Grade retention during basic education in Portugal: determinants and impact on student achievement[1]

Manuel Coutinho Pereira[2] | Hugo Reis[2]

ABSTRACT

The percentage of students repeating a school year is not homogeneous among European countries. In Portugal grade retention is a common practice. This article examines the determinants of grade retention and analyzes its impact on student performance in Portugal, presenting at the same time some results for a set of European countries. In Portugal, individual, family and peer characteristics are important factors explaining grade retention. In particular, students with less maturity and worse economic conditions are more likely to repeat. The effects of school retention are evaluated in the framework of a treatment effects model. Academic performance at a later stage of basic education is negatively affected by repeating at an early stage, which suggests that there will be advantage in implementing alternatives to support students. The short-term effects of repeating at a later stage are positive, although small. In this case, the results do not call into question grade retention.

1. Introduction

Management of students' learning difficulties varies from one country to another. In almost all European countries, according to the legislation, it is possible for a student to repeat a grade during compulsory education, in particular at the basic level.

Grade retention consists in holding students in the same grade for an additional year rather than promoting them along with their age peers. This practice is mostly based on the principle that by repeating a school year, a student has a further opportunity to improve its skills and knowledge. Even in the cases where some remedial measures are provided when students experience problems, retention is proposed as the ultimate measure of support. This practice affects also children that seem immature compared with their age peers. The idea is to give them an opportunity to catch up with the human capital needed in the next grade. Grade retention is also defended by those who claim that education systems are more efficient with less heterogeneous peer groups, and by those who view it as a way to ensure better accountability. However having students repeating grades is costly, entailing the expense of providing an additional year of education, and the cost to society in delaying the student's entry into the labour market. In addition, opponents of grade retention emphasize the psychological effects of such policy. They highlight, in particular, the decreased self-esteem, impaired peer relationship, distancing from school and increased dropout likelihood. Therefore, and taking into account the literature on grade retention, the impact of repeating a grade on students achievement and human capital accumulation is still an open question.

Some countries, like Norway and Iceland, have opted for automatic progression from one year to the other throughout compulsory education, and provide alternative support measures to students in difficulty. In the other European countries, students who have not met the defined criteria must repeat the year. The most common criteria are deficient academic progress,

attendance record, and behaviour. The percentage of students repeating a school year is not homogeneous among European countries even for those with similar regulations. For instance, in the Scandinavian countries and Italy less than 5 percent of the students aged 15 have ever been retained at ISCED 1 or ISCED 2. In contrast, in countries like Portugal, France, Spain and Luxembourg, figures are above 30 per cent. Therefore, in Portugal grade retention is a common practice. Despite some recent decreasing trend, the official figures from the Ministry of Education show that in 2011/12 academic year the retention rate was 15.6 percent in the 3rd cycle of basic education and 11.2 and 4.4 percent in the 2nd and 1st cycles, respectively.

Our article studies the determinants of grade retention at an earlier and a later stage of basic education and its impact on academic achievement. The analysis is performed for Portugal and for a set of European countries, using the OECD's Programme for International Student Assessment (PISA) dataset. According to the International Standard Classification of Education (ISCED), we analyze separately grade retention in two different moments of basic education: ISCED 1 and ISCED 2. In Portugal, the first level corresponds to the 1st and 2nd cycles (1st to 6th grade) and the second level to the 3rd cycle (7th to 9th grade).[3] The effects of school retention are evaluated in the framework of a treatment effects model where selection is endogenous, *i.e.* correlated with performance.

An important issue when evaluating the effect of grade retention is that it is impossible to hold both the grade and age fixed when retaining a student. The PISA dataset, covering students aged 15, is better-suited for measuring the later type of effect, *i.e*. comparing student's cognitive development holding age fixed (see Cooley, *et al.*, 2011, and Schwerdt and West, 2012, for a more detailed discussion of the two approaches). This contrasts with most studies in this area which use samples of students in the same grade.

In what concerns the determinants of grade retention, individual, family and peer characteristics are important factors. In particular, for Portugal, students with less maturity and worse economic conditions are more likely to fail a grade. Socio-economic aspects, school characteristics and differences at the regional and country levels (for example, institutional factors) are important as well.

In this article, we found that school retention during ISCED 1 produces negative effects on student performance in the long-term (defined as more than three years - recall that PISA tests assess student performance at the end of, or just after, ISCED 2). From a policy perspective there seems to be scope for intervention, namely to replace, at least partially, this practice by other forms of student support, which may turn out to be less expensive. The effects of retention at ISCED 2 are positive, despite small, in the short-term (again, taking into account the timing of PISA tests). Although we are not able to address long-term effects in this case, our evidence does not contradict grade retention in more advanced levels of schooling. Finally, it should be noted that students whose socio-economic characteristics make them more likely to repeat are generally also those who gain most (or lose less) with treatment.

Our study belongs to the literature on the impact of retention on educational performance. In contrast to early contributions that did not address endogeneity or selection problems, recent papers provide some evidence of positive effects, in particular in the short-term. For example, Jacob and Lefgren (2009) and Schwerdt and West (2012) (using a regression discontinuity approach) for the US, and with different approaches, Mahjoub (2012), Elodie (2013) and Gary-Bobo, *et al.* (2014) for France. In addition, Baert, *et al.* (2013) using a structural dynamic choice model and Belgian data found some positive effects on performance in the year after retention and also some persistence. All these works essentially perform a same-grade comparison. With

the PISA dataset but using slightly different approaches than the one in this article, Diris (2012) and Garcia-Perez *et al.* (2014) found in general negative effects of retention for a set of OECD countries and Spain, respectively.

The remaining of this article is organized as follows. Section 2 presents the data and summary statistics. Section 3 describes the regression results about determinants of grade retention. Section 4 presents the model used to evaluate the impact on test scores, and section 5 discusses the findings in this regard. Section 6 concludes

2. The database and descriptive analysis

In this study we use the PISA dataset for a group of European countries[4] in 2003 and 2009, years for which there is information on whether a student repeated one or more years during basic education. The student, family and school variables in the regressions are essentially those already employed in previous studies using PISA data, such as Pereira and Reis (2012) (see table 2.3.1 in section 2 and appendix 1 for a complete list of the variables and respective means). There are a couple of additions that are worth highlighting: i) indicators of retention at both levels, ISCED 1 and ISCED 2, computed from the student questionnaire; ii) indicators of kindergarten attendance; iii) indicators of entry age in primary education and its relationship to the cut-off date regulated by law; and iv) an indicator that reflects the maturity of the student. On the basis of the PISA dataset, we also add variables that capture variations at the school level (peers) and at the regional level[5].

2.1. Retention indicator

Table 2.1.1 shows for 2003, 2009 and for the two years together, the percentage of students in the sample who repeated only at ISCED 1, only at ISCED 2, at both levels, and those who have not repeated during basic education. Columns 1 to 3 present the results for the full set of countries, columns 4 to 6 show the results for countries with data available in the two years, and the last three columns present the values for Portugal. The results for 2003 and 2009 are very similar, even when we add further countries in 2009.

The percentage of students who never repeated is around 82 percent, for the full set of countries, compared with only 70 per cent for Portugal. This difference mainly reflects the number of repeaters during ISCED 1, for which the percentage in Portugal is around 13 percent, about twice the average for the other countries. Concerning ISCED 2, Portugal also displays values nearby 13 percent, closer to the average for the other countries that is above 10 percent.

Table 2.1.1 • Grade Retention in Portugal and in a set of European countries

	Whole Sample Percentage (%)			Sample with same countries in both years Percentage (%)			Portugal Percentage (%)		
	2003	2009	Total	2003	2009	Total	2003	2009	Total
Non-repeaters	81.5	82.5	82.0	81.5	81.4	81.5	70.5	70.4	70.4
Repeaters only in ISCED 1	6.5	6.4	6.4	6.5	6.8	6.6	12.8	13.7	13.2
Repeaters only in ISCED 2	10.5	9.7	10.1	10.5	10.2	10.4	12.6	12.8	12.7
Repeaters in both ISCED 1 and 2	1.5	1.5	1.5	1.5	1.6	1.5	4.1	3.2	3.7
Total	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0
Number of countries	20	25		20	20				

Source: PISA 2003 and 2009.
Note: In Portugal, ISCED 1 corresponds to the 1st and 2nd cycles together, and ISCED 2 to the 3rd cycle. For the remaining european countries, the number of grades in each level may be different (see Eurodyce, 2013).

Chart 2.1.1 provides an international comparison of the percentage of students who repeated at least once in basic education. Besides Portugal that has a rate of around 30 percent, countries like Spain, Luxembourg and France stand out with figures higher than 35 percent. In contrast, about half of the countries have percentages below 5 percent, indicating considerable heterogeneity of retention practices. These lower levels also somehow reflect the European trend to limit this practice, fostered by some international organizations such as the OECD (see, for example, OECD, 2012).

