Does the Age of Entry in Primary School affect Student's Achievement?

M. MADEIRA
Nova School of Business and Economics
Universidade Nova de Lisboa

A. B. REIS
Nova School of Business and Economics
Universidade Nova de Lisboa

L. C. NUNES
Nova School of Business and Economics
Universidade Nova de Lisboa

M. C. SEABRA
Nova School of Business and Economics
Universidade Nova de Lisboa

The school starting age is one the factors that influence student's educational attainment. Using a large dataset containing information of students from public schools in Portugal, it was possible to establish a positive relation between the entry age and educational outcomes, such as retentions and exam's grades. However, deferring a child's entry by one year does not seem to provide large benefits for students. This paper complements the analysis by understanding the main characteristics that lead to child's deferment and reinforces the idea that educational policy is a complex issue for Governments to deal with.
Keywords: school starting age, student outcomes, deferment of school entry

INTRODUCTION AND MOTIVATION

The beginning of primary school is an important milestone in children's and their parents' lives. However, the optimal age that children should have at such moment has been the subject of intense debates on the part of many agents due to the multiple effects it can have on student's future development.

Nowadays, most educational systems use cut-off dates to define the eligibility of children to enrol in primary school. These cut-off limits state that children who achieve a given age by a specific date of the year are eligible to enrol in school. In Portugal, the enrolment is mandatory for children who turn six years old by September 15th, while the remaining students, born after the threshold date, are considered conditional – they may start school with five years old if their parents enrol them and if the school has the capacity to accept and include them in an already constituted class.[1]

In a given academic year all students start school on the same day regardless of their birth date, which means that in the beginning there will be students who are older than their colleagues by up to fifteen and a half months of age difference. The relative age difference automatically created in the beginning of the year is believed to have an impact in future outcomes, as relatively older students may be physical or mentally more developed than their younger counterparts, having therefore more learning abilities.

The study of this phenomenon becomes relevant if the differences in maturity have a negative effect in the performance of students that started school at a younger age, placing them in a disadvantageous position. Such initial disadvantage is likely to have even worse outcomes if the differences in maturity caused by age have an impact, not only in the short run, but also in the medium and long term with effects in college participation or labour market outcomes.

Adding to the importance of this analysis for students and their families, this particular issue also carries a significant economic and political consideration for nations: a later entry into primary school may entail later entry into the labour market, thereby decreasing lifetime income and discounts. Due to the inefficiencies and large economic costs involved, state interventions may occur if it leads to better educational outcomes.

The aim of this report is to use the Portuguese data to study the impact of absolute and relative age differences in educational outcomes, particularly in retentions and fourth grade national exams' scores, as several authors developed for other countries.[2]

An additional goal of this report is to understand whether some characteristics make some students more likely to defer their enrolment in primary school. Such analysis becomes relevant as the Portuguese cut-off rule is not strictly imposed, meaning that parents of conditional students have some degree of liberty to choose the year in which their children start school.

1 According to *Despacho Normativo 1-B/2017*

2 Absolute age differences are used when the different groups of students under analysis have about one year of age difference. Relative age differences measure the impact of a few months' difference, using age as a continuous variable in relation to the cut-off date.

The remaining report is structured as follows: in section two is presented a brief summary of the literature related with this topic, which was used as support for this analysis. Section three describes the data and presents the methodology. In section four results are presented and some important outcomes are highlighted. In the last section, the main conclusions are drawn and future extensions of this study are suggested.

LITERATURE REVIEW

The relevance of this subject led several authors to investigate the impact of school starting age in the educational attainment of students in the short and long term, analysing such effect in retentions, exam scores, college enrolment and labour market outcomes.

One of the main concerns of this analysis is whether children that start school at a younger age have less skills than their older counterparts. Becker (1964) was among the first authors to analyse the life cycle skill formation and the determinants of earnings. More recent papers, such as the one by Cunha et al (2006), try to go deeper into the primordial approach by Becker (1964) stating that as younger children are more likely to accumulate less skills by the time they start school, such difference may never be cancelled through the learning process.

