AN ANALYSIS OF PORTUGUESE STUDENTS' PERFORMANCE IN THE OECD PROGRAMME FOR INTERNATIONAL STUDENT ASSESSMENT (PISA)*

*Manuel Coutinho Pereira***

ABSTRACT

This study focuses on the evolution of the scores of Portuguese students in *PISA* cycles from 2003 to 2009. It may be concluded that the variation in scores is strongly influenced by changes in the family background of students and other variables, such as the distribution of students by degrees. The sampling methods used for data collection under the programme tend to amplify these changes. When such factors are taken into account for the analysis, holding constant the score determinants, there has been a continued improvement in student performance over the cycles considered.

1. Introduction

International programmes of educational achievement, such as the OECD *PISA*, provide comparable data over time and between countries which are highly valuable for the evaluation of educational systems and, implicitly, the return on education spending. *PISA* 2009 results, corresponding to the fourth time this programme was administered, were released in December last year and yielded an improvement in the scores of Portuguese students compared to previous editions (which took place in 2000, 2003 and 2006, *i.e*. in cycles of three years). Further analysis of trends in scores, however, requires a confrontation with the evolution of the characteristics of the student population and schools. Firstly, students' socio-economic status has an influence on performance, and any change in status over the editions of the programme should be taken into account. As can be seen, there are other aspects to consider in this context, such as the distribution by grade of the children covered by the programme. *PISA* is a sample survey in which inference is drawn by extrapolation to the population. This appears to magnify the differences between cycles for some student and school variables, and makes an analysis such as this all the more necessary. In contrast, in the presentation of trends in *PISA* results, as in OECD (2010), an unconditional analysis has been favoured (see Gebhardt and Adams, 2007).

This study investigates the change in the scores of Portuguese students throughout the *PISA* surveys, at various points of score distribution, taking into account the changes in the observable determinants. The outcomes for two of the subjects in the programme, mathematics and reading, are examined. This work follows on from Pereira (2010) who analysed the explanatory factors behind Portuguese students' performance in *PISA* 2006 in the European context – assessed from the estimation of education production functions – along with a set of results concerning its variability. This analysis was intended to establish a number of facts of a structural nature, for which no substantial change is expected over the time

* The author thanks Nuno Alves, Maria Manuel Campos, 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 author and do not necessarily coincide with those of Banco de Portugal or the Eurosystem. Any errors and omissions are his sole responsibility.

** Banco de Portugal, Economics and Research Department.

frame covered by the *PISA* cycles, making a comparison with other countries.[1] In contrast, this study focuses on the changes in performance of Portuguese students over the editions of the programme.

When the aim is to compare the outcomes of evaluation tests conducted at different times, as in the research carried out herein, it is essential that the measurement of performance be comparable. If the difficulty of tests varies and the scores are not correspondingly adjusted, the assessment of the change in performance – which is the objective of the analysis – may be jeopardized. In *PISA* care has been taken to ensure the comparability of results by reporting scores in different surveys according to the same scale. This is achieved through the linkage of the assessments for each subject by a set of common items. The degree of difficulty measured for those items is evaluated in each cycle *vis-à-vis* a reference cycle,[2] and any inequality found is used in the construction of a transformation of scores to the scale of the reference cycle (see OECD 2009a, Chapter 12, and Gebhardt and Adams, 2007, for a critical discussion).

This procedure has been followed since *PISA* 2000 and *PISA* 2003 (taken as reference cycles), respectively, for reading and mathematics. The non-comparability of scores in mathematics in the 2000 edition implies their exclusion from this study. It was decided not to consider any data from this cycle at all, including the ones for reading, because the presentation of information about students and schools differs for certain variables relative to the subsequent cycles. Given that, for conditional inference, variables must be available (or constructed according to a common methodology) for each year, the exclusion of *PISA* 2000 also makes it possible to keep a greater number of explanatory variables.

The study begins with a descriptive analysis of how Portuguese students' performance and student and school variables have evolved from *PISA* 2003 to 2009 (Sections 2 and 3). Section 4 sets out a decomposition of scores between the part explained by the change in pupil and school characteristics, and the inequality in scores that would have prevailed, had such characteristics remained identical from one edition to the other. This last component gives a measure of the variation in performance that may be attributed to the educational system. Finally, in Section 5, a detailed analysis is provided for outcomes in public and private schools. The main conclusions are summarized in Section 6.

2. Performance of Portuguese students from *PISA* 2003 to 2009

The population of students in *PISA* consists of 15-year-old students who attend schools in one country and are at least in the 7th grade. In the Portuguese case, most of the students are in the 9th or the 10th grade. The tests are taken by a representative sample from this population. In the sampling process schools are randomly selected in a first stage, and eligible students in each of these, up to a maximum of 40, in a second stage. In Portugal participated in the program 4608 pupils belonging to 153 schools in 2003, 5109 pupils belonging to 173 schools in 2006, and 6298 pupils belonging to 214 schools in 2009. The sample size represented about 5 per cent of the relevant student population. The *PISA* databases include final student weights, reflecting, *inter alia*, sampling probabilities. In addition, scores are reported in the form of values extracted from the estimated distribution of scores assigned to each student (see OECD, 2009b, Chapters 6 and 8).

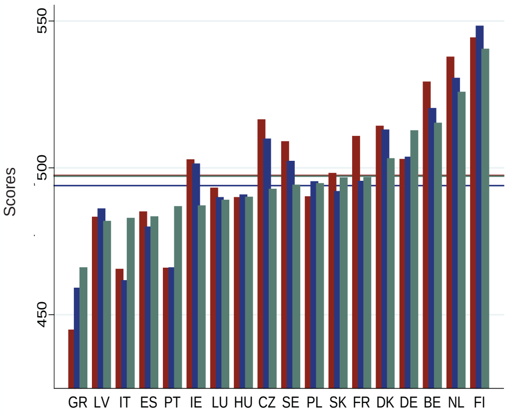
Charts 1A and 1B show the average scores in Portugal and the European Union countries participating in the three editions of the programme and whose data in each of these meet certain quality requirements defined by the OECD (for example, regarding the response rate – see OECD, 2010, Chapter

1 See also Pereira (2010) for a discussion on how empirical studies such as the one presented herein fit in the framework of economics of education literature.