2.2. Test scores

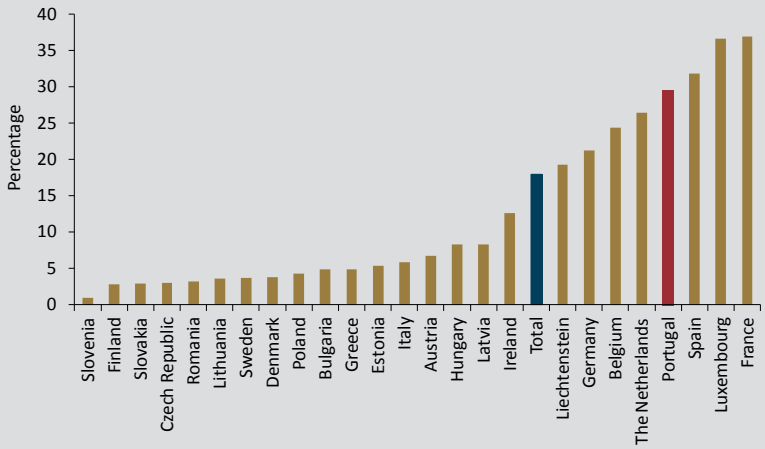
Table 2.2.1 shows the mean and standard deviation of math and reading PISA test scores by type of retention[6], comparing Portugal with the countries considered as a whole. The table shows not only a difference between repeaters and non-repeaters, but also heterogeneity within repeaters. In particular, those who were retained at ISCED 2 have higher scores than those who repeated at ISCED 1, which in turn have better outcomes than those who repeated at the two levels. Comparing Portugal with the overall mean, there are only significant differences for the students who repeated at ISCED 1 or ISCED 2 only, where Portugal has lower values. Regarding the students who never repeated and those who repeated at both levels, the mean scores are very similar to other European countries.

Table 2.2.1 • Test scores by retention levels - mean and standard deviation

	Total		Portugal	
	Mathematics	Reading	Mathematics	Reading
Non-repeaters	508.4	505.0	512.6	520.7
	(87.9)	(86.4)	(71.2)	(65.7)
Repeaters only in ISCED 1	418.0	411.2	396.6	406.4
	(81.6)	(83.6)	(60.8)	(66.8)
Repeaters only in ISCED 2	448.5	442.1	426.2	428.7
	(83.5)	(86.5)	(59.6)	(63.0)
Repeaters in both ISCED 1 and 2	378.8	373.4	372.7	371.0
	(81.1)	(86.3)	(55.4)	(64.3)
Total	497.6	493.7	481.2	488.4
	(91.5)	(91.0)	(84.1)	(82.9)

Sources: PISA 2003 e 2009.
Note: This table presents the mean of test scores by retention level. Standard deviation in brackets.

Chart 2.1.1 • Percentage of students who have repeated a year at least once in basic education



Source: PISA 2003 and 2009.

2.3. Explanatory variables

We end this section with a brief analysis of explanatory variables, in particular individual and family attributes. Table 2.3.1 presents a characterization of these variables by type of retention, comparing Portugal with the average for the countries in the sample.

Individual variables include gender, binary variables for students who attended kindergarten for only one year and more than one year, and two variables related to entry in primary education. The first one - entry age - captures the age attained in the year of entry relative to age 6 (regardless of the official entry age). For example, the variable takes on the value 0 if the student entered school in the year he was 6, and the value 1 if the student entered school in the year he was 7. The second variable - late entry - flags the students for whom it is possible to detect a difference between the actual and the regulatory entry ages, namely when there is a cut-off date during the year that was not met (for example, by a decision of the parents and/or teachers). Note that, despite the existence of such rules in many countries, for some of them, including Portugal[7], this does not show up clearly in the data. Still, for some countries in the sample, late entry may capture specific characteristics of the students that are thereby purged from the estimated impact of retention.

Individual variables include an indicator of the maturity of the student, in terms of relative age, in the spirit of Bedard and Dhuey (2006). This measure of maturity is calculated as the difference in months between the dates in which the student has completed 6 and entered school (using September of the entry year as a reference)[8] :

Maturity= 9 – birth month + 0, if the student entered school in the year that turns 6 years old (entry age = 0);

Maturity = 9 – birth month + 12, if the student entered school in the year that turns 7 years old (entry age = 1);

Maturity = 9 – birth month + 24, if the student entered school in the year that turns 8 years old (entry age = 2).

Table 2.3.1 shows that there are significant differences in some individual characteristics between repeaters and non-repeaters, such as gender and kindergarten attendance. Regarding this last attribute, the Portuguese figures are lower than for the whole sample in both groups. Table 2.3.1 also shows that Portuguese students have, on average, less maturity at entry than their European counterparts. In particular, the difference is about 5 months among those who did not repeat and about 2 months among those who repeated. This is explained by the official later entry age in some countries (age 7), and also by the loose implementation of the cut-off date rule in Portugal (as it follows from the data). Moreover, for the full set of European countries, the average maturity of non-repeaters exceeds that of repeaters. This also holds for Portugal, conditional on the entry age cohorts (this is the relevant type of analysis, by construction of the maturity indicator).

Regarding family variables, students who did not repeat have a higher percentage of more educated parents and better socio-economic conditions (measured by the books at home variable). Among repeaters, those who repeated only at ISCED 2 feature better socio-economic conditions than those who repeated at ISCED 1 only. Moreover, there is a higher percentage of single-parent families among the students who were retained, but no significant difference in this regard between the two groups of repeaters.

Besides student and family variables, we also considered peer and school characteristics, as well as covariates to capture regional variability in socio-economic conditions, and also in attitudes towards education and school retention practices. Finally, the regressions include country fixed--effects that absorb variation arising from institutional differences, particularly regarding retention, and year fixed-effects. The remaining variables used in the analysis are presented in appendix 1, including the respective means for the whole sample and Portugal.

Table 2.3.1 • Summary statistics - student and family characteristics by retention level | Mean and standard deviation

	Non-repeaters		Repeaters					
			Only in ISCED 1		Only in ISCED 2		ISCED 1 and ISCED 2	
	Total	Portugal	Total	Portugal	Total	Portugal	Total	Portugal
Individual characteristics								
Female (b)	0.52	0.56	0.44	0.48	0.41	0.44	0.37	0.37
	(0.50)	(0.50)	(0.50)	(0.50)	(0.49)	(0.50)	(0.48)	(0.48)
Kindergarten – 1 year (b)	0.18	0.18	0.19	0.22	0.14	0.20	0.17	0.19
	(0.38)	(0.38)	(0.39)	(0.42)	(0.35)	(0.40)	(0.38)	(0.39)
Kindergarten – 2 or more years (b)	0.77	0.61	0.71	0.48	0.79	0.57	0.71	0.52
	(0.42)	(0.49)	(0.45)	(0.50)	(0.41)	(0.50)	(0.45)	(0.50)
Entry age	0.52	0.06	0.35	0.23	0.27	0.12	0.12	0.08
	(0.53)	(0.29)	(0.52)	(0.49)	(0.46)	(0.38)	(0.33)	(0.27)
Late entry	0.07	-	0.09	-	0.06	-	0.02	-
	(0.26)	-	(0.31)	-	(0.26)	-	(0.13)	-
Maturity	8.7	3.4	6.4	4.8	5.7	3.7	3.3	2.6
	(6.3)	(4.5)	(6.1)	(6.6)	(5.7)	(5.2)	(4.9)	(4.5)
Family variables								
Books at home (25-200) (b)	0.52	0.53	0.44	0.38	0.50	0.47	0.42	0.34
	(0.50)	(0.50)	(0.50)	(0.48)	(0.50)	(0.50)	(0.49)	(0.47)
Books at home (>200) (b)	0.26	0.20	0.12	0.06	0.17	0.09	0.09	0.07
	(0.44)	(0.40)	(0.33)	(0.25)	(0.38)	(0.29)	(0.29)	(0.26)
Lower secondary education (b)	0.09	0.19	0.15	0.20	0.15	0.25	0.19	0.18
	(0.28)	(0.39)	(0.36)	(0.40)	(0.36)	(0.43)	(0.39)	(0.39)
Upper secondary education (b)	0.46	0.24	0.39	0.16	0.41	0.19	0.36	0.16
	(0.50)	(0.43)	(0.49)	(0.37)	(0.49)	(0.40)	(0.48)	(0.36)
Terciary education (b)	0.42	0.30	0.34	0.17	0.32	0.17	0.26	0.17
	(0.49)	(0.46)	(0.47)	(0.37)	(0.47)	(0.37)	(0.44)	(0.37)
Single-parent family (b)	0.13	0.12	0.18	0.17	0.18	0.17	0.18	0.16
	(0.34)	(0.33)	(0.38)	(0.37)	(0.38)	(0.37)	(0.39)	(0.37)
No parents at home (b)	0.01	0.02	0.02	0.05	0.02	0.03	0.04	0.05
	(0.11)	(0.14)	(0.15)	(0.21)	(0.13)	(0.16)	(0.18)	(0.22)
Immigrant (b)	0.06	0.04	0.20	0.09	0.13	0.06	0.14	0.06
	(0.24)	(0.20)	(0.40)	(0.28)	(0.34)	(0.24)	(0.35)	(0.23)
Language at home different from the test language	0.04	-	0.03	-	0.06	-	0.05	-
	(0.20)	-	(0.18)	-	(0.24)	-	(0.23)	-
Foreign language at home	0.03	0.01	0.11	0.03	0.07	0.02	0.08	0.02
	(0.18)	(0.10)	(0.31)	(0.16)	(0.25)	(0.15)	(0.27)	(0.13)
Educational resources at home	5.04	5.05	4.62	4.45	4.78	4.63	4.37	4.19
	(1.08)	(0.99)	(1.30)	(1.24)	(1.21)	(1.16)	(1.31)	(1.30)
White collar - low skill (b)	0.23	0.28	0.25	0.29	0.28	0.33	0.29	0.31
	(0.42)	(0.45)	(0.43)	(0.46)	(0.45)	(0.47)	(0.45)	(0.46)
White collar - high skill (b)	0.54	0.43	0.39	0.18	0.41	0.23	0.26	0.13
	(0.50)	(0.49)	(0.49)	(0.38)	(0.49)	(0.42)	(0.44)	(0.33)