Among the great contributors to the question of the relative age difference effects are Angrist and Krueger (1991). These authors were concerned about the effect of compulsory school attendance laws on educational attainment for students in the United States.[3] They used an innovative approach, important for most recent literature: by considering that birth dates are unlikely to be correlated with other personal characteristics, they would be able to capture the exogenous variation of age in educational outcomes by using the quarter-of-birth of student as an instrument for student's education. They concluded that older students tend to have less education (less capital accumulation) and subsequently less earnings.

Following Angrist and Krueger's contributions, a vast literature has been produced for several countries to try to understand the effects of age in student's performance. Assuming an exogenous variation of age (due to the different birth dates and cut-off dates), Crawford et al (2007) for England and Kawaguchi (2011) for Japan concluded that students that begin school younger perform worse than their older counterparts in primary school's results and that the relative age effect declines in the following years, being the effect still significant.

Similar analysis were also performed by Fredriksson and Ockert (2004) for Sweden, Puhani and Weber (2007) for Germany and Elder and Lubotsky (2009) for United States. These authors went further in the assumption of exogeneity, and tried to deal with the endogeneity problem by using an instrumental variable framework that defined the expected age of a given individual as an instrument for his actual age. Elder and Lubotsky (2009) concluded that students that start school at a younger age tend to have some disadvantages in the beginning, which fade away with progression in school, and that having older students in the class

[3] Angrist and Krueger (1991) made the analysis for United States, where students are allowed to leave school as soon as they turn sixteen years old. In other educational systems, like the Portuguese one, there are a number of years of schooling that students must complete to be able to drop school.

increase the probability of retentions. The other mentioned authors found similar results, however not only in primary school outcomes, but also in middle and secondary school. For instance, Puhani and Weber (2007) state that students that enter school at seven instead of six years old have a higher probability of attending a more academic rather than vocational track in high school in Germany.

Fertig and Kluve (2005) were also important for the study of this topic, as they perform the analysis for East and West Germany at a time of changing regulation. These authors performed a very complete analysis using age at school entry as a continuous regressor and using a binary variable (distinguishing children that started either at six or seven years old). The empirical methods used by these authors to test the relative age effects were Linear Probability Model and Matching estimations, which corroborated the conclusion that students that start school at six years old perform worse than remaining classmates. However, they concluded that such outcomes were due to unobserved heterogeneity. Therefore, they improved their estimation with an Instrumental Variable approach using relative age as instrument, not finding any effect in this case.

Moreover, Bedard and Dhuey (2006), through the same Instrumental Variable estimation strategy, conducted an analysis for a range of OECD countries and the results are consistent with the ones already presented, concluding simultaneously that students that start school at a younger age have less education than the older ones. These authors were also able to conclude that early disadvantages persist into college enrolment decisions. These long-term results are also obtained by Ponzo and Scoppa (2014)[4], when analyzing for Italian students, but contradicted by Black et al (2011)[5] for Norway.

Another major concern regarding the school starting age relates to the question of early tracking in many countries. Early tracking occurs a few years after primary school and indicates whether students should go for a more academic or vocational program. Muhlenweg and Puhani (2010)[6] are among the many researchers that empirically studied this issue. They concluded that age differences have an impact in early tracking, which means that long-term economic effects are propagated in the long-run and future earnings.

Concerning labour market outcomes, different results were also obtained. While Black et al (2011) state that students that started school at a younger age might have a positive (small) impact on labour market income, Fredriksson and Ockert (2004) state that starting school younger may entail a long-run earning disadvantage and Dobkin and Ferreira (2010)[7] find no effect of age in labour market outcomes.

Despite all ambiguous results, there seems to be a consensus among studies that students that start school at a younger age tend to reach worse educational performances. Whether the analysis is performed through the exogenous variation that birth dates generate, the use of relative assigned age as an instrument, or more advanced frameworks, the tendency is for

4 These authors also use an Instrumental Variable framework.

5 These authors also use an Instrumental Variable framework.

6 These authors also use an Instrumental Variable framework.

7 These authors use a RDD approach

students that start school at an older age to perform better in the different circumstances of analysis.

One shortcoming of all these results is that all of them are under the assumption that there are no heterogeneous effects nor pre-selection of time of child's birth. Peña (2016)'s work for Mexico tries to overcome such weakness and develops a Difference-in-Difference approach (which he later compared with IV and RDD approaches). He concludes that age confers older students an advantage, however, if tested at exact same time, students that start at a younger age would outperform their older counterparts.