2 Note that *PISA* uses the Rasch model, in which a question's difficulty is measured by the proportion of students who answer it correctly; each question is then associated with a point on the scale according to its degree of difficulty. Finally, the student is placed on the point of the scale corresponding to the question to which he/she has a 50 per cent probability of responding correctly.

Chart 1A

PERFORMANCE IN MATHEMATICS | SCORES BY COUNTRY AND OVERALL MEAN, IN 2003 (IN RED), IN 2006 (IN BLUE) AND IN 2009 (IN GREEN)

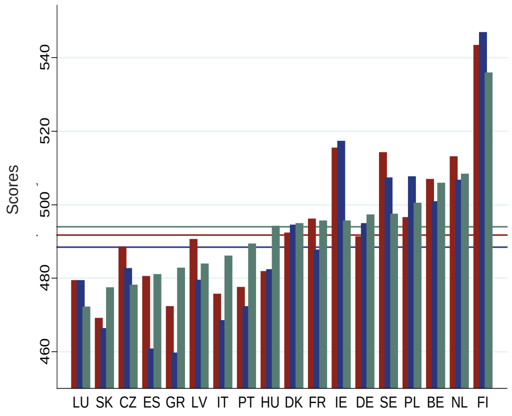


Source: Author's calculations.

Note: Average of the weighted averages for each plausible value.

Chart 1B

PERFORMANCE IN READING | SCORES BY COUNTRY AND OVERALL MEAN, IN 2003 (IN RED), IN 2006 (IN BLUE) AND IN 2009 (IN GREEN)



Source: Author's calculations.

Note: Average of the weighted averages for each plausible value.

1).[3] It also presents the mean for all countries (horizontal lines). It should be remembered that scores are measured by reference to the results for 2000, in the case of reading, and for 2003, in the case of mathematics (with the value 500 corresponding to the respective OECD mean). The countries are ordered according to the *PISA* 2009 results.

The main conclusion to be drawn from charts 1A and 1B is that the mean score for Portugal in the most recent *PISA* was higher than in the two previous editions, in which the results had been fairly close. This trend is particularly visible in reading, placing Portuguese students in an intermediate position in the ranking of EU countries considered. More importantly, the average score does not differ significantly, in statistical terms, from the average in that group of countries.[4] There was a noticeable improvement in performance in mathematics as well. While Portugal continues poorly positioned in terms of the ranking shown in chart 1A, the country clearly caught up with the countries occupying intermediate positions.

In order to complement the picture of score evolution between 2003 and 2009, charts 2A and 2B present the proportion of students in lower and upper score cohorts, respectively, at proficiency level 1 and below and at proficiency level 5 and above. These levels of proficiency, defined under the programme, correspond to increasing levels of difficulty in the questions students must answer (see footnote 2). In particular, students in the lower cohort are deemed to have acquired skills below the minimum level making productive participation in society possible.

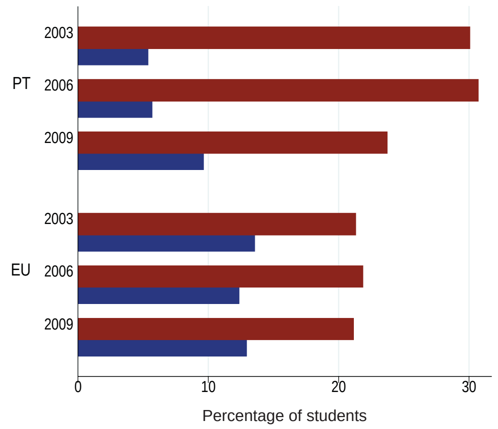
With regard to the proportion of Portuguese students in the lower cohort, a certain increase between 2003 and 2006 was followed by significant declines in 2009, both in reading (a subject for which that proportion became lower than the EU average) and mathematics. In the latter subject there was, at the same time, a significant rise in the proportion of students with very high scores, causing a shift to the right of the score distribution as a whole. In contrast, in reading, the average increase was due to

3 This latter criterion leads to the exclusion of Austria and the United Kingdom from this study. The countries considered are Belgium, Czech Republic, Denmark, Germany, Finland, France, Greece, Hungary, Ireland, Italy, Latvia, Luxembourg, Netherlands, Poland, Portugal, Slovakia, Spain, and Sweden.

4 That is, the 95 per cent confidence intervals (not shown) around the mean score for the population in Portugal and in EU countries intersect.

Chart 2A

PROFICIENCY IN MATHEMATICS IN PORTUGAL AND EUROPEAN UNION COUNTRIES | STUDENTS AT LEVEL 1 AND BELOW (IN RED) AND AT LEVEL 5 AND ABOVE (IN BLUE)

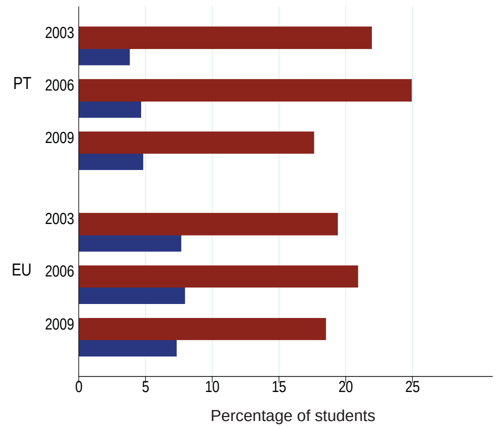


Source: Author's calculations.

Note: Average of percentages for each plausible value.

Chart 2B

PROFICIENCY IN READING IN PORTUGAL AND EUROPEAN UNION COUNTRIES | STUDENTS AT LEVEL 1 AND BELOW (IN RED) AND AT LEVEL 5 AND ABOVE (IN BLUE)



Source: Author's calculations.

