WHAT ACCOUNTS FOR PORTUGUESE REGIONAL DIFFERENCES IN STUDENTS' PERFORMANCE? EVIDENCE FROM OECD PISA*

*Manuel Coutinho Pereira** | Hugo Reis***

ABSTRACT

This paper studies regional differences in students' educational performance and inequality in Portugal. Despite the centralized nature of the Portuguese educational system, there are significant differences across regions. We consider firstly the role of school and family factors. Results suggest that individual and family backgrounds play an important role in explaining both achievement and inequality. School characteristics are also important but only in terms of performance, while the role of "pure" regional effects is limited. From a policy perspective there is scope for school intervention, namely regarding school organization and teachers' responsibilities. Nevertheless, to target educational inequality, educational policy needs to take into account the school-family-community context and should not focus exclusively on schools.

1. Introduction

Despite some improvements in various educational statistics in the last decade, Portugal still ranks low among OECD countries. For instance, only 32 per cent of working-age population has attained at least upper secondary education in 2010 compared to the OECD average of 75 per cent.[1] Furthermore, the high drop-out rate associated with low skills remains a major problem. These disturbing figures are not homogeneous across Portuguese regions. For example, the percentage of working-age population that has attained at least upper secondary education in 2010 goes from 20 per cent in Madeira and Azores to around 45 per cent in Lisboa. The illiteracy rate goes from around 3 per cent in Lisboa and Porto to about 10 per cent in Alentejo. Moreover, indicators of educational achievement in Portugal, such as the results of national examinations, show important territorial variation. It is worth noting that the regional profile of educational outcomes and educational achievement seem to be positively associated. The OECD's Programme for International Student Assessment (PISA) 2009 included, for the first time, detailed information about the Portuguese regional distribution of the students in the sample[2], which confirms the mentioned regional disparities. Therefore, given the highly centralized nature of the Portuguese Educational System, for instance, as far as teacher hiring and pay and definition of curricula are concerned, it is important to understand what is behind such differences.

* The authors thank Nuno Alves, Mário Centeno, Jorge Correia da Cunha, Ana Cristina Leal and José Ferreira Machado for their comments. The opinions expressed in the article are those of the authors and do not necessarily coincide with those of Banco de Portugal or the Eurosystem. Any errors and omissions are the sole responsibility of the authors.

** Economics and Research Department, Banco de Portugal.

1 Among the OECD countries only Turkey presents similar values. For the youngest group (aged 25-34) figures are better (52 per cent) but still well below the OECD average of 82 per cent.

2 The regional breakdown takes as a reference the Nomenclature of Territorial Units for Statistics (NUTS) – level 3 (see Appendix 1).

This paper investigates the determinants of regional differences regarding the level and inequality of students' performance[3], using the standard education production function approach (Todd and Wolpin, 2003), *i.e.* the knowledge of the process by which education is produced. Education production functions provide the means for understanding this process by estimating the effects of the various inputs on student achievement measured by test scores. The explanatory variables are individual characteristics, family background and school resources. We will also examine the relationship between regional disparities and characteristics.

We start by studying students' achievement. One natural explanatory factor for regional heterogeneity is the diverse socioeconomic background of student population across regions. As a first step, one quantifies and nets out the effect of this background on observed gaps in average scores throughout regions. We then investigate to what extent the remaining differences can be ascribed to schools and pure regional factors. In Portugal, one may expect the existence of very little institutional variation (except as far as public versus private schools are concerned, but then the reduced number of the latter in the PISA sample, precludes taking full advantage of this). Nevertheless, there may be differences among schools, for instance, regarding their organizational features and teachers (e.g. schools located in more affluent areas may attract better teachers). Lastly we examine the sources of education inequality. In the spirit of the Coleman Report (1966)[4], and following Carneiro (2008) and Carneiro and Reis (2009), one compares again the role of family and school factors in determining achievement inequality within regions.

Identifying the sources of achievement level and inequality is particularly relevant for the design of public policies targeting students or schools. Such evidence may, for example, lead to a better perception of how equality of opportunity can be achieved. As far as we are aware, for Portugal, this regional field has hardly been explored (beyond the descriptive level of analysis). Despite being a first analysis, our results are a good starting point to get some insights for the debate on the educational system. In particular, the effectiveness of a centralized educational system compared to a decentralized one, namely regarding school organization, responsibility and accountability.

The estimation of education production functions raises a number of issues. Some of the teacher and school characteristics are unobservable, giving raise to unexplained variation of outcomes across schools and regions. At the same time, the effect of socioeconomic composition of schools (both direct and through peers') on outcomes may not be fully captured by family variables. Moreover, differences in achievement across regions may also reflect pure regional factors which are as well unobservable, for instance, the valuation of knowledge and human capital may vary across regions. Finally, family, school and regional characteristics (observable and unobservable) interact and are most likely correlated with each other. In this case, some variables in the education production function may be endogenous and reflect, to some extent, the effect of unobservables. In spite of these caveats, trying to provide an estimate of the relative importance of family, school and regional environment is an interesting and instructive exercise.

Our study contributes to the vast literature on educational performance. In particular, it belongs to the strand of literature devoted to regional analyses of PISA outcomes such as Wössman (2007), for Germany, Bratti et al. (2007), for Italy, and Ferrera et al. (2010), for Spain. Some of these regional studies, notably for Germany, take advantage of within-country institutional variation given the decentralized nature of their educational systems. The article is organized as follows. In Section 2, we describe overall patterns in the data. Section 3 presents the regional analysis of educational achievement. Section 4 examines within- and between-region educational inequality. Section 5 concludes.

---

3 Portuguese achievement levels in PISA showed some convergence to the OECD average between 2000 and 2009 (Pereira, 2011). Unfortunately, we cannot explore the regional dimension of such evidence due to data restrictions.

4 The Coleman Report was a seminal study, for the United States, investigating the relative role of family factors and school resources in achievement, highlighting the importance of students' socieconomic background and social inequality (segregation).

2. The database and descriptive analysis

The PISA 2009 database for Portugal comprises 6298 students belonging to 214 schools of which 209 are assigned to regions of NUTS3. The student, family and school variables used in the regressions throughout the paper are basically those already employed in previous studies using PISA data, such as Pereira (2010, 2011). There are a couple of additions, nevertheless, which are worth highlighting (a full list of the variables used, and respective means by region, is given in Appendix 2).[5] An indicator of year repetition was computed from questions included in the student questionnaire, which allows separating the effect of repetition from the effect of exposure to different curricula (captured by grade, also part of the set of regressors). In this context, the variable age entering the regressions in previous studies becomes redundant (see the discussion in Pereira, 2010, about the interaction of grade and age). A wide group of school variables existing in the PISA 2009 database and covering aspects for which information is normally less readily available was taken on board as well. These variables include, in particular, indicators capturing aspects of students' and teachers' behavior that may affect school outcomes, the way activities of teachers are monitored (e.g. through peer review), and the existence of extra-curricular activities at schools. In addition to using data from PISA, we also comment on the correlation of certain results with regional indicators covering economic characteristics, literacy, attitude towards education, attractiveness of the region and social behaviour.

Regional breakdown

The NUTS3 breakdown divides the Portuguese territory into 28 regions. PISA is a sampling survey and for some of these regions only a reduced number of students and schools were sampled (namely, around 50 students belonging to 2 schools). Therefore, it was necessary to use a more aggregated regional breakdown. At the same time, similarities among some regions of the NUTS3 allow further aggregation without raising big homogeneity concerns. We have aggregated the 28 regions of the NUTS3 into 12 - Norte Interior, Norte Litoral, Grande Porto, Centro Interior, Centro Litoral, Vale do Tejo, Grande Lisboa, Alto Alentejo, Península de Setúbal, Baixo Alentejo, Algarve e Ilhas (Chart 1) - which in our view strike the right balance between aggregation needs and capturing regional variability across Portugal[6]. The correspondence between the NUTS3 and this 12-region breakdown is presented in appendix 1.

Test scores

Chart 2 shows the mean score in PISA 2009 for mathematics and reading by region. We found it useful to show for comparison the mean score in the 2009 national exams at the end of basic education for mathematics and Portuguese language (re-scaled to have the mean of PISA scores). As far as PISA scores are concerned, the results generally correspond to what one would expect, especially in that Lisboa, Porto and coastal regions in-between feature the highest levels of achievement. There are a couple of more unexpected findings though, for instance, the high scores of students in Centro Litoral, slightly surpassing their colleagues in Lisboa and Porto in mathematics, and the low achievement levels of Algarve and Setúbal, in spite of the fact that these regions have relatively favourable development indicators. The difference between the maximum and minimum score across regions (50 to 60 points) is around 2/3

5 Similarly to previous studies, missing values for several regressors were imputed through a regression procedure (see Pereira, 2010, Appendix 2, for more details) taking as core variables grade, age, gender, school location and region.

6 It is worth noting that our regional breakdown is still more disaggregated than those used in studies for other countries, given the respective sizes. For instance, the aforementioned studies for Germany, Italy and Spain are based on breakdowns with, respectively, 16, 18 and 11 regions.