Relevant papers for this analysis try to deal with the endogeneity problem mostly using an instrumental variable approach. However, some recent studies state that two main assumptions might not hold for this practice: the monotonicity and randomness of birth dates. Such criticism is presented in papers by Attar and Cohan-Zada (2017) or Barua and Lang (2008) that intend to estimate the causal effect of entrance age on test scores, isolating this way any effect from the date of births and strictly satisfying the monotonicity assumption.

Another major concern in the literature of the impacts of age in children's outcomes is related with retentions. Portuguese authors' Coutinho and Reis (2014) studied the determinants of grade retention on basic education and later academic development for Portugal (and comparing with other European countries). Using indicators referent to the age, relative to the cut-off date, or a maturity indicator[8], the authors concluded that entering one year later in primary school decreases the probability of being retained, and that the maturity of non-repeaters is greater than the repeaters'. Also, for Portugal, students with less maturity and from a worse socioeconomic environment have a greater probability of repeating a grade.

Departing from the age effects in Bedard and Dhuey's (2008) paper, the distribution of skills across socioeconomic groups may also be a concern. If parents that defer their children's entrance in primary school are the ones from high socioeconomic groups, then the learning process becomes harmful for lower socioeconomic groups: not only they have less economic power, they are also the youngest in class, which may have a double negative effect in terms of skills acquired and children development. Sustaining this point, Dobkin and Ferreira (2010) for California and Texas, state that students born after the cut-off are the ones more likely to delay entrance and that minority groups are more likely to stick with the law, starting earlier more often and creating this way a disproportionate share of these students among the youngest in a given cohort.

DATA AND METHODOLOGY

This report aims at answering one main research question – Does school starting age affect children's performance (in retentions or national exam's scores)? – and one complementary question – Which characteristics influence children's early enrolment?

[8] Based on Bedard and Dhuey (2006)

In order to answer both questions, two main datasets were used. The first one is MISI, useful for both analysis, which contains microdata with detailed information from students since 2006 (whether personal information such as age, gender and parent's core information or academic-related information such as grades, retentions and transfers). For this study only students from public schools were used, as private and state-funded private schools' databases do not display all necessary information.

Concerning the impact of starting age in retentions, only MISI data for public schools is necessary. However, when studying the effect in national exams scores, there is the need to merge the information from MISI with another comprehensive database, denominated ENEB, which contains anonymous information provided by the Ministry of Education regarding students and their exam performance. For this report, only the fourth grade national exams from 2013, 2014 and 2015 were used, as the results are presented in a 0-100 range, rather than the 1-5 range provided for the remaining years.[9] Having a wider range of scores allows for a more precise analysis and perception of results.

As the Portuguese cut-off date is not strictly imposed and parents have some discretion in the choice of deferment or anticipation of their children's start in primary school, it becomes relevant to analyse the students' characteristics that can influence such choice. The significant characteristics will then be used as controls when measuring the effects of school starting age in retentions and exam scores.

Moreover, for the introductory analysis, data used is from children born between September 16th and December 31st and that started primary school in a public school from 2007/2008 academic year until 2015/2016 (the last data available).[10] There are a total of 223,743 observations.[11] The variable *Early_Enrolment*, used in the models explained below and only defined for students born in the specified range, is a binary variable which represents 1, in case the student started school at five years old, and 0, if the student started at six years old (children whose parents' deferred primary school's entry). The variable *age_year6* represents the age of students, measured in months, at the cut-off date in the year they turn six years old - meaning that for this sample, this variable varies between 68.47 and 72 months.

The descriptive statistics, presented in Table 1, also display information regarding gender (*female* is 1, if student is a girl, and 0, if it is a boy), child's place of birth (*foreign* equal to 1, if child was not born Portuguese, and 0 otherwise), parents' place of birth (*foreign_mother* and *foreign_father* defined as the variable *foreign)*, parent's qualification (*qualification_mother* and *qualification_father* are described from 0-9 range in ascending order of parent's qualifications), parent's employment status in the child's first grade (*unemp_mother* and *unemp_father* are equal to 1 if the parent was unemployment in child's first grade, and 0 otherwise), a proxy variable of socioeconomic environment, which describes whether the

9 In 2016 there is not any final evaluation test in 4th grade. Due to political changes, from 2015/2016 onwards for primary school there is only a *Prova de Aferição* in 2nd grade.