Note: Average of percentages for each plausible value.

the better marks of students featuring lower performance levels. The evolution in the quartiles of score distribution for Portugal from 2006 to 2009 (not shown) confirms these findings. Whereas in mathematics there was a similar shift (of around 20 to 21 points) to the right of all quartiles, in reading such displacement was equal to about 26 points in the first quartile and less than 10 points in the third. The dispersion of scores in mathematics remained thus unchanged, while that of scores in reading decreased.

3. Student and school data

In order to put together a group of explanatory variables of the greatest possible extent and common to all the three cycles, the available information about students and schools in the databases was carefully examined. The questionnaires from which this information comes have a similar – although not identical – content over the years. Most of the variables used were directly taken from the databases. However, in some cases these variables were constructed from basic information, namely, the amplitude in the grades offered by school, and the indices of educational resources at home, autonomy in resource allocation and autonomy of curriculum and assessment[5] (see Appendix 1). The set of explanatory variables available – presented in table 1 – covers most of those considered in Pereira (2010). The exceptions are the wealth index (which is largely redundant in that the family background is well captured by other variables) and the variables related to availability of computers, parental pressure on schools to improve standards, the existence of other schools competing for the same students, and students' familiarity with information technologies.[6]

5 The original indices, available in the databases, were not used since it was found that they had not been constructed in a uniform way over time.

6 In addition, the binary variables for the shortage of mathematics and test language teachers were not considered, because, in *PISA* 2006 and 2009, very few Portuguese schools reported the existence of such a shortage.

The most significant changes in the characteristics of the student population, their families and schools for Portugal in the 2003, 2006 and 2009 editions of *PISA* (Table 1)[7] are now considered. As mentioned, students are mainly distributed between the 9**th** and the 10**th** grades, reflecting primarily the month of birth (for those who have never repeated a grade). However, there has been a considerable fluctuation in the distribution between these two grades, with a particularly low number of students in the 10th in *PISA* 2006. As discussed in the next section, the allocation of students to grades greatly influences scores and is therefore of relevance. The sampling process may contribute to it to the extent that there have been departures from representativeness for certain types of schools in the sample, namely those that provide only lower secondary education – courses up to the 9th grade – or upper secondary education – courses from the 10th grade onwards (recall that the sampling process begins with a selection of schools). The fact that the proportion of schools of the first type was higher in *PISA* 2006 than in other cycles indicates that this may have been the case. It is known that such a problem can be mitigated by including the type of school as a sample stratification criterion; this was only the case in *PISA* 2006 (see Table 4.1, Chapter 4 – OECD, 2005, 2009a, 2011). However, even for the schools that offer both lower and upper secondary education courses, there has been some variation over time in the distribution of students by the two grades.[8]

The figures for the variables measuring family context have fluctuated as well. The situation in terms of the educational attainment of parents is more favourable in *PISA* 2009 than in the 2006 cycle, featuring a larger proportion of students whose parents have secondary education as their highest educational level. In particular, there was a large increase in the lower secondary education (9**th** grade) cohort, probably mostly related to qualifications obtained under the *Processos de Reconhecimento, Validação e Certificação de Competências*. In the 2009 cycle, the proportion of students with at least one parent having a tertiary degree also increased. The situation in *PISA* 2006 *vis-à-vis* 2003 in terms of educational attainment of parents was similarly unfavourable, this time with regard to the distribution between tertiary and upper secondary education cohorts. Associated with the variation in qualifications, there has been a fluctuation in the breakdown of parental occupations. For example, the proportion of students with at least one parent in a white collar/highly skilled occupation fell from about 34 to 26 per cent between *PISA* 2003 and 2006, increasing to 36 per cent in *PISA* 2009. Such differences between cycles of the programme with regard to family background variables may be linked to a sample bias towards the selection of schools in rural areas (towns with less than 15 000 inhabitants) in *PISA* 2006 to the detriment of schools located in medium-sized urban areas (towns with between 15 000 and 100 000 inhabitants). In general, one would expect an improvement in parental educational attainment during the period under review, featuring a gradual increase in the number of parents with at least a secondary rather than a primary degree. However, such a trend may, in practice, be obscured by the "noise" introduced by the sampling process.

With regard to school variables, there was a decrease in the proportion of repeaters (in the school as a whole) over the three considered *PISA* cycles. This development is in line with the observed decrease in repetition rates at various educational levels during the last decade (GEPE, 2010). The trend in the indicators of autonomy, compiled from a set of questions answered by schools (see Appendix 1), indicates a decrease in the autonomy of the latter in the choice of curricula and assessment methods. The proportion of the student population attending private schools has increased over time, from just

7 Most variables in the table have a few missing observations for each year. Such observations were imputed by running a regression (for the countries listed in footnote 3) of the variables in question on a set of "key regressors" including the degree, age, gender, school location and country (in the same way as described in detail in Pereira, 2010, Appendix 2). All of the observations for normal hours of the test language in *PISA* 2003 were imputed. In this case the imputation was based on normal hours of mathematics and binary variables for the school location and country.

8 Notwithstanding the school type, it is possible that, given the unequal proportions of students attending the 9th and 10th grades, the sample size does not permit greater accuracy (more so in the case of the 7th and 8th grades). The effective sample size per school is, on average, 32 students (maximum 40).