Chart 1

TWELVE - REGION BREAKDOWN

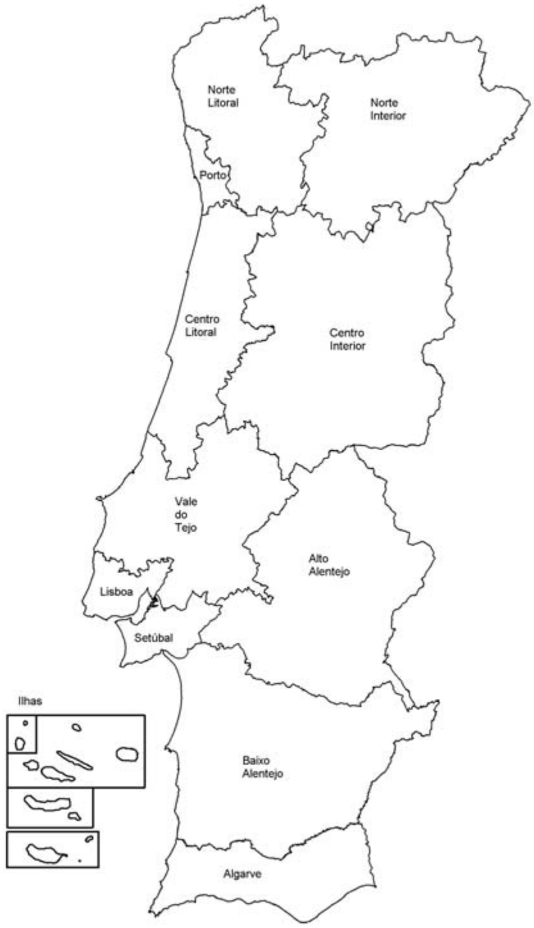
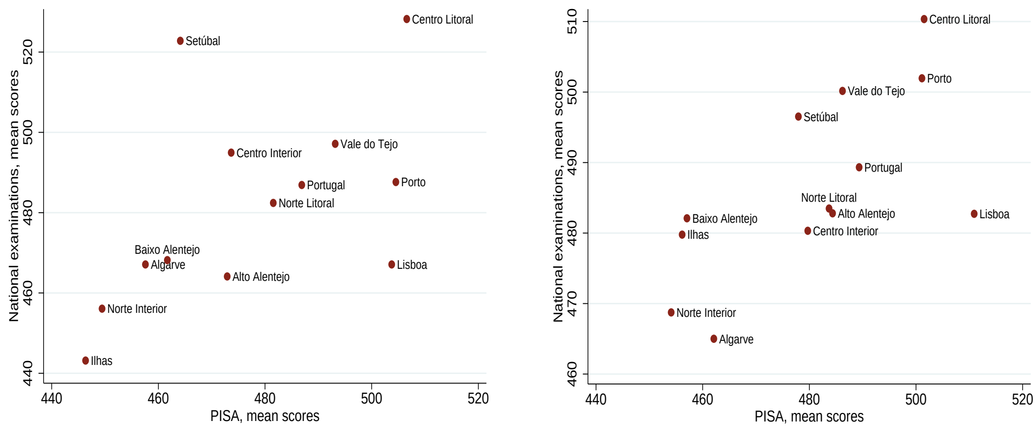


Chart 2

PERFORMANCE BY REGION AND OVERALL TOTAL

Mathematics	Reading



Source: Authors' calculations.

Notes: PISA mean scores are computed averaging out the means for the five plausible values (student data). National examination mean scores are computed from the results by NUTS3 reported in GAVE (2012) and are scaled to have the overall mean of PISA scores; the figure for Ilhas includes Madeira only, as Azores results are not reported.

of one standard deviation both for mathematics and reading, a figure very similar to the one for Spain[7] which has performance levels similar to Portugal.

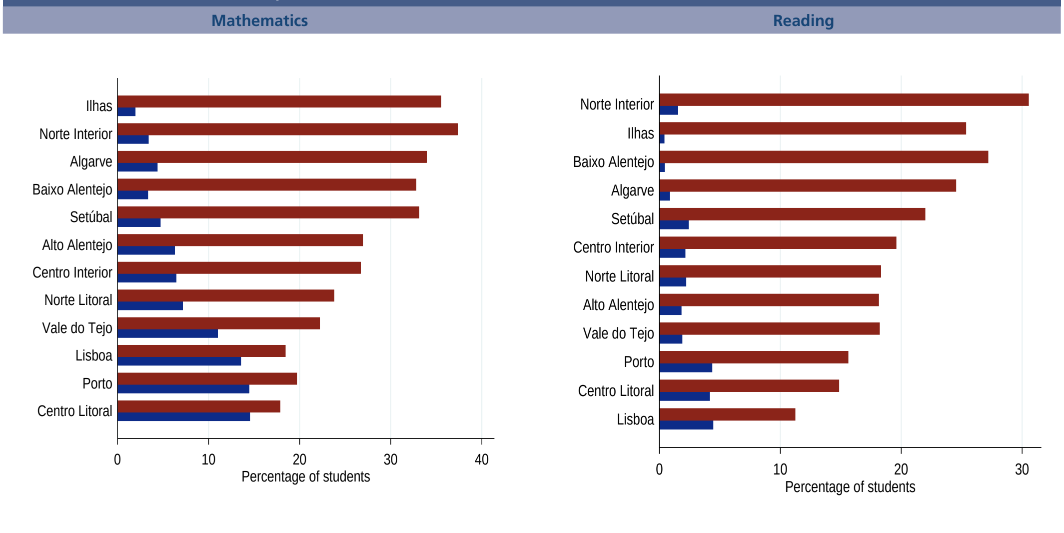
In order to illustrate better what the PISA regional gaps mean in practice, we place the Portuguese regions against the group of 34 OECD countries for which results are available. The best region in mathematics, Centro Litoral, would come just after the 12th country, Iceland, while the worst, Ilhas, would be placed at the bottom of the ranking after the 31st position, occupied by Israel. A similar comparison for reading indicates more marked disparities, with the top-performing region, Lisboa, in the 6th place, slightly above the Netherlands, and Norte Interior, which has the lowest score, after the 32th country, Turkey. In short, there are important differences in schooling outcomes across Portuguese regions, as measured by PISA scores.

We now compare the outcomes in PISA and in the national exams. Chart 2 shows visible correlation between the regional scores in each source (though there are a few exceptions, namely, Lisboa and Setúbal for mathematics, and again Lisboa for reading).[8] Therefore the findings of an analysis based on PISA outcomes, such the one presented here, would most likely remain valid, if the investigation was based on outcomes of formal testing procedures like the national exams. The differences in measured achievement according to the two sources may be accounted for by several reasons. Firstly, PISA is geared toward assessing the acquisition of skills believed useful for productive life, while the national exams evaluate the knowledge of a pre-defined curriculum. Secondly, the target population does not entirely match in the two sources (students aged 15, spreading throughout several grades in PISA vs. students in the 9th grade in the national exams). Thirdly, the PISA survey is based on a sample that covers only a fraction of the relevant student population.

Chart 3 presents the regional breakdown of PISA results in terms of the proficiency levels, which link scores to the actual degree of difficulty of the tasks students can perform (see, for instance, PISA, 2010,

Chart 3

PROFICIENCY BY REGION | STUDENTS AT LEVEL 1 AND BELOW (IN RED) AND AT LEVEL 5 AND ABOVE (IN BLUE)



Source: Authors' calculations.
Note: The percentages shown are computed averaging out the relevant percentages of students for the five plausible values.

7 Considering the regional breakdown presented in the Annex B2 of PISA (2010) and excluding the region Ceuta y Melilla which has much worse outcomes than any other Spanish region.

8 The better educational outcomes of Lisboa in PISA relative to national exams could reflect the fact that the advantage of living in a large city is more visible under PISA-type testing than under curricula-based assessments. Another possible reason could be that a particularly favourable sample of students was gathered for this region in PISA 2009. As regards Setúbal, the very goods results in the 2009 mathematics exam may have been an outlier; in 2011 the region has results at an intermediate-low level.

Chapters 2 and 3). The charts show in red the proportion of students who are not able to perform tasks which enable them to participate productively in society (proficiency level 1 and below), and in blue the share of students who are in a position to complete rather demanding tasks. Regions are sorted according to their average score. There is a high proportion of students at a very low proficiency level, especially in mathematics, in the five worst-performing regions. Furthermore, for mathematics, the decrease in the proportion of students at the lower cohorts, as average performance goes up, is matched by an increase at the upper cohorts. This indicates that the regional distributions are shifting to the right, but are about equally compressed. In contrast, for reading, the increase in the average score is mainly due to the decrease of the number of students at lower performance cohorts, meaning that the regional distributions become somewhat more compressed as the mean increases.

Explanatory variables

We end this section with a brief analysis of regional statistics for the explanatory variables (see Appendix 2). Starting with the repeater indicator, it shows a marked regional variation with values going from 28 per cent in Centro Litoral and Porto to 52 per cent in Algarve. Given the observed regional heterogeneity, it is not reasonable to presume that the indicator is reflecting only disparities in students' innate ability.[9] The condition of repeater may reflect other factors associated to families, schools and even regions (thus although classified as a student variable for convenience, the scope of the repeater indicator is broader). Considering the breakdown by grade, there is also important variation throughout regions: the proportion of students in the 10th grade ranges from 37 per cent in Algarve to 68 per cent in Centro Litoral. There is obvious correlation between the distribution by grades and the repeater condition. However, such distribution is also influenced by PISA sampling procedures (see Pereira, 2011).