10 Once 2006/2007 was the first registered academic year, it was not included in this analysis because it does not allow to observe the previous academic year information.

11 Only students with school starting ages between 5 and 10 years old were considered. The percentage of students born in the specified range that started primary school with more than six years old is 0.73%.

student received a subsidy in first grade (*SASE_beneficiary* equal to 1 if child received the subsidy, and 0 otherwise) and information about computer possession (*computer* equal to 1 if children have a computer at home, and 0 otherwise).

Table 1. Descriptive Statistics for students born between September 16th and December 31st

VARIABLES	Obs.	Mean	SD	Min	Max
Early_Enrollment	**223,743**	**0.825**	**0.380**	**0**	**1**
age_year6	**223,743**	**70.28**	**0.085**	**68.48**	**71.97**
Female	223,743	0.482	0.500	0	1
Foreign	223,743	0.036	0.187	0	1
foreign_mother	223,743	0.110	0.312	0	1
foreign_father	223,743	0.114	0.318	0	1
qualification_mother	221,948	3.028	1.974	0	9
qualification_father	218,839	2.527	1.898	0	9
unemp_mother	223,743	0.135	0.342	0	1
unemp_father	223,743	0.074	0.262	0	1
SASE_beneficiary	223,743	0.554	0.804	0	2
Computer	223,743	0.553	0.497	0	1

For further analysis, only the information from the mother are considered, as from the matrix of correlation of the control variables it can be seen that the variables containing the information regarding parent's qualifications are strongly correlated.[12]

From the table displayed below (Table 2) it is possible to note that students beginning first grade in a public school increased from 2007 to 2008, but have been decreasing consistently from that academic year onwards for the sample of students born after September 15th.[13] Still, the share of students that delay entrance in the first grade has been increasing – from around 16.94% in 2007/2008 to 25.79% in 2015/2016). Such increase in the share of students that defer school by one year is unlikely to be due to school constraints, as the total number of children is decreasing. Parents' choices, local characteristics or Government regulatory changes arise as possible explanations for such phenomenon.

Table 2. Distribution of students according to the school starting age and academic year

Academic Year	Start school with 5 years old	Start school with 6 or more years old	% of students starting at 5 years old	% of students starting at age 6 or more	Total
2007/2008	21,949	4,475	83.06%	16.94%	26,424
2008/2009	23,439	3,891	85.76%	14.24%	27,330
2009/2010	22,204	4,015	84.69%	15.31%	26,219
2010/2011	21,612	3,895	84.73%	15.27%	25,507
2011/2012	21,305	4,182	83.59%	16.41%	25,487
2012/2013	20,568	3,973	83.81%	16.19%	24,541
2013/2014	19,017	4,187	81.96%	18.04%	23,204
2014/2015	18,127	4,857	78.87%	21.13%	22,984
2015/2016	16,361	5,686	74.21%	25.79%	22,047
Total	**184,582**	**39,161**			**223,743**

[12] The correlation between *qualification_mother* and *qualification_father* is strong and equal to 0.6369.

[13] The decrease in the number of students is also captured for the whole population data. When using all population, the number of students slightly increases from 2010 to 2011, but the overall trend is decreasing.

A more detailed analysis of the data showed that the number of students that defer entry is larger for students born in December than for students born in September.[14] This fact can partially be explained by the criteria defined by the Portuguese Ministry of Education and used by schools to define the enrolment of conditional students.[15] According to the criteria, relatively older students are the first ones to be accepted if the school has capacity to integrate conditional students.

When concerned about the short-term impacts of school starting age in retentions and exams, the population considered differs from the one used in the introductory question. In this case, only students from the MISI database that are possible to track during the four years of primary education in a public school in Portugal are used. Students who were transferred for or from abroad, or that moved to a private school in primary level are not considered.[16] Since the analysis was built to track students during primary school, and the expected duration of this stage of basic education is four years, from 2007/2008 to 2015/2016 it is possible to define six theoretical cohorts according to the academic year in which children started first grade. To improve the precision of the analysis, the last cohort (2012/2013 to 2015/2016) was eliminated, because in 2015/2016 academic year it was not performed any final evaluation in fourth grade due to changes in educational policy.

