processes, career guidance employee and presence of computers for instruction) and territorial controls (municipality size and region)[10].

As usual, we also generated indicator functions for observations with missing information for the explanatory variables, in order to control for the non-randomness of the missing values; in the case of missing information, the explanatory variables are fixed as being equal to zero. As a measure of the socio-economic composition of the school we consider the school-average parental education, taking the highest educational level completed by one or other of the two parents[11]. Observations with missing information about the highest parental education have been discarded from the sample (2% of the total sample). Since our school composition measure consists in the school-average value, we also discarded the forty-two observations of students

[10] Notice that we also retain information about the availability of neighbouring schools within the same area and about the students’ age on arrival in Spain (for first generation immigrants). These are included in eq. (6) only.

[11] The final student weight provided in the PISA database has been used in the computation of the school composition variable. This should reduce the imprecision in the school composition measure obtained from PISA data, where (as commented above) not all the students from every school are sampled. Whatever the case, the results are insensitive to the exclusion of the final student weight in the computation of the school-average parental education. Notice also that the mean peer characteristic is usually computed without the contribution of the individual (because this might cause a reflection problem when the average value of the peer group is used as an explanatory variable). In this case, where school-average parental education is only used to define reference and non-reference groups, this complication is not necessary; in any case, the results are virtually unchanged when the average parental education does not include the individual’s contribution (the results are available upon request). See Table 1A in the Appendix for more details about the school composition variable.

that are enrolled in schools with fewer than eight students. In the end, the sample used in the empirical analysis was formed by 19,164 students at 675 different schools.

5. Results

5.1 Level Effects of School Composition

The starting point for this empirical analysis was an estimation of the educational production function as described by eq. (2), in which the school composition measure was allowed to be non linear (dummies for school-average parental education quintiles), but constrained so as to produce only a level effect. As reported in Table 2 in the complete version of the paper (Di Paolo 2010), the results indicate that moving from the first quintile to the second quintile of average parental schooling at the school level had only a slightly significant impact (7 points) on the science test score. However, the *ceteris paribus* comparison between students in the first quintile and those in the third revealed that students in the latter group performed significantly better than the reference group, showing a positive score gap of about 14 points. This positive level effect of school composition fells somewhat when moving to the fourth quintile (11 points). Finally, the test score for students in the most-advantaged group in terms of school composition (fifth quintile) was, on average, 24 points higher than the score for students in the least-advantaged group. This means that an improvement in the school's socio-economic composition had a substantial level effect on individual test scores, and that this appears to be non-linear in the quintiles of average parental human capital at the school level.

The estimates for the remaining control variables are of independent interest, and it is worth briefly commenting on the main findings. The increase in student age was positively associated with the test score, whereas females seemed to obtain worse results than males in science. The effect of the grade attended was as expected, given that students from lower grades than that of the fourth grade (the standard grade at ages 15 and 16) performed significantly worse. Even accounting for the language spoken at home and other family characteristics, first-generation immigrant students performed significantly worse than natives and second-generation immigrants (negative gap of 25 points). An improvement in a family's socio-economic status had a marked positive effect on the science test score, while only maternal education showed a significant and positive effect on a student's competence for the sciences. Children of working mothers performed markedly better than those whose mother did not work, with a *ceteris paribus* average increase of 10 points in the test score. In addition, the number of books and a home's endowment of educational resources also had a significant positive effect on the test score.

An analysis of school control variables revealed the usual results for PISA data — i.e. school characteristic control variables were hardly significant when explaining students' test scores. Therefore, we shall only describe in brief the few variables that displayed statistically significant coefficients. We detected a positive effect of the percentage of girls attending a school, whereas the increase in the ratio of personal computers for instruction to school size had a negative impact on the science test score. After accounting for family characteristics, a school's socio-economic composition and other school characteristics, it was found that public schools performed significantly better than private and public-funded private schools. Finally, students enrolled at schools that can hire teachers autonomously seemed to achieve better results than their counterparts. The evidence obtained from the territorial control variables indicated that being schooled in a large city has a positive effect on science attainment; moreover, the coefficient associated with regional dummies (not shown here) suggested that Catalonia and the Basque Country performed significantly worse than the rest of Spain's regions.