Table 1

STUDENT AND SCHOOL VARIABLES, AVERAGES FOR PORTUGAL[(a)]	2003	2006	2009
Student characteristics			
7th grade	4.2	6.6	2.3
8th grade	10.6	13.1	9.0
9th grade	20.3	29.5	27.9
10th grade	64.9	50.9	60.8
Age (years)	15.9	15.7	15.7
Female	52.4	51.7	51.1
Family background			
Educational resourc. home (index) ⊂ [0,6][(b)]	4.7	5.3	5.0
Books at home < 25	35.1	38.7	36.4
Books at home 25-200	49.1	46.0	48.1
Books at home > 200	15.8	15.4	15.5
Native	95.0	94.1	94.6
Immigrant (1st or 2nd generation)	5.0	5.9	5.4
Test language at home	98.6	97.8	98.4
Foreign language at home	1.4	2.2	1.6
Parents' highest occuppat. level			
Blue collar/low skilled	12.9	12.9	8.9
Blue collar/high skilled	27.9	24.0	21.9
White collar/low skilled	25.3	35.2	33.0
White collar/high skilled	33.9	27.8	36.2
Parents' highest education level			
Primary or less	38.5	38.1	27.0
Lower secondary	16.8	16.2	23.0
Upper secondary	19.4	23.5	24.3
Tertiary	25.3	22.2	25.8
School characteristics			
School size (1000 students)	1.000	0.958	0.937
Proportion of girls	51.5	50.8	50.5
Located in town with less 15 000 people	37.6	42.5	36.2
Located in town with 15 000-100 000 people	42.9	35.9	42.2
Located in city with more 100 000 people	19.5	21.6	21.6
Grade amplitude (max - min grade)	4.7	5.1	5.7
Proportion of repeaters	17.0	14.6	9.8
Autonomy resources (index) ⊂ [0,6]	1.9	1.7	2.2
Autonomy curric./assessm. ⊂ [0,4]	2.4	2.0	1.5
Public school	93.8	91.2	86.2
Private school	6.2	8.8	13.8
School resources			
Class size (students)	22.0	24.0	22.3
Student/teacher ratio	11.0	8.9	8.5
Regular lessons math (hours)	3.2	3.5	4.4
Regular lessons language (hours)[(c)]	3.1	3.2	3.8

Source:

Notes: (a) Weighted averages; figures as a percentage of totals unless otherwise stated (more details about the construction of variables are given in Appendix 1). **(b)** For indices, intervals show the minimum and maximum. **(c)** Figures for 2003 were fully imputed (see footnote 7).

over 6 per cent, in *PISA* 2003, to almost 14 per cent, in *PISA* 2009 (the public/private nature of school has been used as a stratification criterion throughout the various editions). This profile is, however, not corroborated by the figures from other sources (GIASE, 2006 and GEPE, 2011).[9] Also in this case, it may not be possible to achieve greater accuracy given the sample size and the fact that private schools represent a small proportion of the universe of students.

In conclusion, the characteristics inferred for the student population and schools have varied throughout the *PISA* cycles, and the extrapolation from the sample to the population tends to amplify the magnitude of such variation. In this context it is very important to determine their impact on the evolution of student performance.

4. Decomposition of change in performance

4.1. Computation and interpretation of the decomposition

In general, the evolution of a variable explained by a linear regression model can be decomposed into a component relating to the explanatory variables, on the one hand, and to the coefficients associated with them, on the other (see Fortin *et al*., 2011, for a description of the methods used in this context). The linear model that underpins the decompositions performed in this study is the education production function that relates test scores to explanatory variables such as student, socioeconomic and school variables. The change in the dependent variable at its average is traditionally analysed through the Oaxaca-Blinder decomposition, which is based on the estimation of the underlying model by least squares regression. This method makes it possible to differentiate between the contribution of coefficients and that of explanatory variables while, at the same time, directly providing a detailed breakdown of the latter contribution by variable (or sets of variables). This aspect is important in our context, as the regressors are naturally divided up into groups whose contribution should be considered jointly. Three groups of variables are considered in the presentation of results, namely, student characteristics, measures of family context and school characteristics/resources (see Table 1 for the listing). Based on unconditional quantile regressions, developed by Fortin *et al*. (2009), it is possible to perform a similar decomposition at other points of the dependent variable distribution.[10]

The decomposition divides the differential in performance between *PISA* cycles into two terms. The first term is the part that can be attributed to changes in the variables included in the education production function, *i.e.* the characteristics of students, families and schools in each cycle. The second term reflects the changes in the return on the variables, *i.e.* the differential in performance that would prevail, if these variables had remained unchanged from one cycle to the other. The differential in conditional performance, which this second term captures, can be interpreted as originating in the educational system. Note that for the first term, the part concerning the school-related regressors[11] admits a similar interpretation (see also the discussion in the following paragraph about omitted variables). The objec-

9 Which indicate (considering all the students who attend regular courses in the third basic education cycle) that the proportion of students in private schools rose marginally from 12 to 13 per cent over the concerned period.

10 The expressions used to calculate the decompositions are given in the note to table 2 below. The Oaxaca-Blinder decomposition is based on the fact that the least squares estimator of a linear model $y = x\beta + u$ yields the impact (equal to β) on the unconditional expected value of y, $E(y)$, of the variation of $E(x)$, as given by $E(y) = E(E(y \mid x)) = E(x)\beta$. Similarly, the unconditional quantile regression estimates the impact (say, γ) on the unconditional quantile of y, $Q(y)$, of the variation of $E(x)$, *i.e.* $Q(y) = E(x)\gamma$. Note that this property is not shared by the conventional conditional quantile regression of Koenker and Bassett (1978) since, in general, $Q(y) \neq E(Q(y \mid x))$. Hence the decompositions based on these latter regressions require the simulation of counterfactual distributions which, in particular, makes it difficult to obtain a detailed breakdown of the contribution of regressors (see Fortin *et al*., 2011).

11 Except for the indicator of school location.

tive is to eliminate the influence of factors related to family background and data collection, notably the distribution of students by grade, included in student characteristics. While such distribution may be endogenous to the educational system, as it relates to grade repetition, in the data used here such an effect is unlikely to predominate.