Concerning family variables, the pattern of variation seems to be in general the expected one, in line with the relative living standards prevailing in the regions. Ilhas stands out for having a much worse situation than any other region, included the disadvantaged ones, as far as the wealth and educational resources indicators are concerned. As regards parental education and occupations, it is the higher position of Lisboa that stands out, even vis-a-vis the other best-performing regions. For example, the share of students who have at least one parent with tertiary education is 47 per cent in Lisboa, and 28 per cent while in Centro Litoral, the second highest. Regions with low levels of achievement, such as Ilhas and Norte Interior, visibly lag behind in terms of socieconomic indicators, although there are exceptions and performance and socioeconomic variables do not always move in the same direction.

This study considers a large number of school variables. Although there is much heterogeneity in the patterns of variation across regions, a number of general points can be made. Given the centralized nature of the Portuguese school system, it is understandable that some institutional variables point to regional uniformity. Such is the case of the indicators of autonomy in allocation of resources, curricula definition and assessment methods[10], and hours of regular lessons. The average school size has important discrepancies, ranging from around 400 students in Baixo Alentejo to 1200 in Ilhas. Considering this indicator in conjunction with grade amplitude, one can further observe that the size of schools in these two regions is associated, respectively, with the narrower and wider scope of grades offered; in other cases, such as Norte Interior, schools are relatively small despite having a wide scope of grades.

The resource indicators show a mixed picture. The class size shows some variability, ranging from around 19

9 One may suppose that for a large number of students (for instance, if the full population was being considered), innate ability could average out to similar values across regions. In the PISA case, though, the sampling process may introduce some regional heterogeneity in this respect.

10 These indicators are standardized to having mean zero and unit standard deviation across the OECD. Therefore, the figures for Portugal (-0.44 and -0.93, respectively, for the autonomy of resources and curricula/assessment indicators) imply that Portuguese schools enjoy little autonomy for OECD standards.

students in Ilhas to 24 in Porto, assuming larger values in the more populated areas; the same tendency, in this case showing smaller figures, can be observed for the student-teacher ratio. In contrast, schools report uniformly throughout regions an absence of teacher shortage and a high proportion of full-time teachers. As regards material resources, variables related to availability of computers and internet connections do not differ much across regions (except for Ilhas which has a very high figure for the former variable), while the indicator of educational resources at school (that have a broader scope than just IT equipment) reveals more marked gaps across regions. Some of the remaining variables considered measure potentially important explanatory factors, but are at the same time more prone to being affected by the subjective judgment of who filled in the questionnaire. Indicators for student and teacher attitudes that can affect the school climate show some regional heterogeneity, as do the indicator of leadership, measuring the involvement of the management in school affairs, and the indicator of teacher monitoring (tests and peers). The proportion of schools that report parental pressure to raise standards is generally low (the highest figure is 27 per cent for Lisboa) and completely absent in some regions.

3. Regional profile of educational achievement

3.1 The role of individuals and families

We saw in the previous section that students in wealthier regions tend to have better performance and that other variables, such as their distribution between the 9th and the 10th grade, also showed considerable regional variation. In face of this evidence, our investigation starts by quantifying the impact on performance of the student and family variables and determining what remains of the initial regional gaps after these variables are controlled for. We follow the education production function approach, which relates test scores with student, family and school factors. Note that there are unobservable variables that affect test scores and, at the same time, are likely to be correlated with some of those regressors. Hence, estimation results cannot be given a straightforward casual interpretation. Nevertheless, the use of school fixed-effects (i.e. binary variables for each school), as explained below, allows us to control for all observed and unobserved school characteristics, minimising the problems regarding identification of the effects of individual and family characteristics. Moreover, the fixed-effects for the full set of schools within a given region add up exactly to the respective regional fixed effect, and will thus capture regional variability as well. We estimate by OLS (pooling data for all regions) the following education production function:

$$T_{ijr} = \alpha + \beta F_{ijr} + \gamma \phi_{jr} + \varepsilon_{ijr} \quad (1)$$

where T_{ijr} is the test score of student i of school j in region r, F_{ijr} is a vector including regressors for gender, repeater condition, grade and the set of socioeconomic characteristics listed in appendix 2, and ϕ_{jr} is a vector of school fixed-effects. As said, their inclusion allows a more accurate estimation of the coefficients of regressors in F_{ijr}. The conditional mean for a given region can be retrieved as the (weighted) average of the estimated coefficients of the fixed-effects for all schools located there (i.e. averaging out the coefficients of ϕ_{jr} over each region).

We first report briefly on the estimation results for the regression above (see Appendix 3). These are very much the expected ones, with the repeater and grade indicators clearly significant and having the strongest impact on test scores (note that the size of the coefficients can be directly compared for binary variables). Family indicators are as well generally significant and, as it is often the case, the number of books at home stands out as the most important regressor in this set. As far as parental education and occupations are concerned, only the upper categories (respectively, upper secondary or tertiary and white collar/highly skilled) seem to make a difference for test outcomes although with a relatively small impact.[11]

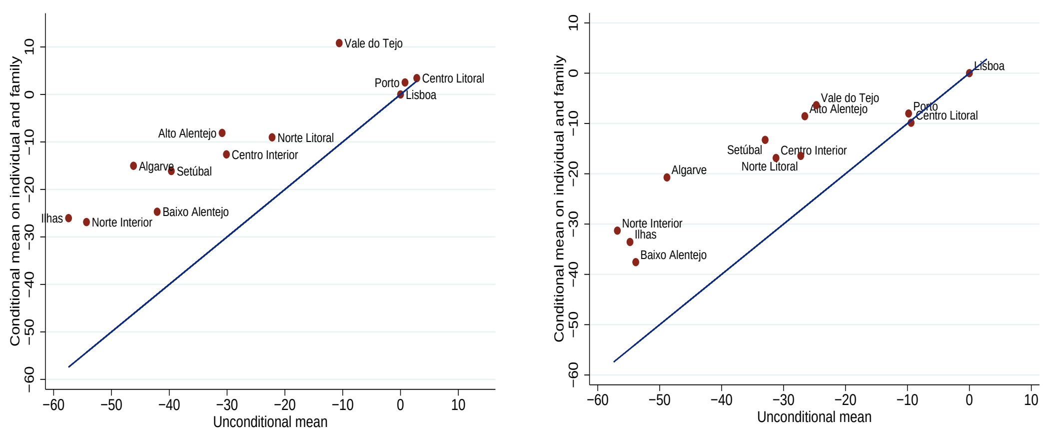
The results are shown in Chart 4, in terms of the gap of each region vis-a-vis a reference region[12] - for which Lisboa was chosen. For the sake of comparison, the corresponding results for the unconditional mean are also shown. When the conditional mean is taken, the gap between Lisboa (or, more generally, the top-performers) and the regions with intermediate to low achievement shortens, albeit remaining negative, both for reading and mathematics. Such regions appear in the charts to the left of the 45° line, and the distance to this line measures the magnitude of the difference between the two means (which is greatest for Ilhas, Norte Interior and Algarve). This reflects a comparatively unfavourable situation vis-a--vis Lisboa as far as socioeconomic composition and/or student variables are concerned. In contrast, the situation of Porto and Centro Litoral in relation to Lisboa barely changes, indicating similar characteristics in terms of the variables which are being held constant. Vale do Tejo builds an exception in that, having already relatively high test scores, it clearly improves the position against the other top-performers, when conditional mean scores are taken (especially in mathematics).

The evidence resulting from Chart 4 indicates that student and family variables although important explain only part of the unconditional regional gaps. Note, in particular, that the initial relative position of the various regions is roughly preserved after student and family variables are controlled for.[13] Nevertheless, some shrinkage of the gaps across regions follows and, in line with this, the respective statistical significance becomes less sharp.[14] For instance, in the unconditional analysis Lisboa's mean in mathematics is significantly different to every region except for the other three in the group of four top performers (Centro Litoral, Porto and Vale do Tejo); in the conditional analysis the mean gap to Lisboa becomes, in addition, not significant vis-a-vis Algarve, Alto Alentejo and Centro Interior. This weakening of the statistical significance of gaps, holding constant the family and student variables, is clearer for reading.

Chart 4

TEST SCORES BY REGION, GAP TO LISBOA

Mathematics | **Reading**



Source: Authors' calculations.

Notes: The y-axis shows the regional averages of the coefficients of school fixed-effects in regression (1), estimated pooling the data for all regions; the x-axis shows the unconditional mean.

11 A more unexpected result concerns the estimated positive influence on scores of belonging to a monoparental family. This variable could be capturing a possible higher socieconomic standing of such families, but this should be controlled for by the other family regressors included in equation (1).

12 We show the results as gaps between regions rather than absolute values because the level of conditional means is of difficult interpretation.