Source: PISA 2003 and 2009.
Note: (b) stands for binary variable. Standard deviation in brackets.

In PISA there is only data on peers (mean of the family variables at the school level) and schools at the time the student takes the test. This issue becomes particularly relevant for students who repeated at ISCED 1, as the majority of students who have failed a grade at ISCED 2 most likely still attend the same school (*i.e.* they have not changed school meanwhile).

Regarding students who repeated at ISCED 1, we construct variables that proxy the characteristics of schools they were attending at the time. Accordingly, for students in PISA currently attending a school that also has ISCED 1, it was assumed that the student has remained in that school throughout. In the other cases, we calculated an average of the characteristics of schools with ISCED 1 by region and type of community where the school is located[9]. In spite of measurement error, this procedure should make up partly for the lack of information regarding schools when students were retained at an early stage of basic education, and mitigate some endogeneity.

3. Determinants of grade retention

The analysis of the determinants of retention is performed separately for students who repeated (once or more) at ISCED 1 and ISCED 2, using a probit model. Students who repeated at the two levels are excluded from both samples[10]. Moreover, in the rest of the article, we consider only 23 countries: France was excluded for not having school data and Slovenia for having less than 1 percent of repeaters.

3.1. Main results

Table 3.1.1 shows the determinants of grade retention at ISCED 1 and ISCED 2 in Portugal and for the countries in the sample as a whole. Regarding ISCED 1, the results in columns 1 and 2 show clearly that individual, family and peer factors are important in explaining retention. In general, all variables are statistically significant in the entire sample; for Portugal, however, kindergarten attendance is not significant. Note that the effects tend to be stronger for Portugal than for the set of European countries in the sample. In Portugal, the probability of repeating at ISCED 1 decreases by around 3.5 percentage points (pp) with an increase of one standard deviation (about five months) in maturity. The same probability decreases by a similar magnitude when the student is female. For the countries as a whole, the effect of maturity is also the most important one, albeit smaller, reaching about 2 pp. Another important variable in the full sample is kindergarten attendance for at least two years, which reduces the probability of repeating at ISCED 1 by around 1.4 pp.

The most important family variables are parental education and books at home (variable measuring differences in income). As with individual attributes, the effects are stronger for Portugal than for the entire sample. In particular, the probability of repeating at ISCED 1 decreases by about 4.5 pp when students have more books at home (around 1.5 pp for all countries). The same probability decreases for Portugal by more than 2 pp if the parents have higher levels of education (about 1 pp for the full sample). Another important factor is family structure. For example, for Portugal the probability of failing a grade increases by about 3.3 pp if the student belongs to a single-parent family.

Results for ISCED 2, in columns 3 and 4, show many similarities to ISCED 1, although the magnitude of the effects is different. As in the previous case, the effects for Portugal are in general stronger than those for all countries. Individual characteristics appear as particularly important. In Portugal, the probability of repeating at ISCED 2 decreases by around 3 pp with an increase

Table 3.1.1 • Grade retention determinants in two levels of basic education

	ISCED 1		ISCED 2	
	Total	**Portugal**	**Total**	**Portugal**
Individual variables				
Female [b]	-0.009	-0.034	-0.026	-0.044
	(0.001)***	(0.007)***	(0.002)***	(0.007)***
Kindergarten – 1 year [b]	-0.002	0.003	-0.004	0.016
	(0.002)	(0.011)	(0.003)	(0.012)
Kindergarten – 2 or more years [b]	-0.014	-0.008	-0.011	0.012
	(0.002)***	(0.009)	(0.003)***	(0.009)
Entry age	-0.008	-	-0.028	-0.012
	(0.002)***	-	(0.003)***	(0.019)
Late entry	0.021	0.039	0.008	-
	(0.002)***	(0.023)*	(0.004)*	-
Maturity	-0.003	-0.007	-0.002	-0.006
	(0.000)***	(0.001)***	(0.000)***	(0.001)***
Family variables				
Books at home (25-200) [b]	-0.011	-0.044	-0.013	-0.015
	(0.001)***	(0.009)***	(0.002)***	(0.008)*
Books at home (>200) [b]	-0.015	-0.046	-0.019	-0.032
	(0.001)***	(0.011)***	(0.002)***	(0.010)***
Lower secondary education [b]	0.000	-0.023	-0.002	0.002
	(0.002)	(0.009)**	(0.003)	(0.009)
Upper secondary education [b]	-0.010	-0.033	-0.009	-0.014
	(0.002)***	(0.009)***	(0.003)***	(0.009)
Terciary education [b]	-0.011	-0.023	-0.015	-0.027
	(0.002)***	(0.011)**	(0.003)***	(0.010)***
Single-parent family [b]	0.008	0.033	0.023	0.027
	(0.002)***	(0.012)***	(0.003)***	(0.012)**
No parents at home [b]	0.019	0.000	0.027	0.004
	(0.006)***	[0.027]	(0.008)***	(0.024)
Immigrant [b]	0.018	-0.063	0.014	0.015
	(0.003)***	(0.012)***	(0.004)***	(0.021)
First language not test language	-0.007	-	0.002	-
	(0.001)***	-	(0.004)	-
Foreign language at home	0.004	0.089	0.002	0.033
	(0.003)	(0.058)	(0.005)	(0.035)
Educational resources at home	-0.005	-0.017	-0.007	-0.009
	(0.000)***	(0.005)***	(0.001)***	(0.004)**
White collar - low skill [b]	-0.007	-0.015	-0.002	-0.002
	(0.001)***	(0.008)*	(0.002)	(0.008)
White collar - high skill [b]	-0.013	-0.04	-0.006	-0.022
	(0.001)***	(0.011)***	(0.002)***	(0.009)**
Peer variables				
Books at home (25-200) - peers	-0.024	-0.235	0.000	-0.280
	(0.013)*	(0.133)*	(0.011)	(0.090)***
Books at home (>200) - peers	-0.016	-0.503	0.002	-0.360
	(0.017)	(0.225)**	(0.014)	(0.133)***
Lower secondary education - peers	-0.018	0.172	0.004	0.094
	(0.022)	(0.137)	(0.021)	(0.101)

Sources: Authors' calculations.
Notes: This table shows the marginal effects from a probit model. (b) stands for binary variable. Standard deviation in brackets. Standard deviation in brackets. * significant at 10%; ** significant at 5%; *** significant at 1%.