One important aspect to take into account when interpreting the coefficient-related component of the decomposition is that the variation in the constant term coefficient will, *inter alia*, pick up the effects of changes in the level of omitted variables.[12] Regressions explaining student assessment outcomes include several statistically significant regressors, but typically fail to explain all of their variability (see, for example, Woessmann *et al*., 2009, Chapter 2, using the *PISA* dataset for a wide range of countries). The coefficient of determination indicates that in the least squares regressions for Portugal – on which the Oaxaca-Blinder decomposition is based – about half of the variance of scores remains unexplained. This should reflect, firstly, the variability in student capabilities, but such factor is expected to remain constant over time and therefore not to greatly affect the decomposition results. The same does not hold, however, for the other unobservable factors that relate to the quality and effectiveness of teaching, such as the role of teachers in the organization of classes and choice of teaching methods. This is probably the worst covered area in the *PISA* database, where there are not, for example, measures of teacher experience.[13] But even if the change in the coefficients is also capturing changes in this type of variable throughout the *PISA* cycles, this is still is consistent with the interpretation of the component at issue as referring to variations in performance attributable to the educational system.

4.2. Results

The decompositions of the variation in mathematics and reading scores at the mean and first and third quartiles are presented in tables 2 and 3, respectively, for 2003-2006 and 2006-2009. It is possible to calculate a detailed decomposition of the coefficient-related contribution corresponding to that of explanatory variables. However, in the presence of binary variables such as for the categories of parental occupations and qualifications, the results are not invariant to the category omitted in the regression (see, for example, Oaxaca and Ransom, 1999). In practice, this invalidates the interpretation of these results, which are therefore not shown.

The approximate stabilization of the average performance of Portuguese students between *PISA* 2003 and 2006 in mathematics and reading, presented in charts 1A and 1B at the beginning of this study,[14] stems from contributions of opposite signs of the coefficients (positive) and regressors (negative), which approximately cancelled each other out. This was also the case at the quartiles, except for the first one in reading, where the increase in return on variables was small and there was a clear reduction in the level of scores. Such a reduction is consistent with the higher percentage of students at lower proficiency levels for this subject, shown in chart 2B. In the contribution of regressors, the most important part is played by student variables, in line with the rise in the proportion of students attending the 9th grade in *PISA* 2006, as well as lower grades – as indicated by the even more negative contribution of student characteristics in the first quartile of score distribution.

The contribution of family variables is almost nil at the mean, and turns into negative at the third quartile, which may have to do with the decline in the proportion of parents with a tertiary degree and

12 The coefficients of the other regressors included in the model will also change to the extent that there is correlation with omitted variables. However, this poses no difficulties in our context, because in the decomposition the coefficients are considered as a whole.

13 The databases include teacher qualification variables, which, however, given their small variability, are of little interest to the analysis.

14 The values in the tables (for the total) differ slightly from those underlying the charts since in the calculation of the latter all observations are used, unlike for the regressions. Indeed, even after the imputation procedure, some missing observations for the explanatory variables remain.

Table 2

DECOMPOSITION OF CHANGE IN SCORES 2003-06 IN THE MEAN AND QUARTILES

	1st Quartile		Mean		3rd Quartile	
	Mathematics	Reading	Mathematics	Reading	Mathematics	Reading
Covariates (1)	-7.1	-7.3	-5.9	-6.1	-5.5	-7.5
	(-11.5,-2.5)	(-12.5,-1.6)	(-9.2,-2.5)	(-9.7,-2.4)	(-10.1,-1.7)	(-12.3,-3.8)
Student	-11.2	-12.6	-10.1	-10.2	-9.7	-9.7
	(-13.7,-8.9)	(-15.4,-9.7)	(-12.0,-8.2)	(-12.3,-8.1)	(-11.9,-7.5)	(-11.9,-7.8)
Family	2.2	3.1	-0.1	0.8	-3.5	-2.9
	(0.4,4.1)	(0.8,5.4)	(-1.4,1.1)	(-0.7,2.2)	(-5.1,-1.9)	(-4.8,-1.3)
School	2.0	2.2	4.3	3.3	7.7	5.1
	(-0.9,5.2)	(-1.5,6.0)	(2.3,6.3)	(1.2,5.7)	(4.2,10.5)	(1.9,8.0)
Coefficients (2)	8.0	1.0	8.1	3.3	10.5	8.0
	(3.1,12.4)	(-4.9,6.3)	(4.8,11.2)	(-0.4,6.6)	(6.3,15.8)	(3.8,12.9)
Total (1+2)	0.9	-6.3	2.2	-2.8	5.0	0.5
	(-3.6,5.2)	(-11.4,-1.9)	(-1.1,5.6)	(-6.3,0.7)	(1.3,9.3)	(-3.1,4.1)

Source: Author's calculations.

Notes: The decompositions are computed as $S(y_{t1})\text{-}S(y_{t0})=(X_{t1}\text{-}X_{t0})b_{t1}+X_{t0}(b_{t1}\text{-}b_{t0})$, where $t1$ and $t0$ index the year, $S(y_{t.})$ are the relevant statistics of test scores, $X_{t.}$ are the averages of the covariates (see Table 1) and $b_{t.}$ are the coefficients obtained by ordinary least squares regressions, for the mean, and unconditional quantile regressions (Fortin *et al.*, 2009), for the quartiles. The regressions are weighted, using the final student weights, and run separately for each plausible value. Bootstrap 95% confidence intervals, on the basis of 1000 replications, in parenthesis.