13 This issue is addressed in more detail at the end of Section 3.2.

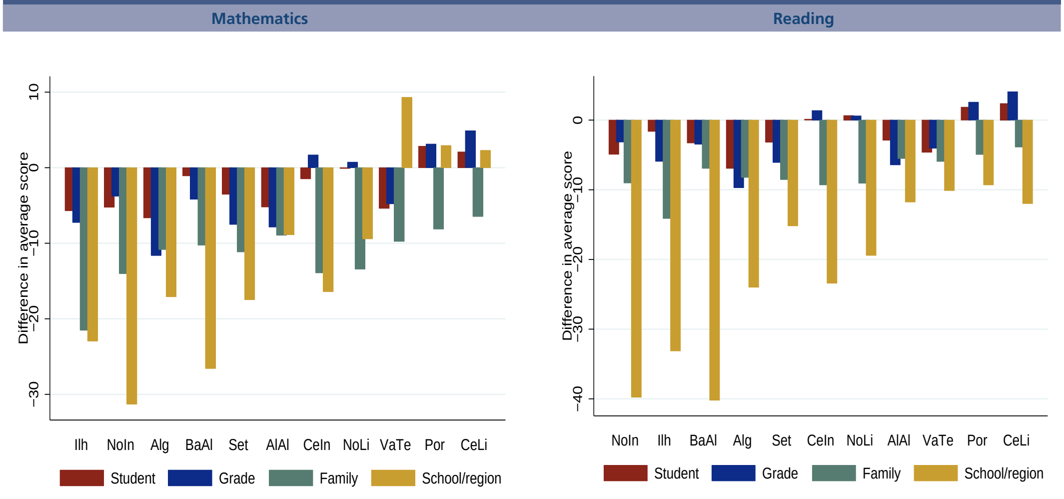
14 Matrices with the significance of mean differences for all pairs of regions are available upon request.

In this case, if one excludes the best-performing region, Lisboa, and the three worst performers, Norte Interior, Ilhas and Baixo Alentejo, the other regions belong to an intermediate group whose mean scores are statistically not different from each other.

We finalize this section by presenting a decomposition of the regional average gaps vis-a-vis Lisboa by means of a Oaxaca-type decomposition, into what is accounted for by student variables proper (gender and repeater indicator), grade indicators and socieconomic variables, i.e. the regressors included in vector F, and an unexplained part which we assign to schools and regions. This unexplained part reflects the difference in the estimated coefficients for the constant and student and family variables between each region and Lisboa, plus the difference accounted for by the school-fixed effects.[15] These results complement the evidence presented in Chart 4, since the difference between the unconditional and conditional gaps is conceptually the sum of the student, grade and family effects, while the remaining gap corresponds to the unexplained part. Chart 5 confirms that the influence of schools and regions (yellow bar) is generally at least as important as the impact of families and individuals (which corresponds to the sum of the remaining bars). The charts indicate for all regions an unfavourable socioeconomic composition vis-a-vis Lisboa. In most of them the distribution of students by grade also contributes negatively to the gap to Lisboa, and in certain cases (notably, Algarve, Setúbal and Alto Alentejo) has an effect comparable to that of family. The role of student variables is essentially driven by the repeater indicator, as the average figure for gender has very little regional variation. Most regions are penalized in the results for having a higher proportion of repeaters than the reference region, particularly those with an intermediate to low level of performance. Note that, as already mentioned, this indicator is most likely capturing a mixture of effects, going from students' innate ability to family, school and regional influences.

Chart 5

DECOMPOSITION OF REGIONAL GAPS TO LISBOA



Source: Authors' calculations.

Notes: The decomposition is based on the estimation of equation (1) by region. The effects of student, grade and family variables are calculated as $\beta_L(F_{ijr}* - F_{ijL}*)$, where $F_{ijr}*$ and $F_{ijL}*$ are, respectively, the averages of *F* regressors in region *r* and Lisboa, and β_L are the respective estimated coefficients for Lisboa. The school/region effect is calculated subtracting the effects of student, grade and family variables from the difference in the unconditional means between region r and Lisboa (this corresponds to $(\beta_r - \beta_L) F_{ijr}* + \gamma_r ø_{jr}* - \gamma_L ø_{jL}* + \alpha_r - \alpha_L$, where β_r, γ_r, γ_L, α_r and α_L are the additional estimated coefficients for region r and Lisboa in equation (1), and $ø_{jr}*$ and $ø_{jL}*$ the average fixed-effects).

15 This Oaxaca-type decomposition deviates from the traditional version in that it includes the school fixed-effects that cannot be compared across regions. Therefore, in our decomposition the unexplained part comprises not only the traditional difference between the coefficients estimated for each region (for the regressors in F and the constant term), but also what is accounted for by the school fixed-effects.

3.2 The role of school characteristics

In this section we want to understand to what extent observable school characteristics explain the remaining regional differences described in previous section, i.e. after controlling for individual and family background. More specifically, we regress the estimated school fixed effects from Section 3.1 ($\hat{\gamma\phi}_{jr}$), on observable school variables[16] (S_{jr}) and regional fixed effects (φ_r); ξ_{jr} represents the usual error term.

$$\hat{\gamma\phi}_{jr} = \eta S_{jr} + \varphi_r + \xi_{jr} \qquad (2)$$

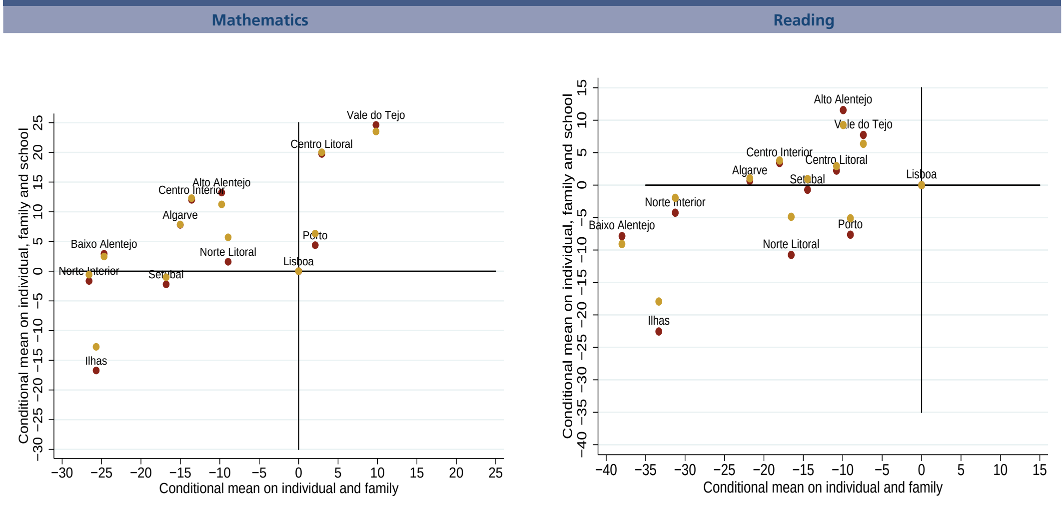
Chart 6 presents, for mathematics and reading, the regional fixed effects (the remaining regional gap) of the two specifications estimated from equation (2). One using exclusively school variables (red dots), and another one in which we add the possible effect of student peers (yellow dots). In general, observable school characteristics appear with the expected sign and are jointly significant (F-test). Results can be found in appendix 3. As before, the regional results represent differences to Lisboa. Notice that, if observable school variables are not enough to explain such differences, it means that there is a role for unobservable school characteristics and pure regional effects.

Firstly, for mathematics, the chart shows an improvement of the position of all regions relative to Lisboa, except for Porto, after controlling additionally for observable school characteristics. These regions appear in the charts to the left of the 45° line (not shown), reflecting a negative contribution of the observable school resources to the respective scores in comparison to Lisboa. In particular, for Norte Interior, Baixo Alentejo and Centro Interior this effect is very strong, which is suggestive, for example, of the low levels of educational resources and parental pressure. Moreover, after controlling for both family and school characteristics, the number of regions with better performance than Lisboa increases considerably, with the gap changing sign in several cases (Alto Alentejo and Centro Interior stand out in this respect).

Note that, the remaining differences among regions are, in general, not statistically significant with the exception of Vale do Tejo and Centro Litoral, on the positive side, and Ilhas on the negative. A student with the same family background and attending a similar school would perform better in Vale do Tejo and Centro Litoral than in any other region. Despite the observed convergence, a pupil with the same

Chart 6

REGIONAL FIXED EFFECTS (GAPS TO LISBOA)



Source: Authors' calculations.

Notes: Red - the regression used only school variables; yellow - the regression included also a proxy for peer effects (average of books at home at the school level). In the x-axis we have equation (1) in which we regress school fixed-effects only on regional fixed effects (conditional on individual and family). In the y-axis we have equation (2) results (conditional on individual, family and school).

16 A full description of the variables can be found in appendix 2.

family background and attending a similar school in Ilhas would still perform worse than in Lisboa and other regions.

The results in reading show a similar pattern. In general, observable school variables are contributing for worse results in the various regions in comparison to Lisboa, except for Porto. The remaining differences among almost all regions fade away, *i.e*. with the same familiar and school context performance would be similar. Only Ilhas and Norte Litoral continue to present statistically significant worse results than other regions: given the family background and school resources, a student there would still perform worse.