Given the rules presented, the analysis of the impact of retentions is done using all students that started school between 2007/2008 and 2011/2012 and that are observed during four years in a public school. The variable *retentions* was constructed as a binary variable that represents 1 if the student failed in any of the grades in primary school, and 0 otherwise.

Concerning the impact in national exams, in order to have a pure effect of age differences, not influenced by other factors (such as retentions), only students that completed primary education in four years are considered.[17]

The control variables used in the estimation for these two analysis are similar to the ones already presented for the introductory question, adding also one variable for the number of students in class and changing the unemployment-related variable. The variables *years_unemp_mother* and *years_unemp_father* are designed as equal to 1 if the parent was unemployed at least one year during child's primary school, and 0 if parents always worked.

For the analysis of the main research question, three different strategies were developed to analyse the impact of school starting age in retentions or national exam's scores:

(1) Use age as continuous variable to explore the relative age differences for students born between January 1st and September 15th. Students in this range are six years

[14] For example, in the 2013/2014 academic year, the percentage of students born in September that defers entry in primary school is of 6.63%, while the percentage of students born in December that are deferred is of 29.22%.

[15] According to *Despacho Normativo 1-B/2017 – Article 10*

[16] The main reason not to consider students transferred for or from abroad during primary school and students that moved from the public to private sector is because they are not possible to track during the four years of primary school. Such cases represent only 1.62% of total students from 2007/2008 academic year to 2015/2016.

[17] Students can be retained in fourth grade due to a poor result in the national exam, but are still analysed here. However, the exam that the same student performs the following academic year will not be considered in the analysis.

old. Even though six years old is the compulsory age, there are students that exceptionally anticipate or delay their entrance in school and that were born in this range. Such students are not considered, nevertheless we remain with 98.47% of the total population (children that started at six years old). The variable that represents student's age, in months, in the beginning of primary school is *starting_age* and it varies between 72 and 80.5 months.

(2) Use data from students that started in a given academic year and that were born either in January or December.[18] With a binary variable (named *January* and that is equal to 1 if student was born in January, and 0 if born in December) the aim is to explore the impact in educational attainment of students that start school at the same time, but that have one year of age difference.

(3) Use students that were born between September 16th and December 31st.[19] These are all the students considered conditional for the civil year in which they would turn six years old. Some of these students started school at five years old (around 85,67%), while others waited one year to start first grade, either through parent's choice or school's incapacity to accept their enrolment, consequently starting at six years old (around 14,33%).

The following table (Table 3) presents the descriptive statistics for all three methodologies, regarding the retentions' analysis.

Table 3. Descriptive Statistics regarding Retentions variable, segmented by the three alternative methods of analysis.

Methodology Scenario	Total Obs.	Nr. of students that never Failed	Nr. of students that failed at least once	% of students that never Failed	% of students that failed at least once
i) Born between January 1st and September 15th	279,897	248,706	31,191	88.86%	11.14%
ii) Born in January or December	58,199	51,369	6,830	88.26%	11.74%
iii) Born between September 16th and December 31st	116,643	99,926	17,247	85.67%	14.79%

As in the previous table presented, Table 4 presents the descriptive statistics for national exams information, segmented by academic year, subject and the three strategies used.

Up to now, the data and the main features of this report were described. However, in order to provide an answer for the research question and study the relation between starting age and the possible outcomes achievable, there is the need to introduce some econometric models. Therefore, an empirical strategy based on a Linear Probability Model was developed, following the approaches of many authors described in the literature review.

[18] Only students that were born in January and started school at 6 years old, and students that were born in December and started primary school with 5 years old.

[19] Students born between September 16th and December 31st should start school either at 5 or 6 years old. However, as in (1), there are students that exceptionally were born in this range and whose starting ages do not correspond to the expected ones. These cases are excluded and 99.42% of information is kept.

Table 4. Descriptive Statistics of National Exams information, segmented by the three alternative methods of analysis.