Table 3.1.1 • Grade retention determinants in two levels of basic education (continued)

	ISCED 1		ISCED 2	
	Total	**Portugal**	**Total**	**Portugal**
Peer variables (continued)				
Upper secondary education - peers	-0.064	-0.048	0.002	0.010
	(0.020)***	(0.140)	(0.017)	(0.096)
Terciary education - peers	-0.054	0.098	-0.033	0.192
	(0.021)***	(0.152)	(0.018)*	(0.104)*
Single-parent family - peers	0.057	0.615	0.044	0.261
	(0.013)***	(0.189)***	(0.014)***	(0.112)**
No parents at home - peers	0.065	0.412	0.079	0.214
	(0.033)*	(0.266)	(0.038)**	(0.199)
Immigrant - peers	0.01	0.3	0.036	-0.053
	(0.016)	(0.155)*	(0.012)***	(0.131)
Educational resources at home - peers	-0.012	-0.132	-0.010	-0.152
	(0.004)***	(0.050)***	(0.004)**	(0.035)***
White collar - low skill - peers	-0.011	0.033	0.004	0.009
	(0.014)	(0.129)	[0.013]	(0.090)
White collar - high skill - peers	-0.04	-0.271	-0.010	-0.109
	(0.014)***	(0.141)*	[0.012]	(0.091)
Other control variables	School variables	School variables	School variables	School variables
	Regional variables	Regional variables	Regional variables	Regional variables
	year fixed-effect	year fixed-effect	year fixed-effect	year fixed-effect
	country fixed-effect	-	country fixed-effect	-
Number of students	233935	8620	236666	8619
	Maturity		Maturity and retention practices	
Joint Test (F-test)	250.8	30.3	608.6	40.5
p-value	(0.00)	(0.00)	(0.00)	(0.00)

Sources: Authors' calculations.
Notes: This Table shows the marginal effects from a probit model. (b) stands for binary variable. Standard deviation in brackets. Standard deviation in brackets. * significant at 10%; ** significant at 5%; *** significant at 1%.

of one standard deviation in maturity. Maturity is also the most important variable for the whole sample, but with a smaller effect of about 1.3 pp. Kindergarten attendance for at least two years is equally important, reducing the propensity to repeat at ISCED 2 by about 1 pp For the whole sample, another relevant factor is entry age: entering the education system one year later (regardless of the official entry age) decreases the probability of repeating at ISCED 2 by about 3 pp. This effect is smaller for ISCED 1 (below 1 pp). For Portugal, this variable does not show up as significant, particularly for ISCED 2, which is due to the loose implementation of the cut-off date rule (see above). Therefore, variables entry age and late entry coincide, and we cannot disentangle the effect of entering later than the date officially prescribed. The effects of family characteristics, such as parental education, number of books at home and single-parent family, are also important, especially in Portugal. All these features have a similar impact on the propensity to repeat at ISCED 2 (about 3 pp).

In addition, the joint tests for individual, family, peer, school, region and country variables, separately, suggest that all these groups of variables influence retention in both education levels analyzed. In Portugal, the effect of schools appears less evident for ISCED 1. These results show that socio-economic aspects are important but are not the only factors explaining grade retention. In particular, the results for the country-fixed effects show that institutional factors may have an important role in the observed differences between countries.

The main findings are robust to different specifications at both levels, ISCED 1 and ISCED 2, namely, interacting with country and year fixed-effects, removing regional variables, or using school fixed-effects. These results are also robust to a change in the group of countries considered by restricting it to countries with higher percentages of retention (see Pereira and Reis, 2014).

4. The treatment effects model

The effects of school retention are evaluated in the framework of a treatment effects model with endogenous selection (see *e.g.* Wooldridge, 2002, Chapter 18, or Blundell and Costa Dias, 2009).

4.1. Methodology

We assume that the performance of each student will depend on whether she/he was retained. Scores with and without treatment are random variables, respectively, S_R and S_{NR} given by

$$S_R = \mu_R + x\,\beta_R + u_R \tag{1}$$

and

$$S_{NR} = \mu_{NR} + x\,\beta_{NR} + u_{NR}. \tag{2}$$

In this general specification, the impact of the determinants observed by the analyst on performance differs depending on whether there is student retention ($x\,\beta_R \neq x\,\beta_{NR}$). In other words, each of these cases gives rise to a distinct education production function. There is observed heterogeneity in the treatment effects, as the treatment interacts with other variables in score determination. In practice, vectors β_R and β_{NR} may differ only for a subset of regressors in x.

At the same time, there are factors that the researcher does not observe, for example, the capabilities and motivation of the student, which also determine performance. Such factors are captured by the error terms in the equations above. In this model, as the error terms are also assumed to be different ($u_R \neq u_{NR}$), grade repetition is allowed to interact with unobserved factors or, equivalently, there is unobserved heterogeneity in its effects.

To complete the model, it is assumed that the mechanism underlying the selection of students to retain is:

$$T^* = x\,\boldsymbol{\pi}_1 + z\,\boldsymbol{\pi}_2 + e, \tag{3}$$

where T* is a latent variable. In general, the covariates considered in the education production functions (vector x) are also part of the selection mechanism. Vector z z includes variables that are not related to school performance (so that they do not enter equations (1) and (2)) but influence the selection of repeaters, *i.e.* the propensity to repeat. These variables function as instrumental variables, and have a key role in the estimation of treatment effects in the presence

of endogeneity. Note that in this model endogeneity is captured by the non-zero covariances between the error terms of the primary equations (1) and (2), and the error term of the selection equation (*i.e.* $\sigma(u_{R'}, e) \neq 0$ and $\sigma(u_{NR}, e) \neq 0$). Indeed, even controlling for the observed determinants of grade retention, selection for treatment remains endogenous as unobserved factors, such as the ability and motivation of the student, play an important role both in the probability of repeating and the explanation of performance.

The treatment indicator, T, is a function of T^* such that $T=1$ if and only if $T^* \geq 0$ and $T=0$ if and only if $T^*<0$. In practice, the analyst observes a realization of the variables $S_{R'}$ when the student fails a grade (*i.e.* $S_R \mid x, T=1$), and S_{NR} otherwise ($S_{NR} \mid x, T=0$). The counterfactual scores $S_R \mid x, T=0$ and $S_{NR} \mid x, T=1$ are not observed. The quantities of interest that can be obtained in this context include the average treatment effect, $ATE = E(S_R - S_{NR})$, and the average treatment effect on the treated $ATET = E(S_R - S_{NR} \mid x, T=1)$. We can also estimate what would have been the effect of treatment on those who were not treated, $ATENT = E(S_R - S_{NR} \mid x, T=0)$.

The estimation of the parameters of interest in the education production functions is carried out by the Heckman control function method (Heckman, 1978, 1979, or Vella and Verbeek, 1999 for a more recent reference), which is consistent under the hypothesis of normality of the error terms. In the process, we also get an estimate of the covariance between the errors in the outcome equations and the selection equation (see *e.g.* Pereira and Reis, 2014).

In the regressions we included the greatest range of control variables available in PISA for the years and countries considered, namely covariates related to the student and family, the current school and peers, as well as regional indicators derived from socio-economic measures and non-cognitive outcomes and year and country fixed-effects (see table 2.3.1 and appendix 1). In the regressions measuring the effect of retention at ISCED 1 we also used the variables described at the end of section 2.3 to capture the school environment that surrounded the students during those grades. Although performance is measured when PISA tests are taken, such ISCED 1 variables should be included to the extent that they determine repetition, in order to eliminate possible sources of endogeneity of the instruments in the outcome equations. However, we considered only the school regressors that proved significant as determinants of repetition and a subset of peer covariates.

The effects of student retention in Portugal are evaluated on the basis of regressions estimated for the extended sample of European countries that we have been considering. As explained in the next section, this solution allows benefiting from a more compelling instrumentation of the treatment indicator than restricting the sample to Portuguese students. However, in order to estimate a specific education production function for Portugal, we interacted all explanatory variables in the model (except the country fixed-effects) with the fixed-effect for Portugal.[11]

With regard to observed heterogeneity, the repetition indicator was interacted with the student and family regressors as well as the country fixed-effects. Furthermore, we allow a specific impact of the first two groups of covariates in the Portuguese case. The regressions are weighted by the student weights in the PISA database.

With the estimates of the different parameters in hand, it is possible to use the theoretical expressions of the treatment effects to estimate them as a function of the variables in x and, in particular, for subgroups of interest within the student population. The average effects of grade

retention for the whole or a subgroup of Portuguese students (namely, those who repeated and those who did not) are obtained as averages over the respective subsamples. One can also calculate the corresponding estimates for all countries underlying the estimation of the model. The standard deviations of the estimates are calculated by bootstrap.

4.2. Motivation for the instrumental variables

The endogeneity in this treatment effects model stems from a probable correlation between grade failure and non-observable individual attributes, so that the latter may be reflected on the measured effect of grade failure. Therefore the indicator of repetition at ISCED 1 was instrumented by a maturity index, in the vein of Angrist and Krueger (1991) and, more recently, Bedard and Dhuey (2006) (see section 2.3 for the details about the construction of this variable), and the indicator of repetition at ISCED 2 by the same maturity indicator and a variable that seeks to capture different regional retention practices.

It is straightforward to argue that maturity affects the probability of grade retention, particularly at ISCED 1, since older children, with more maturity, are likely to perform better. In addition, teachers may be reluctant to hold back older children due to negative stigma effects. Indeed, in the previous section, the indicator of maturity emerges as an important determinant of repetition. However, the main question is whether a measure of maturity is related to unobserved individual attributes, after controlling for all the observable ones (namely, individual, family, peer, school and regional) and country fixed-effects. We assume that the additional variation within each country, conditional on all those variables, is generated by exogenous shocks (*e.g.* an unpredictable event positively or negatively affecting the student). In this case, the individual unobservable attributes will not be related to maturity.