An additional exercise was performed to infer the potential regional differences in terms of peer effects. We add to equation (2) a family background variable at school level to proxy the peer effects.[17] As we can see in Chart 6 (yellow dots), the impact of peer effects seems to be relatively modest, except for Ilhas. In this case, the remaining gap becomes smaller, albeit remaining negative and significant.

Although our observed school variables may not vary exogenously and may reflect the effect of unobserved school variables, it is clear that schools and not only the family background have an important role in determining test scores. The importance of schools emphasizes that there is scope for educational policy to reduce existing differences regarding school resources and organization, notably as far as teachers' role is concerned. In contrast, using PISA 2000 Carneiro (2008) found that school resources were particularly unimportant. One possible explanation for this result is that few teacher variables were available in 2000. Nevertheless, our results do not invalidate that an innovative education policy is needed so that the resources accessible to schools are better used and the role of family should be taken into account.

This article contributes to the discussion of whether educational policy may be more decentralised in terms of school responsibility, organization and accountability. The results suggest that among the observable characteristics policy should focus on quality of educational resources and pay more attention to extracurricular activities. Educational policy should also focus on the allocation of resources by school staff, in particular, attributing more responsibility to teachers and paying attention to the way teachers are monitored (more peer review). Finally, it should be given the correct incentives for more parents' participation in school activities and discussions. In this particular case, families' contribution is likely to be as important as schools'.

A range of past and current reforms in education are underway in Portugal and cover some of the issues mentioned before. In particular, reforms related to school autonomy, teacher appraisal, school leadership and student learning standards (for details see the OECD Report Reviews of Evaluation and Assessment in Education 2012). Nevertheless, it is important to guarantee the enforcement and correct evaluation of the effectiveness of such policies, namely through school and teacher accountability. In terms of educational resources, despite the importance of providing more and better resources to schools, some of the past programs revealed inefficient.

Finally, table 1 presents the correlation at regional level between the three different regional gaps studied in this paper: (i) unconditional regional fixed-effects; (ii) regional fixed-effects after controlling for family background and (iii) remaining regional difference after controlling for observable school variables as well. Interestingly, the two first measures are highly correlated, while after adding school resources the correlation is substantially lower. This is indicative that family regional differences are not enough to change the initial profile of PISA test scores. In contrast, controlling in addition for schools, changes the pattern of the regional gap initially observed, in particular in the case of reading.

17 We used the more than 200 books dummy variable as a proxy of family background at school level.

Table 1

CORRELATION BETWEEN MEASURES OF REGIONAL DIFFERENCES

	Mathematics			Reading		
	Unconditional	Conditional on family	Conditional on family and school	Unconditional	Conditional on family	Conditional on family and school
Unconditional	1			1		
Conditional on family	0.92*	1		0.92*	1	
Conditional on family and school	0.56*	0.69*	1	0.39	0.58*	1

Source: Authors' calculations.
Notes: unconditional: test scores - regional average; conditional on family: test scores - regional average controlling for individual and family characteristics; conditional on family and school: test scores - regional average controlling for individual, family and school characteristics. * statistically significant at 10%.

Remaining regional differences

Despite the fact that most of the unconditional gap is strongly reduced after controlling for family and school resources, it is important to understand what can explain the remaining disparities. Therefore, we perform a simple correlation analysis of these regional differences and the regional environment (Table 2).[18] More specifically, we look at the interior/rural desertification (the inability of some regions to get the best professionals as for example experience of teachers in the regions, number of doctors per habitant), structural educational problems (drop-out and literacy rates) and social behaviour as divorce rate and crime rate. Only drop-out rate differences seem to be of some importance as is also highlighted in Chart 7.[19] In the light of this result, we could interpret the persistent difference of Ilhas to other regions as reflecting a relatively low valuation of education and human capital investment. All other analysed variables do not present any significant correlation, which is in line with the modest role left to a pure regional effect on student performance after controlling for family and school resources.

Table 2

REGIONAL GAP AND REGIONAL CHARACTERISTICS (CORRELATION)

	Mathematics	Reading
GDPpc	-0.23	-0.18
Regional Development Index	0.10	0.31
Illiteracy rate	0.12	0.15
Drop-out rate	-0.49*	-0.58*
Compulsary education	-0.11	0.12
Higher education	-0.13	0.06
Pre-school	0.30	0.19
Teachears experience	0.21	0.44
Doctors per habitant	0.02	0.01
Divorce rate	-0.12	-0.05
Crime rate	-0.08	0.06

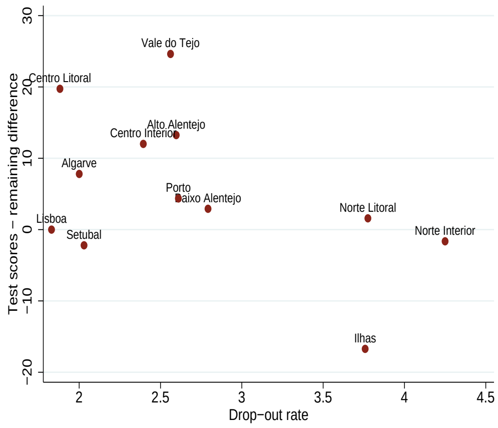
Source: Authors' calculations.
Note: * statistically significant at 10%.

18 As said before, we cannot exclude that these disparities may also reflect school unobservables.

19 Despite the limited number of observations at regional level (12), the drop-out results remain valid after performing some regressions with 2 and 3 variables.

Chart 7

REMAINING REGIONAL DIFFERENCES (MATHEMATICS)



Source: Authors' calculations.

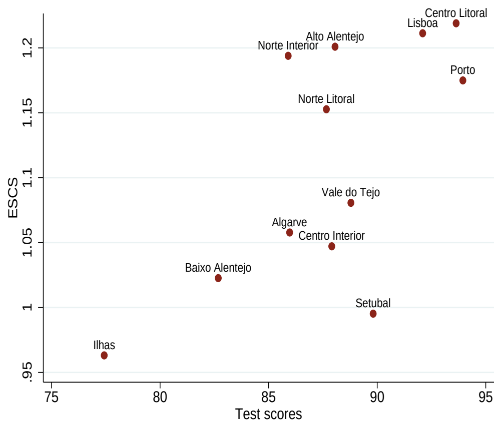
Note: y-axis - remaining differences represent the regional fixed effects estimated in equation (2) and are shown as differences to Lisboa; x-axis: average drop-out rates in percentage by region.

4. Sources of inequality in educational achievement

This section studies inequality in the school performance in the spirit of the Coleman Report. We compare the role of school and family factors in determining inequality within each region. Inequality is a topic of major concern in all open societies and it is likely to emerge well before individuals enter the labour market. Despite the centralised nature of the Portuguese educational system, it is also useful to study if the magnitude of achievement inequality explained corresponds to any difference in regional environment. Chart 8 displays regional standard deviations for test scores in mathematics and for an OECD composite indicator of family background (ESCS), suggesting that higher achievement inequality is associated with more family background inequality.[20]

Chart 8

STANDARD DEVIATION: TEST SCORES AND FAMILY BACKGROUND



Source: Authors' calculations.

Notes: The ESCS is a composite indicator of family background constructed by OECD with PISA data. The R^2 of a regression of this indicator on the family variables used in the article is around 0.9.

20 Notice that higher performance seems to be also associated with more inequality.

We evaluate the sources of inequality in educational achievement among Portuguese regions through a regression-based decomposition approach. We examine the amount of inequality in each region that results from inequalities (i) in family background, (ii) in school resources and organization, and (iii) stemming from poorer families being segregated in worse schools. Then, we relate the importance of each factor (family, school and segregation) to the characteristics of each region such as wealth and development level, region attractiveness, structural educational indicators, and social behaviour.

The measure of inequality we use, the variance, can be easily obtained and decomposed from the estimation of equation (1), by region, as follows:

$$Var(T_{ij}) = Var(\beta F_{ij}) + Var(\gamma \phi_j) + 2Cov(\beta F_{ij}, \gamma \phi_j) + Var(\varepsilon_{ij}) \qquad (3)$$

where the first element represents the contribution of inequality in family characteristics and the second of inequality across schools. The covariance term represents the relationship between school and family factors, i.e. giving an idea if school is exacerbating, being neutral or decreasing initial inequality. In the last case schools are promoting equality of opportunity. The relative contribution can also be easily assessed dividing each element by the total explained variance.

Overall in Table 3 the decomposition shows heterogeneity among Portuguese regions. The variance explained by observable variables ranges from 50 per cent in Ilhas to 62 per cent in Alto Alentejo. Interestingly, these figures are much smaller than differences among European countries, where the differences go from 17 per cent to around 70 per cent (Carneiro and Reis (2008)[21]). Notice that the part left unexplained is still important.