		Born between January 1st and September 15th			Born in January or December			Born between September 16th and December 31st		
Year	Exam	Obs	Mean	SD	Obs	Mean	SD	Obs	Mean	SD
2013	Portuguese	50,207	51.07	16.90	10,207	50.48	16.88	19,695	48.63	16.45
	Maths	50,199	59.80	19.49	10,203	59.33	19.63	19,700	56.89	19.54
2014	Portuguese	46,973	64.68	15.55	9,691	64.61	15.45	18,711	62.25	15.40
	Maths	46,973	58.45	20.05	9,694	58.19	20.12	18,713	55.69	19.92
2015	Portuguese	46,895	67.59	15.07	9,563	67.27	14.89	18,770	65.28	15.06
	Maths	46,961	61.82	18.60	9,584	61.54	18.55	18,809	59.51	18.44

For the introductory question, in which we are concerned on understanding the determinants of early enrolment for students born between September 16th and December 31st the model established is:

(1) $Early\_Enrolment_i = \alpha_0 + a_1.\mathbf{age\_year6} + \alpha_2.Xi + \varepsilon_i$

The variable *Early_Enrolment* is the dependent variable of interest. *Age_year6* is the main explanatory variable, representing a_1 the effect of this variable in *Early_Enrolment*. The vector X_i describes the control variables such as gender, place of birth, parent's qualification and unemployment status, among other already described in the previous section. ε_i describes the error term that captures idiosyncratic shocks or unobserved student characteristics.

For the main research questions are presented three identical OLS models, each one used for a specific type of strategy established and data used:

(2) $Student\_Outcome_i = \beta_0 + \beta_1.\mathbf{starting\_age} + \beta_2.Xi + \varepsilon_i$

(3) $Student\_Outcome_i = \beta_0 + \beta_1.\mathbf{january} + \beta_2.Xi + \varepsilon_i$

(4) $Student\_Outcome_i = \beta_0 + \beta_1.\mathbf{Early\_Enrolment} + \beta_2.Xi + \varepsilon_i$

The dependent variable (*Student_Outcome*) represents either the retentions, already described in the previous section as a binary variable, or the students' scores in the Portuguese or Maths national exam.

Equation (2) is used for the students born between January 1st and September 15th and the main explanatory variable (*starting_age*) describes age at school entry, measured in months and used as a continuous regressor. Equation (3) is used for the analysis of students born either in January or December, and uses the *january* dummy variable to capture the effect of an absolute one year difference in the measured outcomes. Finally, the main explanatory variable of equation (4) is the *Early_Enrolment* variable and its major value is to address the effect of deferring student's entry by one year. In the analysis performed in this case only students born between September 16th and December 31st are used.

Similarly for these equations, the vector X_i describes the control variables and ε_i describes the error term that captures unobserved student characteristics.

The main identification assumption of using the Linear Probability approach is that all the explanatory variables are exogenous, which means that the explanatory variables and the error term are not correlated. If unobserved heterogeneity exists, meaning that the characteristics of older entrants might differ from the characteristics of students that started school at a younger age, the exogeneity assumption is violated which will lead to biased results. For instance, if students that defer have lower ability and the ones who enter younger are more mentally developed (and have therefore more ability), then results may be downward biased.

Despite the limitations of this method, the division in the three strategies established in this report aims at dealing with the unobserved heterogeneity, differing it this way from the simpler analysis performed by other authors. With equation (2), the impact of relative age is measured only for students under the same legal conditions, as they are mandated to start school at six years old. Equation (3) aims at testing the absolute age difference in order to enable comparisons between absolute and relative age effects. In equation (1) and (4), only students born after the threshold date are used. These students are all under the same legal conditions as well, and if they all entered in school in the year they turned six years old, they would all be the youngest. However, some parents postpone their children's start. Equation (4) was, therefore, developed to analyse the benefits of such decision for students under the same conditions.

RESULTS AND DISCUSSION

After the description of the empirical strategy, in this section are presented the main results of the estimation of models. Introductory question results are firstly introduced and the main research question outcomes are presented afterwards.

The results regarding the predictors of early enrolment are presented in table 5. It is possible to infer that children that are born closer to the cut-off date of September 15th are more likely to start school at five years old. This corroborates the previous comment made: the percentage of students born in December that defer primary school start is much higher than those of children born in September.

Regarding the remaining variables there is evidence that suggest that students with parents with higher qualifications are more likely to start school earlier. One possible explanation for this outcome can be that such parents may be more capable of preparing their children for the primary school and consider that, even at a younger age, they are mature enough to start school. On the other hand, the outcome table suggests that students of worse socioeconomic environment are more likely to defer school entry by one year. The inference about the socioeconomic environment of students derives from the interpretation of the subsidy or unemployment variables.