As an alternative approach to retention at ISCED 2, which is closer to the moment the PISA tests are taken, we use an additional variable that captures cross-country differences in grade retention practices. For this purpose, we take the average level of regional retention[12], controlling at the same time for socioeconomic characteristics and attitudes towards education in those regions. Our hypothesis is that after controlling for these variables as well as school characteristics and country fixed-effects, the variation in retention across regions will mainly capture differences in retention practices. In some countries, for instance, the Netherlands, the regulation of grade retention is not centralized, but partly defined at the regional level. If our instrument captures such differences, it is possible to argue that it affects the likelihood of repetition. Again, the question is whether the instrument is related to unobservable determinants of test scores at a later point in time. It is assumed that the regional repetition level affects individual performance only through the repetition indicator or; in other words, the unobservable determinants of test scores do not vary in response to changes in retention practices.

Despite being strong, these assumptions are tenable given the detailed information we have, respectively, by student and region (see Pereira and Reis, 2014, for further discussion about the validity of the instruments). The results for determinants of grade repetition and significance of the instruments presented in table 3.1.1 show that in our main specifications the instruments (maturity for ISCED 1 and maturity plus average retention for ISCED 2) are important explanatory factors of the likelihood of repetition. Furthermore, the main findings in the article are robust to using the instruments separately.

To the extent that our model includes country fixed-effects, the use of the abovementioned instruments is essentially based on variability within countries as a source of exogenous variation.[13] With regard to retention practices, given that the Portuguese education system is very centralized,

such exogenous variation (notably, regional or between schools) will tend to be lower than in other countries in the sample. Concerning maturity at the beginning of compulsory education, the variability in the distribution throughout the year of the time the required age is reached may be larger in countries with higher educational decentralization, as the regulations in this area tend to be vague and admit numerous exceptions.[14] Therefore the instrumentation of repetition indicator benefits from estimating a multi-country regression.

5. Empirical results

This section presents the estimates of the average treatment effects (ATE) for the Portuguese students in PISA, as well as for repeaters (ATET) and non-repeaters (ATENT), obtained by the Heckman control function method. We consider grade failure at ISCED 1 and ISCED 2, and scores in reading and math, as a measure of performance. We present, in addition, the treatment effects estimated by ordinary least squares.[15] Naturally, in the presence of endogeneity and interaction of grade failure with unobservables, this estimator is biased. Nevertheless, a comparison of the least squares estimates with estimates taking into account selection for treatment and unobserved heterogeneity provides an additional cross-checking of the findings.

5.1. Impact of student retention on test scores

The estimates in table 5.1.1 differ substantially depending on whether we consider grade failure at an initial stage of basic education or later on. We estimate a negative and statistically significant effect of retention during ISCED 1 on the performance of Portuguese students, both for those who were treated (ATET), and for those who were not, if they had been (ATENT). Contrary to what might be expected, the impact of treatment is (slightly) more negative among the first of these groups, although the difference is not statistically significant. In relative terms, in the case of reading, the estimates in table 5.1.1 correspond to about -19 and -14 percent of the average score, respectively, for repeaters and non-repeaters. In contrast, for retention at ISCED 2, the findings are clearly differentiated depending on the group considered. We estimate a positive and statistically significant effect on the performance of repeaters, although of a small magnitude (about 4

Table 5.1.1 • Average impact of grade retention during basic education in Portugal

		ISCED 1				ISCED 2			
		Reading		Mathematics		Reading		Mathematics	
Heckman control function	ATE	-74.3	(6.2)***	-83.4	(6.0)***	-15.7	(4.9)***	-21.4	(4.7)***
	ATET	-76.5	(6.5)***	-87.9	(5.9)***	15.7	(5.3)***	14.0	(5.1)***
	ATENT	-73.9	(6.6)***	-82.6	(6.5)***	-21.4	(5.3)***	-27.8	(5.1)***
	$\sigma(u_R, e)$	-7.1	(3.6)**	-8.0	(3.6)**	-27.2	(2.9)***	-24.2	(2.7)***
	$\sigma(u_{NR}, e)$	1.1	(4.0)	3.9	(3.7)	-53.3	(3.1)***	-53.0	(2.9)***
Ordinary least squares	ATE	-83.0	(2.7)***	-92.2	(2.4)***	-72.8	(2.2)***	-74.2	(2.2)***
	ATET	-76.1	(2.3)***	-83.6	(2.2)***	-68.5	(2.2)***	-69.3	(2.2)***
	ATENT	-84.3	(2.9)***	-93.8	(2.6)***	-73.5	(2.2)***	-75.1	(2.3)***

Source: Authors' calculations.
Note: The treatment effects are obtained as described in section 4.1; the table presents averages for the Portuguese students. The parameters σ(uR,e) and σ(uNR,e) are estimated commonly for all countries. Standard deviations obtained by bootstrap in parentheses. * Significant at 10%; ** Significant at 5%; and *** significant at 1%.

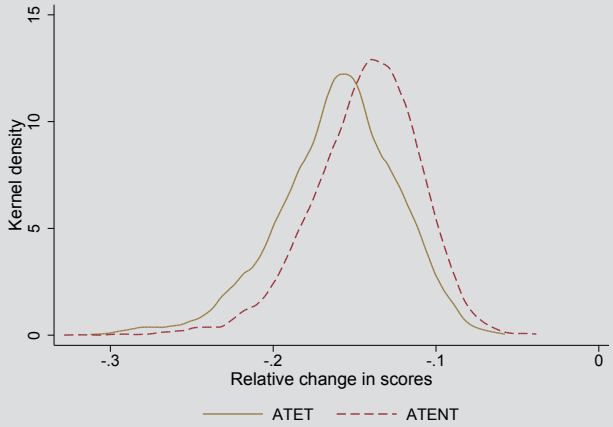
percent of the respective average score for reading). For non-repeaters, we still estimate a negative impact if the treatment had occurred, but a less important one (about -4 percent of the average group score for reading). Evidence on the basis of math and reading scores is very similar.

The evidence presented in table 5.1.1 is complemented with charts 5.1A and 5.1B that show the distribution by the student population of treatment gains in terms of reading performance, for repeaters and non-repeaters (the corresponding charts for math give similar indications). Note that when measuring the treatment gains for a given student it is preferable to use the relative variation in scores than the absolute. To this end, we take as a reference the situation in the absence of retention (which may be observed or counterfactual, depending on the group of students at issue). Therefore in the case of a student i who was held back, we take the ratio to the score excluding the estimated gains from treatment, *i.e.* $ATET_i \ / \ (S_i - ATET_i)$, where S_i is the observed score. In the case of a student who did not fail the grade, the gain is given by $ATENT_i \ / \ S_i$.

Chart 5.1A shows that the distribution of the effects of grade failure at ISCED 1 for the treated is slightly displaced to the left relative to the distribution of the potential effects on the untreated. Thus, the findings in this respect presented in table 5.1.1 for the mean hold in fact for the entire distribution. Furthermore, it is estimated that the impact of repetition at ISCED 1 is negative over the whole distribution for both groups of students. Regarding retention at ISCED 2, chart 5.1B indicates clear differences between the distributions of effective treatment gains for repeaters and potential gains for non-repeaters. While for the latter group there are losses virtually throughout the entire distribution, it is estimated that most repeaters reap benefits from treatment.

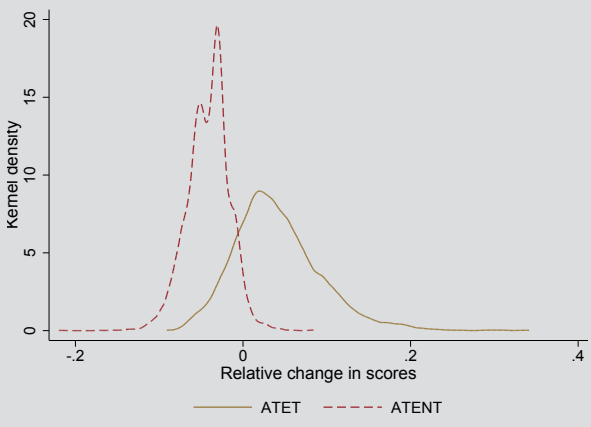
Pereira and Reis (2014) show that these findings are robust to changes in the instruments[16], despite some sensitivity in the size of estimated of treatment effects.

Chart 5.1 A • Distribution of the effects of grade retention at ISCED 1 for repeaters (ATET) and non-repeating (ATENT) in Portugal, scores in reading



Source: Authors' calculations.
Note: Distribution of the ratio $ATET_i/(S_i\text{-}ATET_i)$ for repeaters, and the ratio $ATENT_i/S_i$ for non-repeaters, where S_i is the observed score and $ATET_i$ and $ATENT_i$ the effect of the treatment on student *i*.

Chart 5.1 B • Distribution of the effects of grade retention at ISCED 2 for repeaters (ATET) and non-repeating (ATENT) in Portugal, scores in reading



Source: Authors' calculations.
Note: Distribution of the ratio $ATET_i/(S_i\text{-}ATET_i)$ for repeaters, and the ratio $ATENT_i/S_i$ for non-repeaters, where S_i is the observed score and $ATET_i$ and $ATENT_i$ the effect of the treatment on student *i*.