Table 3

DECOMPOSITION OF CHANGE IN SCORES 2006-09 IN THE MEAN AND QUARTILES

	1st Quartile		Mean		3rd Quartile	
	Mathematics	Reading	Mathematics	Reading	Mathematics	Reading
Covariates (1)	16.9	13.9	15.2	11.7	14.2	6.9
	(12.4,21.2)	(9.3,18.7)	(11.8,18.6)	(8.3,15.0)	(9.4,18.0)	(2.9,10.5)
Student	16.7	16.3	12.0	11.3	7.3	6.4
	(14.0,19.2)	(13.8,19.0)	(10.1,13.7)	(9.5,13.0)	(6.0,8.6)	(5.2,7.5)
Family	2.0	2.6	2.3	2.8	3.0	3.0
	(0.8,3.4)	(1.4,3.9)	(1.1,3.4)	(1.8,3.8)	(1.6,4.7)	(1.7,4.4)
School	-1.9	-4.9	0.9	-2.4	3.8	-2.5
	(-5.2,1.3)	(-8.1,-1.4)	(-1.6,3.4)	(-4.6,-0.2)	(0.1,6.8)	(-5.8,0.7)
Coefficients (2)	12.0	19.3	12.2	11.7	10.9	4.9
	(7.2,16.9)	(13.9,24.4)	(8.8,15.3)	(8.4,15.0)	(6.4,16.3)	(0.3,9.5)
Total (1+2)	28.9	33.2	27.4	23.4	25.1	11.8
	(24.7,32.7)	(28.5,37.8)	(24.3,30.4)	(20.2,26.2)	(21.1,29.0)	(8.1,15.5)

Source: Author's calculations.
Note: See note to Table 2.

a white collar/highly skilled occupation. The variation in performance attributable to school variables is positive, particularly in intermediate and upper score levels, reflecting the changes in various regressors, notably, the increase in the number of hours of regular classes, the reduction in autonomy in the choice of curricula and assessment methods and, for mathematics, the higher proportion of private schools. Table 2 shows a more favourable picture regarding the comparison of results in *PISA* 2003 and 2006 than charts 1A and 1B. Indeed, there is an improvement in the return on variables, i.e. in conditional performance, at most points of score distribution (to which a positive contribution of the school variables adds).

In the evolution of scores between *PISA* 2006 and 2009 (Table 3), both the coefficients and regressors make positive contributions, which thus reinforced each other. Therefore, the improvement in marks associated with the coefficients falls short of the overall figure. As expected, the part of the variation in performance attributable to student characteristics is now positive, its magnitude being particularly large in the lower half of score distribution. The part of that variation relating to the coefficients has a similar profile, and thus the combined effect is a more pronounced improvement in performance in that segment of the distribution – which is in line with the decrease in the percentage of students at lower proficiency levels (Charts 2A and 2B). The contribution of socio-economic variables was consistently positive throughout the score distribution – as a result of the more favourable situation in *PISA* 2009 regarding these variables – but its magnitude is less important than that of student characteristics. The impact of school regressors on the score variation is relatively small in absolute terms. The sign of this impact is not uniform along the distribution of scores in the case of mathematics, while in the case of reading it is negative. In particular, the increase in the proportion of private schools (by 5 p.p.) inferred for the population in 2009, which, as noted, may have to do with the sampling process, has little influence on the variation of performance. Indeed, the coefficient of the indicator of private school for 2009, used in the decomposition, is very small – especially at the mean (see the next section).

With the caveat that the period of time under consideration is not too long, evidence indicates an improvement in conditional performance of students in the last two editions of *PISA*, which can be attributed to the educational system. Among the factors that may, tentatively, be put forward to explain such an evolution, the gradual introduction of national examinations[15] is likely to have played an important role. Economics of education literature advocates that central exams, external to schools, are a very effective way of setting the right incentives for academic success for the various agents. On the one hand, they enhance accountability of schools, teachers and pupils and, at the same time, allow informed decision-making. Empirical multi-country studies that have addressed this topic have found a higher performance level in school systems with central exams, common to the various points of its distribution and family contexts (see, for example, Woessmann, 2002).

5. Performance in public and private schools in PISA

PISA outcomes can be used to make a comparison of scores between public and private schools and, indirectly, assess teaching quality in these institutions. In this context, the programme data have the advantage of being accompanied by information on the socioeconomic status of students, which can be taken into account in the analysis. At the same time, the availability of data for the three cycles makes it possible to examine the issue on a sounder basis.

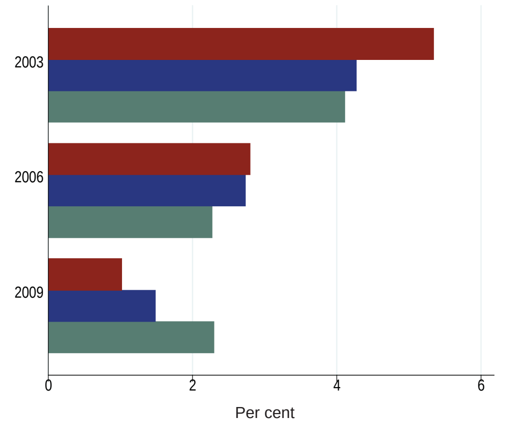
In this analysis attention is restricted to students in the 10th grade (which is the largest group) as a way of ensuring that the conclusions are not affected by a different distribution by grades of students in public and private schools. As a starting point, charts 3A and 3B present the difference for some descriptive statistics of scores between the two types of institutions.

With regard to scores in mathematics, private schools have outperformed their public counterparts throughout the *PISA* cycles, both at the mean and the quartiles. However, the marks in both types of schools have drawn closer over time, primarily owing to the improvement in attainment in public schools. In *PISA* 2009 the differential was already relatively small (around 2 per cent), and on the threshold of statistical significance. This trend has been more marked for lower score levels, and reversed the profile of inequality across the distribution from 2003 to 2009: while the difference in *PISA* 2003 was highest at the bottom of the distribution, this happens at the top in *PISA* 2009. For reading, the differential between private and public schools in the 2006 cycle was negative (but clearly non-significant in statistical terms) and atypical relative to the 2003 and 2009 cycles. One possibility would be a particularly unfavourable sample of private schools in that year (the average score inferred for the population falls *vis-à-vis* 2003), but such a conjecture is not corroborated by the results in mathematics. Since in 2009 the difference is positive but relatively small, it may be concluded that there has been little disparity in reading performance between public and private schools in the two most recent editions of *PISA*.

15 At the end of upper secondary education, in the nineties, and of lower secondary education, since 2005. There have been national exams at the end of the 4th and 6th grades as well, but which have no consequences in terms of student assessment.