5.2 Accounting for Shape Effects and for Selection Bias

The results obtained from the estimation of eq. (2) suggest a significant and positive effect of the school's socio-economic composition. However, as previously highlighted, this result may merely represent partial or incomplete evidence, given that we implicitly constrained the impact of the school-average parental education so as to affect only average attainments (i.e. the intercept of the educational production function). In order to capture any other potential slope effect produced by an improvement in the school endowment of parental human capital, we implemented the innovative methodology described above in section 3[12].

Panel A of Table 3 contains the estimated value of our measure of school composition spillovers (eq. 5). We computed IEX_j separately for each quintile of the school-average parental education and also calculated the mean value for all the quintiles (except that of the first, which is the reference category). The results from the semi-parametric methodology confirmed that the effect of the parental education of the peer group was substantial and clearly non-linear. As in the previous case, moving from the least-advantaged group to the second quintile of the school's socio-economic composition had almost no effect on individual test scores (almost 5 points, but not statistically different from zero), whereas the step to the third quintile produced a positive increase of about 12 points. However, the movement to higher quintiles generated substantial (and positive) slope effects, which were hidden by the implicit constrains of eq. (2). Indeed, school composition effects could be quantified into 26 additional test score points for students

[12] The estimates of the educational production functions for the reference group (eq. 3 and eq. 7) are not reported here for reasons of space, but are available upon request; in general, the results are conventional and qualitatively similar to those reported in Table 2.

in the fourth quintile of the average parental schooling and up to 71 points for students in the highest quintile. Additionally, the mean value for all the non-reference groups was also statistically significant, approaching 28 test score points.

[TABLE 3 ABOUT HERE]

However, these results may well be biased by the fact that students with a better endowment of unobserved abilities are more likely to enrol in the better schools (in terms, that is, of their socio-economic composition). In order to reduce this potential selection bias, we first estimated eq. (6) using an ordered probit model, the dependent variable of which was the five quintiles of the actual school-average parental education. The estimates (see Table 2A in the Appendix) indicate that immigrant pupils who arrived in Spain at an earlier date are significantly more likely to be enrolled in schools where their schoolmates' parental education is higher; moreover, conditional on region and municipality size, the chances of being in better schools is also higher for those who reside closer to other schools. In general, the variables included provide a satisfactory explanation of the probability of being in each of the quintiles of the school-average parental education. Subsequently, we used the predicted linear score to obtain a proxy of the school-average parental human capital that was independent of the students' unobserved characteristics.

When students were sorted into reference and non-reference groups according to the predicted linear score, the evidence concerning school composition spillovers was markedly different. As reported in the lower panel of Table 3, the mean effect for all the non-reference groups was statistically non existent, which is the result of a clear convexity of school composition effects with respect to the different quintiles of $\overline{PE^s}$. In fact, students from the second quintile of the proxied average parental education were penalized by about 10 points with respect to pupils at the least-advantaged schools (i.e. the reference group), and the spillover effects for students in the third quintile were not statistically different from zero. In addition, when the selective sorting of pupils into schools was accounted for, the effect of school composition was strongly reduced for students enrolled in the better endowed schools (about 15 test score points for both the fourth and the fifth quintiles). This evidence suggests that, especially for students in the highest quintile of the average parental schooling, there is a considerable sorting process in their favour with respect to less-advantaged students. Summing up, a significant contextual peer effect was still detected, but it seemed to generate a positive and modest spillover only in those schools where the average level of parental education was higher. However, the process of student sorting would seem to be even more important than the externality produced by the socio-economic origins of an individual's schoolmates.

5.3 Robustness Checks

The evidence presented above sought to determine whether a more favourable school composition produced better individual results in the science test score (*ceteris paribus*). The baseline results seem to suggest that the exogenous characteristics of an individual's schoolmates (the contextual peer effect, here defined in terms of parental education) exert a positive externality on the individual acquisition process of competence in this field. This spillover was even higher when we considered not only the level effect, but also the whole slope effects in the educational production function. However, these results are likely to have been confounded by the presence of an endogenous sorting process that allocated the students with a better endowment of unobserved abilities to the better schools (in terms of their socio-economic composition). On attempting to reduce this potential bias, the results are markedly different: there was a small and positive effect of the school's socio-economic composition only in those schools where the average level of parental education was considerably higher.