It becomes relevant to mention at this stage one possible limitation of this analysis, as results can be somewhat biased due to the fact that children that start school at a younger age from high socioeconomic families may have the opportunity to attend a private school if the enrolment is rejected by the public system. Despite the limitation, the data available does not enable the study of this interesting topic which, nevertheless, presents an interesting result.

The difference between the three models presented in table 5, is the control for academic year fixed effects and location of school fixed effects. The results are similar which contributes to the robustness of the analysis. For this estimation, as well as the following ones, in order to account for the variation of the error components in the model, robust standard errors were used in regressions. The introduction of fixed effects increases the coefficient of determination (R^2). When regressing for each academic year observations, the conclusions taken are similar when compared with pooled observations.

Table 5. Results of OLS Estimation for predictors of *Early_Enrolment*

Dependent Variable: *Early_Enrolment* **Independent Variables**	**Model 1**	**Model 2**	**Model 3**
age_year6	**0.0699*****	**0.0701*****	**0.0703*****
female	0.0428***	0.0420***	0.0419***
foreign	-0.1550***	-0.1460***	-0.1490***
foreign_mother	-0.0652***	-0.0151***	-0.0171***
qualification_mother	0.0056***	0.0089***	0.0116***
unemp_mother	-0.0333***	-0.0333***	-0.0249***
SASE_beneficiary	-0.0293***	-0.0278***	-0.0270***
computer	0.0104***	0.0151***	0.0177***
Constant	0.9360***	0.9180***	0.9070***
Municipal FE	-	Yes	Yes
Academic Year FE	-	-	Yes
Observations	221,948	221,948	221,948
R-squared	0.059	0.155	0.182

*Statistically significant at 10%, **Statistically significant at 5%, ***Statistically significant at 1%

Concerning the short-term impact in retentions and national exam's outcomes, which are the main objective of this report, the results are presented separately by the three types of strategies used and that were described in the methodology and data section.

Only the main coefficient of interest and R^2 are displayed. For the following tables of results the upper part displays the coefficients without control for city and class fixed effects, while the lower part controls for such fixed effects.

Students born from January 1st to September 15th

Table 6 displays the main results for students born in the range of January 1st to September 15th. The selection of these students was due to the fact that they are mandated to enrol by the time they turn six years old. The main explanatory variable used is the age in months as continuous regressor, which means that the aim is to analyse the relative age effects.

From the results in table 6, it is possible to infer that students that start school at an older age are less likely to fail one school year in primary school. The value of coefficient is small but highly significant, as being one month older decreases the probability of retention by 0.56 percentage points.

Table 6. OLS Estimation for the impact of relative age in retentions and exam's scores

		2013 National Exam		2014 National Exam		2015 National Exam	
Dependent Variable	***Retentions***	***PT***	***MAT***	***PT***	***MAT***	***PT***	***MAT***
Main Independent Variable: ***starting_age***	-0.0056 ***	0.444 ***	0.448 ***	0.488 ***	0.521 ***	0.469 ***	0.444 ***
R-squared	0.068	0.090	0.094	0.090	0.106	0.094	0.101
		2013 National Exam		**2014 National Exam**		**2015 National Exam**	
Dependent Variable	***Retentions***	***PT***	***MAT***	***PT***	***MAT***	***PT***	***MAT***
Main Independent Variable: ***starting_age***	-0.0057 ***	0.451 ***	0.468 ***	0.496 ***	0.527 ***	0.458 ***	0.434 ***
R-squared	0.187	0.312	0.344	0.331	0.370	0.326	0.374

*Statistically significant at 10%, **Statistically significant at 5%, ***Statistically significant at 1%
Note: The upper part of the table displays the coefficients without control for city and class fixed effects and the lower part controls for such fixed effects

The results for fourth grade national exams of 2013, 2014 and 2015, suggest that students that started at an older age tend to score better than their younger counterparts. The interpretation for exams is similar. For instance for the 2013 Portuguese exam, a student who is one month older is likely to, on average, score 0.444 points higher. All coefficients presented in the table are significant at 1% confidence level. The study of three different years of exams enriches the analysis and contributes to its robustness.