Despite different magnitudes, student and family characteristics play a crucial role in all regions, while school features have a smaller impact on educational inequality. Variance decompositions depend not

Table 3

VARIANCE DECOMPOSITION (VARIANCE EXPLAINED BY DIFFERENT COMPONENTS)

	Portugal	Algarve	Alto Alentejo	Baixo Alentejo	Centro Interior	Centro Litoral	Ilhas	Lisboa	Norte Interior	Norte Litoral	Porto	Setubal	Vale do Tejo
Mathematics													
Var(F)	3302.4	3248.0	3525.9	3354.9	2848.9	3169.3	3771.6	3036.5	3409.6	3585.8	2628.9	4039.6	3526.6
Var(S)	738.7	382.8	1067.7	632.4	561.2	585.4	525.4	837.6	190.6	567.5	991.3	169.1	500.5
Cov(F,S)	451.2	220.2	-200.4	-392.7	489.3	895.1	-944.2	359.6	16.0	155.7	1074.9	144.0	350.0
Var(exp)	4492.4	3851.0	4393.3	3594.6	3899.4	4649.9	3352.8	4233.6	3616.3	4309.0	4695.1	4352.7	4377.1
Var(unexp)	3515.5	3192.1	2714.2	2849.9	3738.5	3550.4	3370.8	3429.3	3013.3	3486.4	3374.0	3300.1	3356.4
Reading													
	Portugal	Algarve	Alto Alentejo	Baixo Alentejo	Centro Interior	Centro Litoral	Ilhas	Lisboa	Norte Interior	Norte Litoral	Porto	Setubal	Vale do Tejo
Var(F)	2753.5	2860.2	2716.2	2971.9	2469.7	2740.6	3375.7	2299.9	3278.4	3038.1	1850.3	3598.1	3506.5
Var(S)	594.4	395.2	356.9	876.4	574.9	477.7	387.8	490.8	546.9	602.6	823.7	328.6	196.1
Cov(F,S)	542.9	594.8	299.2	-715.7	756.2	851.6	-320.1	422.1	942.4	111.2	1088.7	217.4	28.8
Var(exp)	3890.7	3850.2	3372.3	3132.6	3800.8	4069.8	3443.4	3212.8	4767.7	3752.0	3762.7	4144.1	3731.4
Var(unexp)	3260.5	2967.7	2415.9	2900.1	3714.4	3266.4	3178.3	2938.8	2796.5	3061.3	3379.7	3161.8	3031.5

Source: Authors' calculations.
Notes: Var(F): individual and family contribution to total test score variance; Var(S): school contribution to total test score variance; Cov(F,S): relationship between school and family factors; Var(exp): variance explained by equation (3); Var(unexp): unexplained variance as a result of equation (3).

21 Carneiro and Reis (2009) used the 2003 PISA dataset.

only on the variance of the regressors but also on the coefficients themselves. In our case, student and family variables are important to explain differences in achievement but their variance does not change much across regions. Therefore, the higher contribute of pupil and family inequality in certain regions stems from a larger impact of these variables on school performance (as estimated by the coefficients).

In addition, the covariance term presents also distinct results, suggesting the existence of regions more stratified than others in educational terms. In particular, Porto and Centro Litoral present the highest level of segregation, while in Ilhas and both Baixo and Alto Alentejo there is a negative association between observable student/family and school characteristics. In these cases the figures are mainly influenced by the coefficients and not by the covariance level.[22] In the former regions, schools seem to exacerbate initial inequality, while in the latter, schools contribute to decrease inequality. This may be due to several reasons. On the one hand, if students with better individual characteristics and/or from richer families tend to sort into better schools, this correlation will be positive. On the other hand, if a government tries to compensate inequalities in family background and provide extra support to failing schools[23], there may be a negative correlation between school and family features. Both phenomena are likely to be present in our results.

Given the heterogeneity among Portuguese regions it is instructive to document how the importance of each factor is related to some features of each region (Table 4). Using the same characteristics of the previous section, results suggest that regions where school contributes to increase initial inequality are associated with: better structural educational outcomes, higher development and higher inequality in teachers' experience. This result may be, to some extent, related to the availability of more schools in these areas, despite the relatively absence of school choice in the Portuguese educational system.[24] Opposite features are presented by more disadvantaged regions, where schools seem to contribute to reduce inequality of opportunities.

Table 4

CORRELATION BETWEEN VARIANCE DECOMPOSITION AND REGIONAL CHARACTERISTICS

	VAR(F)	VAR(S)	COV(F,S)
GDPpc	-0.12	0.46	-0.10
Regional Development Index	-0.51*	0.32	0.49*
Illiteracy rate	0.43	0.13	-0.49*
Drop-out rate	0.45	-0.14	-0.52*
Compulsary education	-0.28	0.07	0.37
Higher education	-0.45	0.23	0.42
Pre-school	0.09	0.21	-0.28
Teachears experience (years)	-0.32	0.30	0.43
Teachears experience (standard deviation)	-0.55*	0.22	0.72*
Doctors per habitant	-0.67*	0.42	0.57*
Divorce rate	-0.13	-0.01	0.25
Crime rate	-0.12	-0.26	0.25

Source: Authors' calculations.
Note: * statistically significant at 10%.

22 The strong positive and negative results in Porto and Ilhas, respectively, are reflecting the impact of individual variables (repeater status and grade).

23 In Portugal, there are several initiatives and programs with that aim. For instance, accompanied study at schools and the national program supporting educational development in socially segregated and excluded areas - Educational Territories of Priority Intervention (Territórios Educativos de Intervenção Prioritária).

24 This is in line with those that advocate that school choice may increase segregation, by moving good peers to other schools, and may produce competition in irrelevant attributes if parents are careless about educational outcomes. In contrast, those in favour of school choice advocate that school choice may create incentives for schools to increase productivity, offering a product closer to students demand, and expand the choice set for poor students.

Summing up, most inequality is within schools (driven by individual and family factors), and not between schools, which means that schools by themselves cannot explain a large portion of the observed disparities. Therefore, education policy measures alone may be not enough to address achievement inequality, as regional gaps in educational opportunities and outcomes have a wider scope. Policies that focus on poverty and related issues are expected to be more successful than purely educational policy.

5. Conclusions

This article studies educational achievement and inequality throughout Portuguese regions, using data from the OECD PISA 2009. The main findings are the following.

• There are important regional differences in educational performance as measured by PISA scores, and their pattern seems to broadly match the one revealed by scores in national exams. A descriptive analysis indicates that territorial gaps appear to conform to discrepancies in socioeconomic characteristics and educational outcomes across Portugal.

• As expected, student and family variables explain part of the unconditional gaps. Specifically, regions with intermediate to low achievement levels are penalized by an unfavourable socioeconomic composition, a higher proportion of repeaters and a prevalence of students in the 9th or lower grades vs. the 10th grade. Holding these variables constant, there is a shrink of the initial differences and a fading of their statistical significance, although the starting relative position of regions is not substantially changed.

• Schools are found to play an important role in explaining performance differences across the territory. Therefore, when school observables are brought into the analysis, the gaps close further and there are noticeable modifications in the original ranking of regions.

• The role played by schools suggests that there is room for policy interventions to improve their contribution in the regions lagging behind. In particular, the enhancement of school autonomy in the allocation of resources, teacher participation and monitoring, and involvement of parents appear to be fruitful areas of intervention.

• The scope for an important contribution of pure regional factors seems small, although evidence hints at a potential influence of regional disparities concerning the way education is valued.

• The analysis of inequality in educational achievement also reveals some territorial heterogeneity across Portugal. The driving forces behind such inequality seem to be mostly related to students and families rather than schools.

• There is some evidence that schools tend to exacerbate inequality in educational achievement in the more developed regions, and the opposite in the less developed ones. These findings may be related, among other factors, with wider school availability in the first case, as well as the impact of programs targeting the performance of students coming from socially segregated backgrounds in the second.

References

Bratti, M., Checchi, D. and Filippin, A. (2007), "Territorial Differences in Italian Students' Mathematical Competencies: Evidence from PISA 2003," *IZA Discussion Papers 2603*, Institute for the Study of Labor (IZA).

Carneiro, P. (2008), "Equality of opportunity and educational achievement in Portugal", *Portuguese Economic Journal*, vol. 7(1), pp. 17-41.

Carneiro, P. and Reis, H. (2009), "Sources of inequality in educational achievement: an international comparison", UCL manuscript.

Coleman, J., Campbell, E., Hobson, C., McPartland, J., Mood, A., Weinfeld, F. and York, R. (1966), "Equality of educational opportunity", *US government printing office*, Washington, DC.

Ferrera, J., Cebada, E. and González, D. (2010), "Factors affecting regional attainment: Evidence from Spanish PISA 2006 results", *Regional and Sectoral Economic Studies*, vol. 10(3).

GAVE (2012), Exames Nacionais - Relatório 2011, Gabinete de Avaliação Educacional, Lisboa: Ministério da Educação.

OECD (2010), *PISA 2009 Results: What Students Know and Can Do*, Volume I, Paris: Organization for Economic Co-operation and Development.

OECD (2012), *Education at a glance*, Paris: Organization for Economic Co-operation and Development.

Pereira, M.C. (2010). "Educational Attainment and Equality of Opportunity in Portugal and in Europe: The Role of School Versus Parental Influence", *Summer Economic Bulletin*, Banco de Portugal, pp.23-45.

Pereira, M.C. (2011), "An analysis of Portuguese students' performance in the OECD Programme for International Student Assessment", *Autumn Economic Bulletin*, Banco de Portugal, pp.123-136.