5.2. Interpretation of results

As mentioned above, the treatment effects in this article are estimated from a sample of students with approximately the same age and attending different grades. Retained students attend a lower degree than they would without retention. This tends to impact negatively on estimates of treatment effects reflecting differences in curricula (adding to other potential impacts such as through instruction time, teacher quality, or even educational trajectories). It is not possible to isolate a specific effect of grade retention via the academic degree on performance in PISA, but its magnitude should be substantial.[17]

The estimates presented in table 5.1.1 indicate, firstly, a very negative effect of retention at ISCED 1 on performance in PISA. As the PISA tests are taken around the end of ISCED 2, this is a long--term effect (say, between 4 and 9 years after grade failure). Such result contradicts the common perception that grade failure at an early stage of the educational career may be beneficial (or, at least, relatively less harmful), to the extent that it gives the child an opportunity to catch up with the necessary maturity levels[18] (for instance, Tomchin and Impara, 1992, for the United States). Note that the empirical evidence tends to contradict this perception, with several studies finding negative effects of early retention on long-term performance (Baenen, 1988, Pagani *et al.*, 2001, among others). The results in this article are thus generally in line with such literature, but the evidence presented in table 5.1.1 provides yet other elements of interest in this regard.

The least squares estimate of the impact of grade failure at ISCED 1 has a similar magnitude as the estimates correcting for the endogeneity of treatment. In fact, the mean treatment effect estimated by the Heckman control function is only slightly lower (in absolute terms) than the strongly negative least squares estimate, both for reading and math. The respective confidence intervals intersect. In the case of retention during ISCED 2, by the contrary, the correction of endogeneity brings about a much less negative mean treatment effect. Consistently, the covariances between the error terms in outcome equations and selection equation for ISCED 1 regressions, are negative but on the brink of non-significance (parameter $\sigma(u_R, e)$) or not significant at all (parameter $\sigma(u_{NR}, e)$). Again this in contrast to the corresponding estimates for retention at ISCED 2, which are negative[19] and highly significant, as expected in the presence of endogeneity. Remember that these parameters capture the co-movement between the unobservable factors in the primary and selection equations. While such results apply to the countries in the sample as a whole and not specifically to Portugal, it should be noted that the unobservable factors should be uncorrelated with country-related aspects (captured by the inclusion of respective fixed-effects).

Results suggest a low degree of endogeneity in the selection of students for retention during ISCED 1. In other words, unobservable factors that determine the performance of students at a later stage of the educational career seem not to play an important role in the selection of students for retention at an early stage of it. The fact that the actual losses of repeaters (ATET) are slightly higher than the potential losses of non-repeaters (ATENT) also indicates a failure in selection for treatment. Note that this phenomenon is not related to observable characteristics of students. With regard to those, it is shown in section 5.3 that the students who benefit most from treatment are generally those more likely to be chosen, in accordance with the evidence in section 3.

In conclusion, the findings in this article indicate that early retention - a practice particularly prevalent in Portugal (see Table 2.1.1) - seems to be detrimental to educational performance in the long run. Thus, there may be an advantage in replacing to some extent this practice for alternative support programs. In Portugal, an example of this type of actions is the program «Mais Sucesso Escolar» whose goal is to support projects to prevent and combat school failure in basic education.

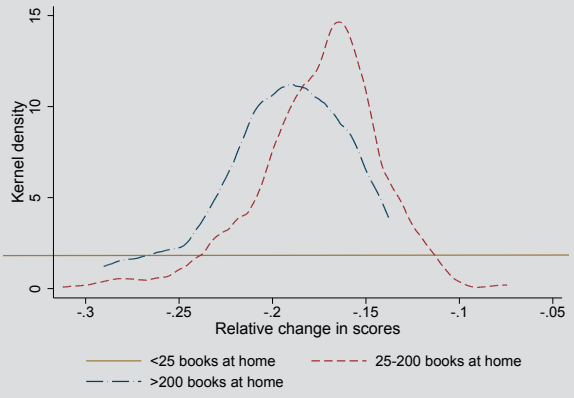
Furthermore, the results suggest that the mechanism of selection could not effectively scrutinize at an early stage of school life which children would benefit from treatment in the long-term. Two sorts of reasons may explain this last result. On the one hand, some unfavorable characteristics of younger students, notably, immaturity, may be overcome in later stages, and cease to play an important role for performance. The data nonetheless exclude that this process is a consequence of treatment, *i.e.* grade failure at ISCED 1 could have led students to change personality traits with a negative impact on performance. Indeed, if this had happened, repeaters should reap more benefits (less losses) from treatment than non-repeaters, which is not the case. On the other hand, the selection of repeaters at ISCED 1 is based on less information and entails more subjective judgment by the participants in the process (notably, teachers and parents) than at a later stage. Recall, for example, that at the early grades of ISCED 1 such a decision is primarily based on the assessment of a single teacher.

With regard to grade failure during ISCED 2, the estimates in table 5.1.1 indicate a small positive effect on performance in Portugal. Our results are in line with recent studies such as Gary-Bobo *et al.* (2014) for France and Baert, *et al.* (2013) for Belgium. Older literature as Hagborg, *et al.* (1991) tended to find negative effects of repetition, particularly at a later stage. Given that PISA tests are taken around the end of ISCED 2, in this case grade retention precedes the tests by around 1 to 3 years. Therefore one cannot rule out that the measured positive effects of grade failure are limited to the years immediately after treatment. There are several studies concluding that performance gains from retention are confined to the short-term (*e.g.* Mantzicopoulos *et al.*, 1992, and Roderick and Nagaoka, 2005). PISA assesses the use of school knowledge from a practical standpoint, not as formal testing. This suggests the possibility of persistence in treatment gains, but one cannot draw firm conclusions on this point.

5.3. Effects of retention by characteristics of repeaters

In this section we present some evidence on the effects of grade retention on treated students conditional on their observable characteristics. We consider socio-economic background of students, gender and living with parents. Only the findings for reading scores are discussed, as math scores give very similar indications.

Chart 5.3.1 A • Distribution of the effects of grade retention at ISCED 1 by the socioeconomic background of repeaters, scores in reading



Source: Authors' calculations.
Note: Distribution of the ratio $ATET_i/(S_i\text{-}ATET_i)$, where S_i is the observed score and $ATET_i$ the effect of the treatment on student i (repeater), for the categories of the variable books at home.

Chart 5.3.1 B • Distribution of the effects of grade retention at ISCED 2 by the socioeconomic background of repeaters, scores in reading

Source: Authors' calculations.
Note: Distribution of the ratio $ATET_i/(S_i\text{-}ATET_i)$, where S_i is the observed score and $ATET_i$ the effect of the treatment on student i (repeater), for the categories of the variable books at home.

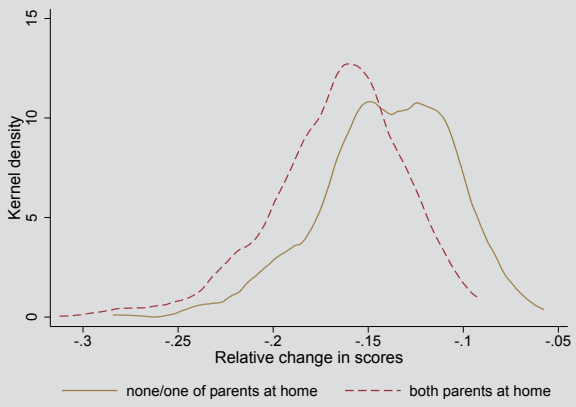
The socio-economic background is measured by the number of books at home (an indicator commonly used in this context) considering three cohorts. Charts 5.3.1A and 5.3.1B show that the gains (losses) of repetition are larger (smaller) for students from disadvantaged families for both levels of education considered. For ISCED 2, the treatment produces gains of 0.3, 3.2 and 5.6 percent, on average, in the scores of treated students, from the highest cohort to the lowest. The difference between the outer cohorts is statistically significant. Regarding retention at ISCED 1, the corresponding figures are -19.4, -17.5 and -14.7 percent; in this case, negative treatment effects throughout the distribution are estimated even for the most disadvantaged students. Recall that in section 3 we showed that students from disadvantaged families were more likely to be selected for retention; it turns out that such students also benefit most (or lose least) with treatment.

With regard to gender, the change in scores of female students subjected to treatment is more positive than that of male students (not shown). The treatment gains are, respectively, 6.0 and 2.5 percent, on average, for retention at ISCED 2 (-14.6 and -17.4 percent at ISCED 1). The differences between such estimates are, however, not statistically significant. Although male students tend to benefit less (or lose more) with treatment once chosen, the probability of being so is relatively higher (see section 3).