Whatever the case, these results might still be biased if the variables used as exclusion restrictions had been systematically related to a student's unobserved ability. Recall that the validity of these results is based on the assumption that, having controlled for the father's and mother's education, the family's socio-economic status, migration status, language spoken at home and other family characteristics, the presence of one or more available schools and the arrival age for immigrant children are unrelated to the unobserved abilities. Unfortunately, there is no formal way to prove the validity of this assumption, given that it involves elements that are, by definition, unobservable. Even so, we have provided an intuitive falsification test, which helps us to corroborate our excludability assumption. This test is based on the idea that if the excluded variables had had some effect on the test score equation (even via correlation with the unobservable), including them in the equation for the reference group would have modified the results obtained with our measure of peer effects. In fact, the logic behind the exclusion restrictions is that these variables only affect the test score (for the reference group) through their effect on the probability of being in each quintile of the school-average parental education.

First, we performed several statistical tests to analyse the significance of the variables excluded from the educational production function; the results (not shown here) suggest that both variables (individually and jointly) do not differ from zero at any conventional level of significance. Moreover, we gradually included the dummies for school availability and age on arrival in the test score equations (3) and (7) and, then, we recomputed the measures of school composition effects (5) and (8), without and with the endogenous sorting correction respectively. The results, reported in Table 4, showed *i*) the baseline measure of spillovers to be indistinguishable from the original one computed without the excluded variables; in addition, *ii*) the results were only slightly different (but identical in statistical terms) when students were re-sorted into reference and non-reference groups according to eq. (6) and the two variables were included in the test score equation for the reference group. In principle, if the excluded variables had had an effect on a student's test score and/or had contributed to explain a school's composition effects, we would have observed a marked alteration in the measure proposed in this paper. The evidence that can be drawn from the fact that when school availability and age on arrival are included in the test score equation and no significant changes are observed makes the excludability assumption made in this paper more reliable. Whatever the case, it should be borne in mind that, were another sorting mechanism to be operating — especially with respect to schools' unobservable characteristics — the results could still contain some bias and must be considered with caution.

6. Conclusions

Drawing on PISA 2006 data (primarily the science test score), this paper has investigated the effects of school composition on Spanish secondary schools. A novel methodology has been implemented to measure the spillovers produced by one specific exogenous characteristic of a student's schoolmates, namely we treated the highest level of education completed by the parents as a measure of the school's socio-economic composition. The proposed methodology relaxes the implicit constraint — common to any peer effect study — whereby the contextual element can only affect the average outcome through an intercept shift (i.e. a level effect).

When accounting for all the changes in the educational production function parameter generated by an improvement in the school's socio-economic composition (level and slope effects), it was found that school composition effects are substantial and significantly higher than those obtained with the constrained specification. More specifically, the results indicate that the effect of moving from the least-advantaged schools (those in the first quintile of the school-average parental education) to better endowed schools improves the science test score in a non-linear way, with the positive effect being particularly pronounced for pupils from top schools — i.e. those enrolled in schools where most of the parents had completed upper-secondary or tertiary education (fifth quintile of the school-average parental education).

However, this preliminary evidence should not be understood as being the pure contextual effect of the school's socio-economic composition, given that it might be confounded by the presence of correlated effects. This paper has explicitly attempted to deal with the endogenous selection process whereby students endowed with higher unobserved abilities are allocated to better schools (in terms of their socio-economic composition). This was achieved by re-sorting students according to a predicted linear score obtained from an ordered probit model, which estimates the probability of membership of each quintile of school-average parental education.

It is argued that, by proxying parental human capital, students can be re-sorted in a way that is uncorrelated with unobserved individual characteristics.