Chart 3A

PERFORMANCE IN MATHEMATICS IN PUBLIC AND PRIVATE SCHOOLS (STUDENTS IN THE 10TH GRADE) | SCORES IN PRIVATE SCHOOLS RELATIVE TO PUBLIC, PERCENTAGE DIFFERENCE AT THE 1ST QUARTILE (IN RED), THE MEAN (IN BLUE) AND THE 3RD QUARTILE (IN GREEN)

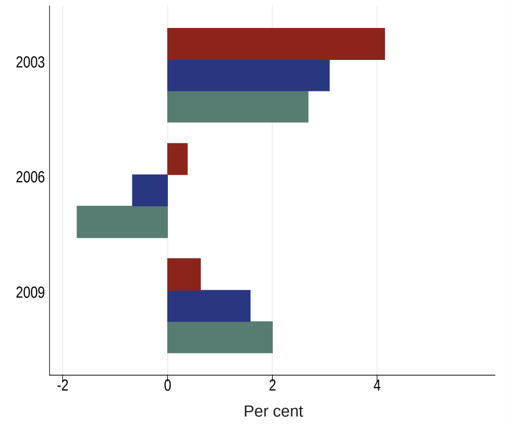


Source: Author's calculations.

Note: Coefficient of the indicator variable for private school in (weighted) least squares and conditional quantile (Koenker e Bassett, 1978) regressions, also including the constant as a regressor.

Chart 3B

PERFORMANCE IN READING IN PUBLIC AND PRIVATE SCHOOLS (STUDENTS IN THE 10TH GRADE) | SCORES IN PRIVATE SCHOOLS RELATIVE TO PUBLIC, PERCENTAGE DIFFERENCE AT THE 1ST QUARTILE (IN RED), THE MEAN (IN BLUE) AND THE 3RD QUARTILE (IN GREEN)



Source: Author's calculations.

Note: Coefficient of the indicator variable for private school in (weighted) least squares and conditional quantile (Koenker e Bassett, 1978) regressions, also including the constant as a regressor.

The differential between private and public schools considering the same statistics for scores, but conditional on family context, is now presented. Charts 4A and 4B show the coefficients of the binary variable for private school in least squares and quantile regressions,[16] also including family background variables and the school location indicator. As expected, the differentials controlling for the socioeconomic composition, more favourable in private schools, are smaller compared to those shown in charts 3A and 3B. Such a reduction is, however, not uniform over the three *PISA* cycles considered, being more substantial in 2003 and 2009 than in 2006. Indeed, the aforementioned composition is more homogeneous between the two groups of schools in this latest edition.

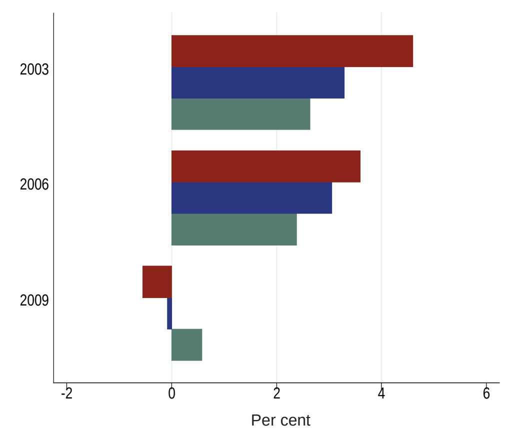
With regard to mathematics scores (Chart 4A), the gap between private and public institutions, after controlling for family background, is similar in the 2003 and 2006 cycles, because the "correction" of that background's influence was more important in 2003. On the other hand, in 2009, the gap virtually disappeared (and became, in addition, statistically not significant). The outcomes in reading (Chart 4B) reinforce the interpretation that the classifications in public and private schools differed little in the last two editions of the programme: the differences are, except for the third quartile in 2006, of a small magnitude and statistically not significant.

In conclusion, some indication of a better performance of private schools in the older editions of *PISA* has been fading. In the 2009 cycle, in particular, there was no relevant difference between scores in public and private schools, whether in mathematics or reading, controlling for the socio-economic status of students. A qualification applying to the whole analysis relates to the fact that the sample contains a small number of private schools, and the findings may be disturbed by non-representativeness of those selected. It may thus be problematic to extrapolate to the universe of students on this matter. In addition, the outcome of national 9th grade exams indicates, even in recent years, larger differences between the performance of students in the two types of schools than those presented in charts 3A

16 In this exercise quantile regressions of Koenker and Basset (1978) were used, as the aim is now to investigate the effect of changes in regressors on the quantiles of the conditional distribution of scores.

Chart 4A

PERFORMANCE IN MATHEMATICS IN PUBLIC AND PRIVATE SCHOOLS (STUDENTS IN THE 10TH GRADE) – CONSTANT FAMILY BACKGROUND |
SCORES IN PRIVATE SCHOOLS RELATIVE TO PUBLIC, PERCENTAGE DIFFERENCE AT THE 1ST QUARTILE (IN RED), THE MEAN (IN BLUE) AND THE 3RD QUARTILE (IN GREEN)

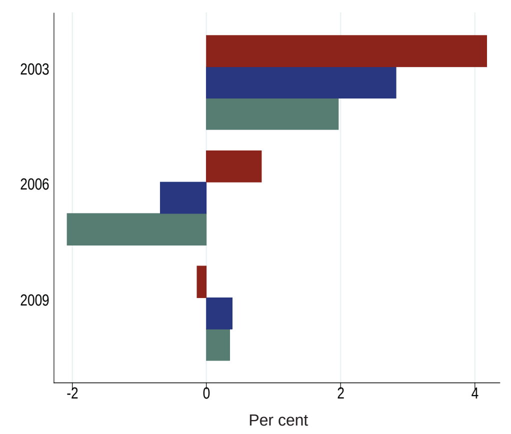


Source: Author's calculations.