Finally we consider differentiated effects of grade repetition depending on the situation of repeaters in terms of living with their parents (Charts 5.3.2A and 5.3.2B). Scores of those who do not live with both parents vary more positively with treatment (this is clearer for ISCED 1 than for ISCED 2). Treatment brings about changes of -14.0 and 5.5 percent in the scores of repeaters, respectively, at ISCED 1 and ISCED 2 (-16.6 and 3.7 percent for those who live with both parents).

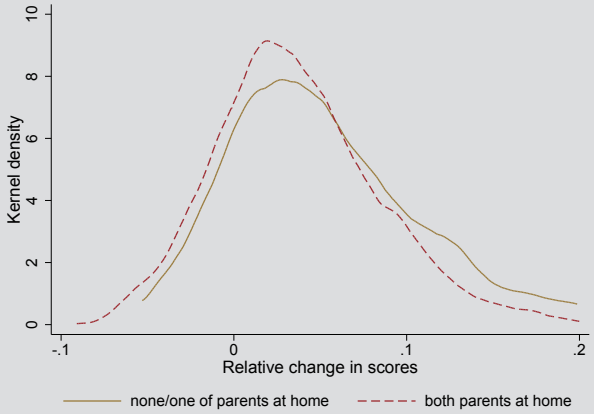
In general, the variation in the effects of repetition as a function of the analyzed characteristics goes in the same direction, whether repetition occurs at ISCED 1 or 2. Moreover, with regard to family variables, selection tends to encompass those who benefit more (or lose less) with treatment. Thus, as noted, the slightly more negative estimate for ATET vis-a-vis ATENT obtained for ISCED 1 (Table 5.1.1 and Chart 5.1A) seems to reflect the interaction with unobservable characteristics.

Chart 5.3.2 A • Distribution of the effects of grade retention at ISCED 1 by the situation of repeaters in terms of living with parents, reading scores



Source: Authors' calculations.
Note: Distribution of the ratio $ATET_i/(S_i\text{-}ATET_i)$, where S_i is the observed score and $ATET_i$ the effect of the treatment on student *i* (repeater), for the categories of the variable living with parents.

Chart 5.3.2 B • Distribution of the effects of grade retention at ISCED 2 by the situation of repeaters in terms of living with parents, reading scores



Source: Authors' calculations.
Note: Distribution of the ratio $ATET_i/(S_i\text{-}ATET_i)$, where S_i is the observed score and $ATET_i$ the effect of the treatment on student *i* (repeater), for the categories of the variable living with parents.

5.4. Comparison with the estimates for a set of European countries

We now make a comparison between the effects of grade failure in Portugal and in the group of European countries that were the basis for estimations. Table 5.4.1 shows the average treatment effects for the entire sample.[20]

In general, the results for the set the European countries considered are broadly consistent with those obtained for Portugal. On the one hand, grade repetition at ISCED 1 has very negative effects for those submited to treatment, in contrast to the positive effects of repetition at ISCED 2. Moreover, the difference between the ATET and ATENT for ISCED 1 is statistically not significant (the estimate of the second parameter being less negative than the estimate of the first one). The magnitude of treatment benefits for repeaters at ISCED 2 in table 5.4.1 is, however, higher than previously estimated for Portugal. More generally, Pereira and Reis (2014) show that country fixed-effects interact strongly with the indicator of repetition. This may be explained, among other factors, by the differences among policies of support to repeaters implemented in the different education systems, an issue discussed in more detail in that study.

Table 5.4.1 • Average impact of grade retention during basic education in a set of European countries

		ISCED 1				ISCED 2			
		Reading		Mathematics		Reading		Mathematics	
Heckman	ATE	-59.9	(8.8)***	-64.4	(8.9)***	4.1	(6.4)	-6.9	(6.0)
control	ATET	-63.7	(7.2)***	-76.5	(6.7)***	49.0	(5.6)***	42.0	(5.2)***
function	ATENT	-60.1	(8.5)***	-65.0	(8.6)***	7.8	(6.0)	-2.8	(5.6)

Source: Authors' calculations.
Note: The treatment effects are obtained as described in section 4.1; the table presents averages for all students in the sample. Standard deviations obtained by bootstrap in parentheses.
* Significant at 10%; ** Significant at 5%; and *** significant at 1%.

6. Conclusions

This article studies the determinants of grade retention during basic education and analyzes its effects on student performance in Portugal compared to Europe, using data from OECD PISA 2003 and 2009. The main conclusions are as follows.

- The results point to the importance of individual characteristics, family and peers as determinants of grade retention. In particular, students with less maturity and worse economic conditions are more likely to repeat in Portugal.

- Despite the importance of socio-economic aspects, there are other factors that help explain grade failure. The results suggest the importance of school characteristics, regional differences and differences at the country level (*e.g.* institutional features).

- The long-term effects of repetition at ISCED 1 on student performance in Portugal are negative, which suggests that there will be advantage in replacing this practice, at least partially, for alternative ways of supporting students who have learning difficulties in the early stages of school life.

- Students who are most likely to be retained during ISCED 1 on the basis of their observable

characteristics (*e.g.* socio-economic factors) tend to be less penalized with treatment. In contrast, at this early stage of schooling, there seems to be difficulties in identifying the students whose unobservable characteristics would enable them to achieve gains in performance over the long-term.

- The effects of short-term repetition at ISCED 2 in Portugal are positive albeit small. Therefore, despite the uncertainty about the long-term effects, our results do not call into question the practice of retention for higher levels of schooling. In addition, there is an alignment between selection for treatment and treatment benefits, both in regard to observable and unobservable characteristics of students.

- The results for the set of European countries considered in the sample are broadly consistent with those for Portugal, but there is a more positive effect of student retention at ISCED 2 in the first case.

REFERENCES

Angrist, J. and **Krueger**, A. (1991), "Does Compulsory School Attendance Affect Schooling and Earnings?", *Quarterly Journal of Economics*, 106(4), pp.979-1014.

Baenen, N. R. (1988). Perspectives after five years – Has grade retention passed or failed?, *Paper presented at the Annual Meeting of the American Educational Research Association*, New Orleans, LA.

Baert, S., **Cockx**, B. and **Picchio**, M. (2013), "On Track Mobility, Grade Retention and Secondary School Completion", Ghent University, manuscript

Bedard, K. and **Dhuey**, E., (2006), "The Persistence of Early Childhood maturity: International Evidence of Long-Run Age effects", *Quarterly Journal of Economics*, 121(4), pp.1437-1472.

Blundell, R. and **Costa-Dias**, M. (2009), "Alternative Approaches to Evaluation in Empirical Microeconomics", *Journal of Human Resources*, 44 (3), pp.565-640.

Cooley, J., **Navarro**, S. and **Takakashi**, Y. (2011) "How the Timing of Grade Retention Affects Outcomes: Identification and Estimation of Time-Varying Treatment Effects", *Centre for Human Capital and Productivity Working Papers 20117*, University of Western Ontario

Diris, R. (2012) "Estimating the Impact of Grade Retention: A Heterogenous Analysis", manuscript.

Eurydice (2013) "The structure of the European education systems 2012/13: schematic diagrams", European Commission.

Elodie, A. (2009) "Is grade repetition a second chance?", Toulouse School of Economics, manuscript

Fertig, M., and J. Kluve (2005). The effect of age at school entry on educational attainment in Germany, IZA DP 1507.

Garcia-Perez, J., **Hidalgo-Hidalgo**, M., and Robles-Zurita J. (2014) "Does grade retention affect students' achievement? Some evidence from Spain", *Applied Economics*, 46(12), pp.1373-1392.

Gary-Bobo, R., **Gousse**, M, and **Robin**, J.M. (2014) "Grade Retention and unobserved heterogeneity", Cemmap working paper, CWP14/14

Hagborg, W. J., G. **Masella**, P. **Palladino**, and J. **Shepardson**. (1991). "A Follow-Up Study of High School Students with a History of Grade Retention", *Psychology in the Schools*, 28(4), 310–317.

Heckman, J. J. (1978). "Dummy endogenous variables in simultaneous equation systems", *Econometrica*, 46, pp. 931-960.

Heckman, J. J. (1979). "Sample selection bias as a specification error", *Econometrica*, 47, pp. 153-161.

Jacob, B. and **Lefgren** L., (2009), "The Effect of Grade Retention on High School Completion" *American Economic Journal: Applied Economics*, 1(3), pp.33-58.

Mahjoub, M. (2009), "The treatment effect of Grade Repetitions", Paris School of Economics, manuscript.

Mantzicopoulos, P. Y., D. C. **Morrison**, S. P. **Hinshaw**, and E. T. **Carte** (1989). "Nonpromotion in kindergarten: The role of cognitive, perceptual, visual-motor, behavioral, achievement, socioeconomic, and demographic characteristics", *Educational Research Journal*, 26(1), 107–121.