When school-composition spillovers were recomputed on the basis of this artificial re-sorting, the evidence was significantly different. The externalities produced by the parental human capital of schoolmates were drastically reduced and they were moderately positive only when the school socio-economic composition was comparatively high (in the fourth and fifth quintiles). Moreover, additional evidence concerning the asymmetries of the effects of school composition (see Di Paolo 2010) revealed major differences between male and female and between high and low performance students. It seems that the results of male students are more closely affected by endogenous sorting than they are by the exogenous characteristics of their schoolmates; by contrast, the results of their female counterparts are more sensitive to the positive contextual effect given that school composition effects were greater when “cleaned” by self-selection. Furthermore, the subgroup of low performance students appears to be positively affected by an improvement in school-average parental background, even after accounting for the presence of endogenous sorting.

Having said this, it is important to bear in mind the potential pitfalls of this study, which are linked primarily to the limitations of the database drawn upon. First, it is quite likely that the reported effects of school composition are a lower bound of the true impact, given that the limited sampling of students in PISA might cause some attenuation bias in our estimate (compare Micklewright et al. 2010). Second, in the case where selection is made on the basis of a school’s unobserved characteristics (i.e., those that are not captured by the extensive list of school controls included here), the measure of school composition effects might still contain some bias. Third, if the variables used as exclusion restrictions are in some way correlated with unobserved student abilities, the methodology adhered to here for reducing the bias generated by endogenous sorting would not be effective at all.

Whatever the case, and even taking these potential limitations into consideration, the evidence presented above makes important contributions to the on-going public debate concerning school laws and the (re)allocation of certain types of student into (other) schools. First, the relevance of endogenous student sorting raises the question as to just how equitable and efficient the zoning laws regulating access to Spain’s secondary schools are. This becomes a matter of urgency when it is seen that, with the self-selection of students ruled out, the positive impact of enhancing a school’s socio-economic composition is only possible when the average parental educational background is comparatively high. This result would seem to suggest that the zoning laws are actually impeding students of a low socio-economic background from benefiting from a more favourable socio-economic school environment, given that they appear to lead to the concentration of such students in disadvantaged school environments. This is because families of lower socio-economic standing tend to locate systematically in certain residential areas and, as such, their children have restricted access to the "better" schools in terms of their socio-economic composition. Although these results do not in themselves justify the suppression of the aforementioned zoning laws, they might be seen as a justification for seeking to compensate less advantaged students.

Secondly, evidence pointing to an asymmetric effect in favour of low performance students might be deemed an argument in favour of the introduction of more flexible school admission policies. Indeed, a more equitable student mix, achieved by reallocating marginal, low-performance students into schools with a higher than average socio-economic composition, might well reduce inequalities in educational achievement and even increase overall results. Whatever the case, a more detailed examination of the relationship between existing school laws, school segregation and the effects of school composition is essential if we are to clarify any of this evidence. These represent interesting questions for future research, particularly if more exhaustive data can be drawn upon.

References

Ammermueller, A. & Pischke, J.S. (2009). "Peer Effects in European Primary Schools: Evidence from the Progress in International Reading Literacy Study". Journal of Labor Economics, University of Chicago Press, vol. 27(3), pages 315-348, 07.

Brunello, G., De Paola, M. & Scoppa, S. (2010). "Peer Effects In Higher Education: Does The Field Of Study Matter?". Economic Inquiry, Western Economic Association International, vol. 48(3), pages 621-634, 07.

Hoxby, C. (2000). "Peer Effects in the Classroom: Learning from Gender and Race Variation". NBER Working Papers 7867, National Bureau of Economic Research, Inc.

De Paola, M. & Scoppa, V. (2010). "Peer group effects on the academic performance of Italian students". Applied Economics, Taylor and Francis Journals, vol. 42(17), pages 2203-2215.

Di Paolo, A. (2010). "School composition effects in Spain". XREAP working paper 2010-13, available at http://ideas.repec.org/p/xrp/wpaper/xreap2010-13.html

Feinstein, L. & Symons, J. (1999). "Attainment in secondary school". Oxford Economic Papers, Oxford University Press, Vol. 51 (2), pages 300-321.

Fertig, M. (2003). "Educational Production, Endogenous Peer Group Formation and Class Composition - Evidence from the PISA 2000 Study". IZA discussion paper No. 714, Institute for the Study of Labor, Bonn.