Santiago, P., Donaldson, G., Looney, A. and Nusche, D. (2012), *OECD Reviews of Evaluation and Assessment in Education: Portugal 2012*, Paris: Organization for Economic Co-operation and Development.

Todd, P. and Wolpin, K. (2003), "On the Specification and estimation of the Production Function for Cognitive Achievement", *Economic Journal*, 113, F3-F33.

Wössman, L. (2007), "Fundamental Determinants of School Efficiency and Equity: German States as a Microcosm for OECD Countries," *IZA Discussion Papers 2880*, Institute for the Study of Labor (IZA).

APPENDICES

Appendix 1

CORRESPONDENCE BETWEEN THE NUTS3 AND THE 12-REGION BREAKDOWN

	12-region breakdown	Stud. weights population	Schools in sample
Alentejo Central	Alto	0.022	12
Alto Alentejo	Alentejo		
Alentejo Litoral	Baixo	0.019	9
Baixo Alentejo	Alentejo		
Lezíria do Tejo	Vale do	0.074	18
Médio Tejo	Tejo		
Oeste			
Algarve	Algarve	0.029	22
Baixo Mondego	Centro	0.097	19
Baixo Vouga	litoral		
Pinhal Litoral			
Beira Interior Norte	Centro	0.070	18
Beira Interior Sul	Interior		
Cova da Beira			
Dão Lafões			
Pinhal Interior Norte			
Pinhal Interior Sul			
Serra da Estrela			
Alto Trás-os-Montes	Norte	0.036	8
Douro	Interior		
Grande Lisboa	Lisboa	0.178	29
Península de Setúbal	Setúbal	0.068	11
Ave	Norte	0.230	35
Cávado	Litoral		
Entre Douro e Vouga			
Minho Lima			
Tâmega			
Grande Porto	Porto	0.142	20
Madeira	Ilhas	0.035	8
Açores			

Appendix 2 (continue)

DESCRIPTIVE STATISTICS (AVERAGES)

Student variables	Portugal	Algarve	Alto Alentejo	Baixo Alentejo	Centro Interior	Centro Litoral	Ilhas	Lisboa	Norte Interior	Norte Litoral	Porto	Setubal	Vale do Tejo
9th grade(b)	0.27	0.39	0.25	0.31	0.29	0.23	0.24	0.27	0.31	0.27	0.27	0.22	0.36
10th grade(b)	0.58	0.37	0.51	0.52	0.60	0.68	0.53	0.59	0.53	0.60	0.63	0.53	0.48
repeater(b)	0.35	0.52	0.43	0.39	0.37	0.28	0.39	0.33	0.48	0.32	0.28	0.41	0.46
female(b)	0.51	0.49	0.55	0.46	0.51	0.52	0.58	0.51	0.48	0.52	0.49	0.51	0.52

Source: PISA database.
Note: (b) stands for binary variables.

Family variables	Portugal	Algarve	Alto Alentejo	Baixo Alentejo	Centro Interior	Centro Litoral	Ilhas	Lisboa	Norte Interior	Norte Litoral	Porto	Setubal	Vale do Tejo
wealth (ind.)	0.49	0.49	0.56	0.46	0.32	0.56	-0.05	0.54	0.34	0.48	0.59	0.46	0.55
educ. resourc. home (ind.)	0.18	0.05	0.15	0.13	0.22	0.27	-0.08	0.30	0.14	0.12	0.20	0.12	0.21
books at home 25-200(b)	0.48	0.46	0.51	0.48	0.47	0.48	0.33	0.51	0.44	0.45	0.51	0.48	0.53
books at home > 200(b)	0.15	0.15	0.15	0.11	0.10	0.21	0.07	0.26	0.10	0.10	0.15	0.12	0.16
immigrant status(b)	0.05	0.11	0.02	0.05	0.03	0.04	0.00	0.12	0.01	0.02	0.03	0.16	0.03
foreign lang. at home(b)	0.02	0.05	0.01	0.02	0.01	0.01	0.01	0.03	0.01	0.01	0.01	0.02	0.01
blue collar/ high. skilled(b)	0.22	0.15	0.15	0.20	0.30	0.21	0.31	0.09	0.26	0.34	0.20	0.17	0.22
white collar/ low. skilled(b)	0.34	0.47	0.37	0.50	0.35	0.34	0.44	0.32	0.33	0.26	0.31	0.46	0.40
white collar/ high. skilled(b)	0.35	0.31	0.34	0.25	0.25	0.39	0.13	0.53	0.24	0.30	0.40	0.32	0.28
lower sec. educ.(b)	0.23	0.25	0.22	0.28	0.27	0.22	0.26	0.18	0.21	0.24	0.21	0.28	0.27
upper sec. educ.(b)	0.24	0.28	0.28	0.34	0.23	0.27	0.19	0.24	0.25	0.17	0.25	0.34	0.26
tertiary educ. (b)	0.26	0.25	0.26	0.20	0.15	0.28	0.12	0.47	0.14	0.20	0.26	0.23	0.20
one parent home(b)	0.11	0.14	0.12	0.13	0.09	0.09	0.08	0.17	0.13	0.09	0.11	0.11	0.09
no parents home(b)	0.02	0.02	0.03	0.05	0.04	0.02	0.03	0.02	0.05	0.02	0.02	0.04	0.03
ESCS (ind.)	-0.32	-0.38	-0.27	-0.39	-0.63	-0.18	-1.05	0.23	-0.73	-0.56	-0.24	-0.30	-0.42

Source: PISA database.
Note: The ESCS index is used in variance decompositions only. (b) stands for binary variable.

Appendix 2 (continue)

DESCRIPTIVE STATISTICS (AVERAGES)

School variables	Portugal	Algarve	Alto Alentejo	Baixo Alentejo	Centro Interior	Centro Litoral	Ilhas	Lisboa	Norte Interior	Norte Litoral	Porto	Setubal	Vale do Tejo
school size (1000 stud.)	0.94	0.71	0.61	0.41	0.51	0.77	1.20	1.06	0.71	1.10	1.05	0.98	0.82
percentage of girls	50.5	49.6	52.7	44.9	50.0	50.2	49.4	49.6	50.6	51.6	51.0	50.1	51.0
located town 15-100 inh.(b)	0.42	0.84	0.73	0.38	0.33	0.32	0.52	0.18	0.64	0.44	0.53	0.47	0.45
located town > 100 ihn.(b)	0.22	0.01	0.00	0.00	0.00	0.27	0.21	0.66	0.00	0.15	0.18	0.09	0.03
grade ampl. (max-min)	5.7	4.4	5.2	4.9	5.0	5.9	6.2	5.5	6.4	5.7	6.0	5.9	5.9
percentage of repeaters	0.10	0.10	0.14	0.11	0.10	0.08	0.13	0.10	0.09	0.07	0.08	0.16	0.13
non-native speak.>10%(b)	0.02	0.27	0.00	0.00	0.00	0.05	0.00	0.00	0.07	0.00	0.00	0.08	0.00
auton. resources (ind.)	-0.44	-0.64	-0.57	-0.62	-0.40	-0.34	-0.62	-0.47	-0.40	-0.51	-0.13	-0.58	-0.58
auton. curric./ assess.(ind.)	-0.93	-1.05	-0.96	-1.09	-1.05	-1.01	-0.94	-0.88	-0.96	-0.85	-0.90	-0.98	-0.97
private school(b)	0.14	0.01	0.04	0.21	0.13	0.20	0.04	0.12	0.08	0.18	0.23	0.00	0.10
student record(b)	0.16	0.00	0.16	0.09	0.14	0.00	0.04	0.34	0.30	0.09	0.27	0.09	0.06
parental pressure(b)	0.13	0.00	0.06	0.00	0.00	0.09	0.00	0.27	0.00	0.15	0.19	0.08	0.12
school competition(b)	0.79	0.73	0.68	0.62	0.76	0.90	0.25	0.93	0.72	0.83	0.78	1.00	0.53
perc. comp. with web	0.95	0.90	0.87	0.98	0.96	0.94	1.00	0.97	1.00	0.95	0.94	0.92	0.96
comp. - school size ratio	0.56	0.52	0.57	0.79	0.60	0.43	1.03	0.57	0.72	0.53	0.46	0.54	0.60
extra-curric. activ. (ind.)	0.29	0.20	-0.33	-0.32	0.16	0.50	0.44	0.09	-0.49	0.52	0.51	0.28	0.11
educ.resources sch. (ind.)	-0.17	-0.26	-0.07	0.01	-0.04	-0.39	-0.39	-0.08	-0.32	-0.13	-0.04	-0.45	-0.26
teacher particip. (ind.)	-0.78	-0.82	-0.61	-0.94	-1.00	-0.85	-0.39	-0.72	-0.98	-0.83	-0.73	-0.69	-0.74
teacher shortage (ind.)	-0.80	-0.77	-0.41	-0.93	-0.68	-0.91	-0.96	-0.71	-0.77	-0.82	-0.80	-1.02	-0.82
teacher behav. (ind.)	0.13	0.08	0.02	0.20	-0.16	0.00	-0.55	-0.11	-0.05	0.48	0.60	-0.16	-0.05
perc. full-time teachers	0.87	0.86	0.88	0.81	0.81	0.89	0.85	0.88	0.77	0.89	0.84	0.94	0.87
leadership (index)	-0.15	-0.42	-0.15	0.12	-0.13	0.11	-0.26	0.05	-0.43	-0.18	-0.09	-0.65	-0.25