Ministério da Educação (2013) "Educação em Números - Portugal 2013", Direção Geral de Estatísticas da Educação e Ciência (DGEEC)

OECD (2012), "Equity and Quality in Education: Supporting Disadvantaged Students and Schools", *OECD Publishing*.

Pagani, L., **Tremblay**, R., **Vitaro**, F., **Boulerice**, and B., **Mcduff**, P. (2001). "Effect of grade retention on academic performance and behavioral development". Development and Psychopathology. 13, 297–315.

Pereira, M. C. and **Reis**, H. (2012), "Diferenças Regionais no Desempenho dos Alunos Portugueses: Evidência do Programa PISA da OCDE", *Boletim Económico de Inverno*, Banco de Portugal.

Pereira, M. C., and **Reis**, H. (2014). The effects of grade retention on student achievement: evidence from a multi-country dataset, Working Paper do Banco de Portugal (forthcoming).

Peterson, S. E., J. S. **DeGracie**, and C. R. **Ayabe** (1987). "A Longitudinal Study of the Effects of Retention/Promotion on Academic Achievement" *American Educational Research Journal*, 24(1), 107–118.

Robertson, E. (2011) "The Effects of quarter of birth on academic outcomes at the elementary school level", *Economics of Education Review*, 30, pp.300-311

Roderick, M., and J. **Nagaoka**. (2005). "Retention under Chicago's High-Stakes Testing Program: Helpful, harmful, or harmless?", *Educational Evaluation and Policy Analysis*, 27(4), 309–340.

Schwerdt, G. and **West**, M., (2012) "The Effects of Test-Based Retention on Student Outcomes over Time: Regression Discontinuity Evidence from Florida", *Program on Education Policy and Governance Working Papers Series - PEPG* 12-09

Tomchin, E. M., and J. C. **Impara** (1992). "Unraveling teachers' beliefs about grade retention", *American Educational Research Journal,* 29, 199-223.

Vella, F. and M. **Verbeek** (1999). "Estimating and interpreting Models with endogenous treatment effects", *Journal of Business and Economic Statistics*, 17(4), 473-478.

Wooldridge, J. M. (2002). Econometric Analysis of Cross Section and Panel Data, Cambridge, MA: The MIT Press.

Notes

1. The authors would like to thank Isabel Horta Correia, Jorge Correia da Cunha and Vanda Cunha for their comments and discussion. The analyses, opinions and findings of this article represent the views of the authors, which are not necessarily those of the European Central Bank, the Eurosystem of Central Banks, the Banco de Portugal or any of the institutions with which we are affiliated.

2. Banco de Portugal, Economic Research Department.

3. The number of grades in each level is similar but not necessarily the same across countries. For more details, see Eurydice (2013).

4. These countries are: Austria, Germany, Belgium, Denmark, a Slovakia, Spain, Finland, France, Greece, Hungary, Ireland, Italy, Latvia, Liechtenstein, Luxembourg, the Netherlands, Poland, Portugal, the Czech Republic, and Sweden, for which there is information available for 2003 and 2009. In 2009 there are also data for Bulgaria, Slovenia, Estonia, Lithuania and Romania.

5. Using means at the level of the PISA stratification variable that has in most cases a regional component. For countries where this is not the case, we use means at the level of the school location variable (village, small town, town, city and large city).

6. In this article we use the mean of the five plausible values for each discipline. Note that in the estimations the variance is generally calculated by bootstrap.

7. Despite some exceptions, in Portugal the cut-off date is September 15 of the year the student turns out 6 years old.

8. In most countries the school years starts in September.

9. We are assuming the existence of low mobility of students across schools and regions, which should be reasonable for most countries as we are considering basic education levels only.

10.The study of the impact of grade retention on performance for this group of students would require a slightly different methodological approach than the one presented in section 4, notably multiple treatment effects.

11.Thus one is just imposing a common influence of the instruments on the propensity to repeat in Portugal and the other countries in the sample.

12.For the calculation of this variable, we took averages at the level of the PISA stratification variable (which in most cases has a regional component) and the school location indicator.

13. More accurately, in the case of maturity (measured in months) as the model includes, in addition to the country fixed-effects, the entry age in compulsory education (in years), the source of exogenous variation is the variability of maturity within country and entry age cohorts.

14. For example, for Germany, Kluve and Fertig (2005) write *«Regulations determining maturity, and hence enrolment and deferment decisions, are somewhat vague: in some cases parental application is sufficient, in some cases approval by the school and/or a public health officer is required, and sometimes decisions are based on a test. This leads to the fact that there is possible variation in enrolment practices over time and across federal states, and even between neighboring schools.»*

15.Note that even for this estimator *ATET* and *ATENT* differ due to the existence of interactions of grade retention with various covariates.

16.For example, the consideration of retention practices in the regressions for ISCED 1, or the maturity indicator, separately, for ISCED 2.

17.It is possible to get an idea of such magnitude by running regressions for repeaters and non-repeaters of scores on the academic degree and covariates included in the education production function (but excluding entry age which is precisely meant to capture, along with grade failure, the impact of academic degree). It is estimated that attending the degree immediately above can bring gains of around 40 points for repeaters (ISCED 1 and ISCED 2) and 30 points for non-repeaters (reading scores).

18.This perception by teachers about beneficial effects of retention typically covers the kindergarten and early primary grades.

19.The negative sign of these covariances stems from the fact that important unobservable factors explaining performance (say, ability and motivation of the student) also play a role in the selection equation, but with the opposite effect.

20.The differences to the estimates presented in Pereira and Reis (2014) are justified by the fact that the latter are based on a model without interactions for Portugal.

Appendix 1 - Summary Statistics - peer, school and regional characteristics (mean)

	ISCED 1		ISCED 2	
	Total	Portugal	Total	Portugal
Peer variables				
Books at home (25-200) - peers	0.52	0.49	0.52	0.49
Books at home (>200) - peers	0.25	0.16	0.24	0.16
Lower secondary education - peers	0.09	0.20	0.09	0.20
Upper secondary education - peers	0.45	0.22	0.46	0.22
Terciary education - peers	0.41	0.26	0.40	0.26
Single-parent family - peers	0.14	0.14	0.14	0.14
No parents at home - peers	0.01	0.03	0.01	0.03
Immigrant - peers	0.07	0.05	0.07	0.05
Educational resources at home - peers	5.00	4.86	4.99	4.87
White collar - low skill - peers	0.23	0.29	0.24	0.29
White collar - high skill - peers	0.53	0.36	0.52	0.36
School variables				
located town 15000- 100000 inh. (b)	0.40	0.43	0.40	0.43
located town > 100000 inh. (b)	0.28	0.21	0.29	0.21
Percentage of girls	24.54	24.12	25.37	24.76
Grade amplitude (max-min grade)	7.53	6.28	5.19	5.16
Class size	22	21	23	22
Student/teacher ratio	12.60	9.49	12.98	9.83
Percentage of computers with web	0.82	0.82	0.86	0.77
School educational resources	0.00	0.00	0.00	-0.11
Teacher participation	0.21	-0.33	0.10	-0.43
Teacher shortage	-0.06	-0.53	-0.11	-0.65
School Size	540	808	630	983
Private school (b)	0.23	0.20	0.15	0.10
Ability grouping between classes (b)	0.64	0.62	0.61	0.52
Admission criteria: residence (b)	0.40	0.48	0.37	0.38
Admission criteria: academic record (b)	0.18	0.02	0.22	0.01
Admission criteria: feeder schools (b)	0.14	0.02	0.18	0.01
Admission criteria: Religious philosophy	0.14	0.21	0.10	0.14
School responsibility: resource allocation	2.92	2.37	2.68	2.01
School responsibility: curriculum and assessment	3.08	2.32	2.94	1.99
Student absenteeism (b)	0.09	0.08	0.10	0.12
Students with disruptive behaviour (b)	0.07	0.07	0.06	0.07
Students skipping classes (b)	0.06	0.06	0.06	0.09
Students being bullied (b)	0.02	0.00	0.01	0.00
Math regulat lessons	3.31	3.85	3.19	3.79

Source: PISA 2003 and2009.
Note: (b) stands for binary variable.

Appendix 1 - Summary Statistics - peer, school and regional characteristics (mean)

	Total	Portugal
Regional characteristics		
Books at home (25-200) - region	0.52	0.49
Books at home (>200) - region	0.24	0.16
Lower secondary education - region	0.10	0.20
Upper secondary education - region	0.46	0.22
Terciary education - region	0.40	0.26
Single-parent family - region	0.14	0.14
No parents at home - region	0.01	0.03
Immigrant - region	0.07	0.05
First language not test language - region	0.04	-
Foreign language at home - region	0.04	0.01
Educational resources at home - region	4.98	4.86
White collar - low skill - region	0.23	0.29
White collar - high skill - region	0.52	0.35
Attitude towards school - region	-0.04	0.31
Relation with other students - region	-0.09	0.31
Retention practices	0.09	0.19

Source: PISA 2003 and 2009.
Note: (b) stands for binary variable.