Foster, G. (2006). "It's not your peers, and it's not your friends: Some progress toward understanding the educational peer effect mechanism". Journal of Public Economics, Elsevier, vol. 90(8-9), pages 1455-1475, September.

Gorard, S., Taylor, C., & Fitz, J. (2003). *Schools, Markets and Choice Policies*. RoutledgeFarmer.

Hanushek, E.A., Markman, J.M., Kain, J.F. & Rivkin, S. (2003). "Does peer ability affect student achievement?". Journal of Applied Econometrics, John Wiley & Sons, Ltd., vol. 18(5), pages 527-544.

Lauzon D. (2004), "Variance estimation with plausible value achievement data: Two STATA programs for use with YITS/PISA data". Information and Technical Bulletin, Statistics Canada, Ottawa.

Lavy, V. & Paserman, M.D. & Schlosser, A. (2008). "Inside the Black of Box of Ability Peer Effects: Evidence from Variation in the Proportion of Low Achievers in the Classroom". NBER Working Papers 14415, National Bureau of Economic Research, Inc.

Macdonald, K. (2008). "PV: Stata module to perform estimation with plausible values". Statistical Software Components S456951, Boston College Department of Economics, revised 05 Sep 2008.

Manski, C.F. (1993). "Identification of Endogenous Social Effects: The Reflection Problem". Review of Economic Studies, Blackwell Publishing, vol. 60(3), pages 531-42, July.

Manski, C.F. (2000). "Economic Analysis of Social Interactions". Journal of Economic Perspectives, American Economic Association, vol. 14(3), pages 115-136, Summer.

McEwan, P.J. (2003). "Peer effects on student achievement: evidence from Chile". Economics of Education Review, Elsevier, vol. 22(2), pages 131-141, April.

Micklewright, J., Schnepf, S.V. & Silva, P.N. (2010). "Peer effects and measurement error: the impact of sampling variation in school survey data". DoQSS Working Papers 10-13, Department of Quantitative Social Science - Institute of Education, University of London.

Organisation for Economic Co-operation and Development (OECD) 2009. *PISA 2006 technical report*. Paris: OECD.

Rangvid, B.S. (2007). "School composition effects in Denmark: quantile regression evidence from PISA 2000". Empirical Economics, Springer, vol. 33(2), pages 359-388, September.

Raymond, J.L. & Roig, L. (2010) “Intrafirm human capital externalities and selection bias”, Applied Economics, Taylor and Francis Journals (forthcoming).

Robertson, D. & Symons, J. (2003). "Do Peer Groups Matter? Peer Group versus Schooling Effects on Academic Attainment" Economica, London School of Economics and Political Science, vol. 70(277), pages 31-53, February.

Sacerdote, B. (2001). "Peer Effects With Random Assignment: Results For Dartmouth Roommates". The Quarterly Journal of Economics, MIT Press, vol. 116(2), pages 681-704, May.

Schneeweis, N. & Winter-Ebmer, R. (2007). "Peer effects in Austrian schools". Empirical Economics, Springer, vol. 32(2), pages 387-409, May.

Zimmerman, D.J. (2003). "Peer Effects in Academic Outcomes: Evidence from a Natural Experiment". The Review of Economics and Statistics, MIT Press, vol. 85(1), pages 9-23, November.

TABLES AND FIGURES

Table 3: School composition spillovers (level + shape effects)

Panel A SCHOOL COMPOSITION SPILLOVERS – BASELINE (eq. 5)		
School-Average Parental Education	*IEX_j*	*t-Statistic*
quintile 1	*reference group*	
quintile 2	4.876	1.153
quintile 3	12.105	2.235
quintile 4	26.112	3.010
quintile 5	71.215	7.802
mean	27.718	6.920
Panel B SCHOOL COMPOSITION SPILLOVERS – SELF-SELECTION CORRECTED (eq. 8)		
School-Average Parental Education	*IEX_j**	*t-Statistic*
quintile 1	*reference group*	
quintile 2	-9.808	-2.546
quintile 3	2.962	0.552
quintile 4	15.431	3.698
quintile 5	15.072	3.706
mean	1.211	0.468

Note: the test score equations for the reference groups contain the same control variables as those included in table 3 and have been estimated with the STATA command "pv", which means the characteristics of the PISA sample can be taken into account, see section 4.