Note: Coefficient of the indicator variable for private schools in (weighted) least squares and conditional quantile (Koenker e Bassett, 1978) regressions, also including the socioeconomic variables, the school location indicators and the constant as regressors.

Chart 4B

PERFORMANCE IN READING IN PUBLIC AND PRIVATE SCHOOLS (STUDENTS IN THE 10TH GRADE) – CONSTANT FAMILY BACKGROUND |
SCORES IN PRIVATE SCHOOLS RELATIVE TO PUBLIC, PERCENTAGE DIFFERENCE AT THE 1ST QUARTILE (IN RED), THE MEAN (IN BLUE) AND THE 3RD QUARTILE (IN GREEN)



Source: Author's calculations.

Note: Coefficient of the indicator variable for private schools in (weighted) least squares and conditional quantile (Koenker e Bassett, 1978) regressions, also including the socioeconomic variables, the school location indicators and the constant as regressors.

and 3B. In the 2009 national exams, private schools outperformed public schools on average by about 9 and 16 per cent, respectively, in Portuguese and mathematics (*Público*, 2009).

6. Conclusions

This study presents an analysis of the evolution of Portuguese students' scores in the OECD *PISA* cycles of 2003, 2006 and 2009. The main conclusions are as follows.

- After a relative stabilization between the 2003 and 2006 editions, Portuguese students' marks improved considerably in the 2009 edition, both in mathematics and reading. This allowed a progression towards intermediate positions in the ranking of EU countries, especially in the latter subject.

- The variation in scores between *PISA* cycles has been substantially influenced by the changes in determinants, particularly with regard to the family background of children and, more importantly, the distribution of students by grades.[17] Such changes in determinants have been partly caused by the use of sampling methods for data collection.

- Keeping the student characteristics and family background constant, there has been a steady improvement in scores over the considered *PISA* cycles, which can be attributed to the educational system. The positive impact of the return on the variables was, in the cycle of 2006 compared to 2003, offset by unfavourable changes in the distribution of students by grades and family context. On the contrary, between 2006 and PISA 2009, the two components reinforced each other, resulting in a sharp increase in marks.

- An analysis of scores in public and private schools in *PISA* indicates a tendency for a fading of the differences between both types of educational institutions. However, given the small number of private schools in the sample, the extrapolation of these findings to the universe of students appears to be problematic.

17 This suggests that a comparison between *PISA* results over various editions, even in descriptive terms, should be made according to the student grade, as a simple way of controlling for such changes.

Bibliography

Fortin Nicole, Thomas Lemieux and Sergio Firpo (2009). "Unconditional quantile regressions". *Econometrica*, 77(3): 953–973.

Fortin Nicole, Thomas Lemieux and Sergio Firpo (2011). "Decomposition Methods in Economics", in *Handbook of Labor Economics*, O. Ashenfelter e D. Card (eds.), Part 1, Volume 4A. Amsterdam: North-Holland.

Gebhardt, Eveline and Raymond Adams (2007). "The influence of equating methodology in reporting trends in PISA", *Journal of Applied Measurement*, 8(3): 305-322.

GEPE (2010). *A Educação em Números - Portugal 2010. Lisboa: Gabinete de Estatística e Planeamento da Educação*.

GEPE (2011). *Estatísticas da Educação - Anos Letivos de 2005/06 a 2008/09, tables available online at http://estatisticas.gepe.min-edu.pt/index.jsp. Lisboa: Gabinete de Estatística e Planeamento da Educação*.

GIASE (2006). *Séries Cronológicas, Alunos - 1985-2005. Lisboa: Gabinete de Informação e Avaliação do Sistema Educativo*.

Jornal Público (2009). *Ranking, Ensino Básico e Secundário. Suplemento à edição de 17 de Outubro de 2009*.

Koenker, Roger and Gilbert Basset (1978). "Regression quantiles". *Econometrica*, 46(1): 33–50.

Oaxaca, Ronald L. and Michael R. Ransom (1999). "Identification in detailed wage decompositions". *Review of Economics and Statistics*, 81(1): 154-157.

OCDE (2005). P*ISA 2003 Technical Report*. Paris: Organization for Economic Co-operation and Development.

OCDE (2009a). *PISA 2006 Technical Report*. Paris: Organization for Economic Co-operation and Development.

OCDE (2009b). *PISA Data Analysis Manual*. Paris: Organization for Economic Co-operation and Development.

OCDE (2010). *PISA 2009 Results: Learning trends - Changes in Student Performance since 2000*. Volume V. Paris: Organization for Economic Cooperation and Development.

OCDE (2011). *PISA 2009 Technical Report* (*preliminary version*), available online at http://www.oecd.org/. Paris: Organization for Economic Co-operation and Development.

Pereira, Manuel C. (2010). "Educational attainment and equality of opportunity in portugal and in europe: the role of school versus parental influence". Banco de Portugal, Economic Bulletin - Winter.

Woessmann, Ludger (2002). "How central exams affect educational achievement: international evidence from TIMSS and TIMSS-Repeat". *Harvard University Program on Education Policy and Governance Working Paper No. PEPG/02-10.*

Woessmann, Ludger, Elke Luedemann, Gabriela Schuetz, and Martin West (2009). School Accountability, *Autonomy and Choice around the World*. Cheltenham, UK: Edward Elgar.

Appendix 1: Definition of the explanatory variables calculated by the author

Educational resources at home. Index calculated from students' answers to six questions about household possession of the following items: a desk to study, a quiet place to study, a computer for schoolwork, educational software, books to help with schoolwork and a dictionary.

Grade amplitude. Calculated as the difference between the minimum and maximum grades taught at schools.

Autonomy of resources. Index computed from schools' answers to six questions about who has the responsibility for: teacher hiring, teacher firing, setting initial salaries, setting salary increases, formulation of the overall school budget, and changing allocations inside the budget.

Autonomy of curriculum and assessment. Index computed from schools' answers to four questions about who has responsibility for: defining student assessment policies, choosing the textbooks used, defining curricula, and choosing the courses offered.