Appendix 2 (continuation)

DESCRIPTIVE STATISTICS (AVERAGES)

School variables	Portugal	Algarve	Alto Alentejo	Baixo Alentejo	Centro Interior	Centro Litoral	Ilhas	Lisboa	Norte Interior	Norte Litoral	Porto	Setubal	Vale do Tejo
student behav. (ind.)	0.03	-0.43	-0.25	-0.11	-0.15	-0.18	-0.68	-0.16	0.04	0.47	0.36	-0.42	0.07
teac. monitor.: tests(b)	0.51	0.31	0.38	0.75	0.38	0.30	0.71	0.73	0.42	0.54	0.40	0.53	0.50
teac. monitor.: peers(b)	0.80	0.85	0.63	0.87	0.73	0.90	0.77	0.89	0.51	0.69	0.85	0.89	0.81
teac. monitor.: sen. staff(b)	0.22	0.21	0.36	0.12	0.23	0.30	0.36	0.27	0.07	0.13	0.14	0.34	0.28
teac. monitor.: external(b)	0.04	0.05	0.00	0.00	0.09	0.08	0.32	0.00	0.00	0.02	0.04	0.00	0.03
class size (students)	22.3	21.2	19.6	19.8	19.7	22.4	19.3	23.2	19.6	23.0	24.0	22.8	21.5
student-teacher ratio	8.5	7.6	7.4	7.6	6.7	8.6	7.0	9.0	7.0	9.6	9.3	8.1	7.5
reg. lessons math. (hours)	4.4	4.1	4.3	4.8	4.6	4.3	5.0	4.8	4.1	4.0	4.5	4.4	4.4
reg. lessons lang. (hours)	3.8	3.6	3.8	4.5	4.0	3.5	4.7	4.0	3.4	3.5	3.8	3.7	3.9

Source: PISA database.
Note: (b) stands for binary variable.

Regional variables	Portugal	Algarve	Alto Alentejo	Baixo Alentejo	Centro Interior	Centro Litoral	Ilhas	Lisboa	Norte Interior	Norte Litoral	Porto	Setubal	Vale do Tejo
GDPpc - 2008	15647	15883	13299	18626	10959	15089	17653	25353	10799	10946	15726	11459	13581
reg. develop. ind. - 2010	100.0	97.0	98.4	94.4	96.6	99.5	93.4	109.8	94.8	97.7	99.8	98.7	96.7
illiteracy rate (%) - 2011	5.2	5.4	10.0	11.3	8.2	6.4	5.8	3.0	9.5	5.3	3.1	3.9	6.4
drop-out rate (%) - 2001	2.8	2.0	2.6	2.8	2.4	1.9	3.8	1.8	4.3	3.8	2.6	2.0	2.6
comp. educ. (%) - 2001	38.0	39.0	31.7	28.7	27.3	36.6	32.2	53.9	26.6	27.1	43.4	48.0	33.3
higher educ. (%) - 2001	8.6	7.0	6.1	4.9	5.3	8.4	6.4	15.1	5.8	4.9	10.8	8.9	5.9
pre-school (%) - 2007/2008	78.3	78.0	92.2	98.4	97.4	85.5	83.3	75.3	94.1	74.0	69.2	58.1	88.2
divorce rate - 2010	17.2	15.8	18.0	16.9	16.6	17.1	15.3	18.8	18.2	17.0	17.8	17.1	17.2
doctors (per 1000) - 2010	3.9	3.0	2.8	1.7	2.0	5.5	2.6	6.6	2.2	1.9	6.9	2.4	1.6
teacher exp. (years) - 2005	2.6	3.0	2.4	2.0	1.9	2.6	2.8	2.8	2.2	2.2	3.1	3.0	2.6
crime rate (%) - 2011	39.4	57.3	27.8	28.2	27.1	36.0	35.0	48.1	32.0	29.9	38.8	43.2	36.5

Sources: INE for all variables except teacher experience that was computed from data in database *Recursos Humanos da Administração Pública 2005*.

Appendix 3 (continue)

REGRESSION (1) – THE ROLE OF STUDENT AND FAMILY VARIABLES		
	Mathematics	Reading
repeater(b)	-53.1	-44.6
	[3.8]***	[4.2]***
female(b)	-28.3	22.1
	[1.8]***	[1.8]***
9th grade(b)	50.0	44.9
	[4.2]***	[3.4]***
10th grade(b)	74.9	71.0
	[5.8]***	[6.0]***
wealth (index)	0.7	-2.2
	[1.4]	[1.3]*
educat. resources home (index)	5.1	2.8
	[1.5]***	[1.0]***
books at home 25-200(b)	15.3	12.1
	[2.6]***	[2.3]***
books at home > 200(b)	31.3	25.4
	[3.4]***	[3.2]****
immigrant status(b)	-12.3	-11.5
	[5.2]**	[4.5]**
foreign language at home(b)	16.9	1.0
	[9.1]*	[8.5]
blue collar/highly skilled(b)	-2.3	-6.6
	[4.6]	[4.2]
white collar/ lowly skilled(b)	0.5	2.4
	[4.2]	[3.6]
white collar/highly skilled(b)	10.6	11.7
	[5.1]**	[4.2]***
lower secondary education(b)	0.5	4.6
	[3.1]	[2.9]
upper secondary education(b)	7.2	6.9
	[3.1]**	[3.1]**
tertiary education(b)	12.3	6.5
	[3.1]***	[3.4]*
one parent home(b)	14.3	12.5
	[3.7]***	[3.2]***
no parents home(b)	-2.5	-11.2
	[6.7]	[6.6]*
Observations	5913	5913
R-squared	0.56	0.55

Source: Authors' calculations.
Notes: (b) stands for binary variable. Computed on the basis of the 5 p-values for test scores. Standard errors in brackets. The regressions include also school-fixed effects which are not shown. * significant at 10%; ** significant at 5%; *** significant at 1%.

Appendix 3 (continue)

REGRESSION RESULTS

	Mathematics	Reading
percentage of computers with web	8.4	1.1
	[16.3]	[14.4]
computers - school size ratio	10.3	0.5
	[6.9]	[6.1]
percentage of girls(b)	0.1	0.4
	[0.3]	[0.3]
school size (1000 students)	10.2	11.1
	[5.6]*	[4.9]**
class size (students)	2.5	1.3
	[0.8]***	[0.7]*
student-teacher ratio	-1.1	-1.1
	[1.2]	[1.0]
private school(b)	-16.1	-7.0
	[10.9]	[9.5]
extra-curricular activ. (index)	2.1	5.8
	[2.3]	[2.0]***
educ. resources school (index)	4.5	3.9
	[2.5]*	[2.2]*
teacher participation (index)	7.4	4.3
	[3.8]**	[3.3]
teacher shortage (index)	6.2	5.8
	[4.0]	[3.6]
teacher behaviour (index)	-0.2	-0.2
	[2.8]	[2.4]
parental pressure(b)	9.3	8.8
	[5.4]*	[4.7]*
located town 15-100 inh.(b)	1.6	1.7
	[4.3]	[3.8]
located town > 100 ihn.(b)	9.3	10.4
	[5.9]	[5.2]**
school competition(b)	-6.1	0.6
	[5.1]	[4.4]
percentage of full-time teachers	21.1	23.1
	[17.6]	[15.5]
regular lessons (hours)	0.8	1.3
	[3.3]	[2.6]
leadership (index)	1.6	1.8
	[2.8]	[2.5]
student behaviour (index)	3.8	2.9
	[2.6]	[2.3]
teacher monitoring: tests(b)	0.8	-4.1
	[4.0]	[3.5]
teacher monitoring: peers(b)	11.9	7.9
	[5.0]**	[4.3]*
teacher monitoring: senior staff(b)	-1.8	-2.8
	[4.9]	[4.3]
teacher monitoring: external(b)	-6.3	-3.9
	[14.2]	[12.5]
autonomy resources (index)	7.7	3.3
	[4.0]*	[3.5]
autonomy. curricula/ assess. (index)	-5.0	7.9
	[7.1]	[6.2]

Appendix 3 (continuation)

REGRESSION (2) – THE ROLE OF SCHOOL VARIABLES		
	Mathematics	Reading
non-native speakers > 10 %(b)	-1.2	-17.4
	[10.9]	[9.6]*
student record consideration(b)	9.2	7.0
	[5.4]*	[4.7]
percentage of repeaters	-19.7	-6.4
	[28.7]	[24.9]
grade amplitude (max-min grade)	0.9	-0.5
	[0.9]	[0.8]
Constant	337.4	365.3
	[33.7]***	[30.0]***
Observations	209	209
R-squared	0.42	0.44
F- Test (all school variables)	2.72	3.04
p-value	0.00	0.00

Notes: (b) stands for binary variable. Standard errors in brackets. The regressions include also regional-fixed effects which are not shown.* significant at 10%; ** significant at 5%; *** significant at 1%.