Table 4: Falsification test for the validity of the exclusion restrictions

Panel A SCHOOL COMPOSITION SPILLOVERS – BASELINE (eq. 5)			**School Availability included in eq. (3)**		**Age on Arrival included in eq. (3)**		**School Availability and Age on Arrival included in eq. (3)**	
School-Average Parental Education	*IEX_j*	*t-Statistic*	*IEX_j*	*t-Statistic*	*IEX_j*	*t-Statistic*	*IEX_j*	*t-Statistic*
quintile 1	*reference group*		*reference group*		*reference group*		*reference group*	
quintile 2	4.876	1.153	4.590	1.049	4.841	1.148	4.544	1.041
quintile 3	12.105	2.235	12.643	2.288	12.089	2.207	12.639	2.261
quintile 4	26.112	3.010	26.181	3.119	26.069	3.005	26.132	3.116
quintile 5	71.215	7.802	71.454	7.713	71.111	7.807	71.322	7.716
Mean	27.718	6.920	27.826	6.944	27.669	6.909	27.768	6.933
Panel B SCHOOL COMPOSITION SPILLOVERS – SELF-SELECTION CORRECTED (eq. 8)			**School Availability included in eq. (7)**		**Age on Arrival included in eq. (7)**		**School Availability and Age on Arrival included in eq. (7)**	
School-Average Parental Education	*IEX_j**	*t-Statistic*	*IEX_j**	*t-Statistic*	*IEX_j**	*t-Statistic*	*IEX_j**	*t-Statistic*
quintile 1	*reference group*		*reference group*		*reference group*		*reference group*	
quintile 2	-9.808	-2.546	-8.789	-2.249	-11.068	-2.880	-10.063	-2.583
quintile 3	2.962	0.552	4.947	0.913	1.361	0.254	3.327	0.616
quintile 4	15.431	3.698	16.940	3.998	13.132	3.105	14.592	3.401
quintile 5	15.072	3.706	17.413	4.325	12.671	3.210	14.988	3.841
Mean	1.211	0.468	2.727	1.040	-0.470	-0.183	1.022	0.393

Appendix

Table 1A: School composition variable

School socio-economic composition variable: SCHOOL-AVERAGE PARENTAL YEARS OF EDUCATION (HIGHEST)			
	num. obs.	**mean**	**s.d.**
quintile 1 (ref. group)	3854	9.144	3.840
quintile 2	3827	10.752	3.905
quintile 3	3833	11.725	3.698
quintile 4	3823	12.699	3.479
quintile 5	3827	14.562	2.913

Note: the years of education are based on the OECD's standard (ISCED97) for levels of completed education; calculations include the final student weight.

Table 2A: Ordered Probit estimation results (eq. 6)

DEPENDENT VARIABLE: 5 quintiles of the school-average parental education		
	Coefficient	**z-Statistic**
age on arrival	-0.034	*-4.930*
no other schools	*Ref. Cat.*	
one school	0.348	*2.180*
more than one school	0.635	*4.040*
small village	*Ref. Cat.*	
village	0.423	*2.260*
town	0.391	*3.370*
city	0.927	*8.490*
large city	1.175	*3.970*
Other Regions	*Ref. Cat.*	
Andalusía	-0.514	*-2.060*
Aragón	0.285	*1.360*
Asturias	0.669	*3.320*
Cantabria	0.582	*3.070*
Castilla y Leon	0.492	*2.280*
Cataluña	0.270	*1.140*
Galicia	0.285	*1.240*
La Rioja	0.414	*2.030*
Navarra	0.720	*3.620*
País Basco	0.927	*5.020*
cut-off point 1	0.420	
cut-off point 2	1.136	
cut-off point 3	1.743	
cut-off point 4	2.434	
Number of Observations		19164
Pseudo R^2		0.098

Robust standard error with school-clusters (675 schools).