Students born in January or December

For this second scenario, students born in both extreme tails of the civil year are used (i.e. students born in January or December), have practically one year of age difference and start all in the same academic year. This study is relevant to corroborate the interpretation of results from previous analysis. However, the main distinction from these two scenarios is that, while in the first case the relative age effects were being studied, in this case it is the absolute difference instead. Nevertheless, the expected results are the same since we are comparing the older versus younger students' performance.

Table 7 displays the results of this estimation, using the variable *January* as main explanatory variable, which distinguish the students that started school at six years old (born in January) or five years old (born in December).

Table 7. OLS Estimation for the impact of absolute age in retentions and exam's scores

		2013 National Exam		2014 National Exam		2015 National Exam	
Dependent Variable	***Retentions***	***PT***	***MAT***	***PT***	***MAT***	***PT***	***MAT***
Independent Variable: *january*	-0.0547 ***	4.876 ***	5.083 ***	4.871 ***	4.834 ***	4.639 ***	4.201 ***
R-squared	0.069	0.100	0.099	0.090	0.105	0.105	0.106
		2013 National Exam		**2014 National Exam**		**2015 National Exam**	
Dependent Variable	***Retentions***	***PT***	***MAT***	***PT***	***MAT***	***PT***	***MAT***
Independent Variable: *january*	-0.0527 ***	5.580 ***	5.990 ***	5.102 ***	5.032 ***	3.900 ***	3.812 ***
R-squared	0.354	0.561	0.584	0.578	0.597	0.577	0.605

*Statistically significant at 10%, **Statistically significant at 5%, ***Statistically significant at 1%
Note: The upper part of the table displays the coefficients without control for city and class fixed effects and the lower part controls for such fixed effects

Even though the data and the explanatory variable used are different, the main conclusions are in good agreement with the first case, as expected. Students that are one year older are less likely to be retained in any year of primary school and are more likely to perform better in national exams. Being one year old older is expected to decrease the probability of retentions by 5.47 percentage points. Similarly for exams, and using again the 2013 Portuguese exam, students that are one year older are expected to score, on average, practically more 5 points than younger colleagues. All coefficients are statistically significant at 1% confidence level.

While it was mentioned that there is no agreement regarding the impact of school starting age in the educational attainment, from the literature review it was possible to infer that a relevant amount of papers seem to conclude that younger entrants tend to perform worse than their colleagues. The analysis in this report seems to be in agreement with those author's views.

Students born from September 16th to December 31st

From this last scenario, more important than studying the age effect, it is possible to study the question of whether the option to defer or anticipate the entry in primary school brings some benefits or not. For that purpose, only students that can be considered conditional are used, as their parents have some power to decide the timing of enrolment of their children

Table 8 displays the main results for this analysis in which was used, as main explanatory variable, the binary variable *Early_Enrolment* that distinguishes students that started at five years old from the ones that delayed school entry and only started at six years old.

Table 8. OLS Estimation for the impact of early enrolment in retentions and exam's scores

		2013 National Exam		2014 National Exam		2015 National Exam	
Dependent Variable	***Retentions***	***PT***	***MAT***	***PT***	***MAT***	***PT***	***MAT***
Independent Variable: *Early_Enrolment*	-0.0347 ***	-0.941 ***	0.915 **	-1.916 ***	0.072	-1.865 ***	0.530
R-squared	0.073	0.076	0.085	0.073	0.099	0.084	0.096
		2013 National Exam		**2014 National Exam**		**2015 National Exam**	
Dependent Variable	***Retentions***	***PT***	***MAT***	***PT***	***MAT***	***PT***	***MAT***
Independent Variable: *Early_Enrolment*	-0.0347 ***	-0.713	0.986 *	-1.820 ***	-0.563	-1.577 ***	-0.405
R-squared	0.261	0.421	0.454	0.442	0.477	0.437	0.483

*Statistically significant at 10%, **Statistically significant at 5%, ***Statistically significant at 1%
Note: The upper part of the table displays the coefficients without control for city and class fixed effects and the lower part controls for such fixed effects

The analysis of retentions presents evidence that students that start school earlier are less likely to be retained during primary school attendance. From table 8, being one year younger (which in this case means, starting school as conditional) is expected to decrease the probability of being retained by 3.47 percentage points. This result contradicts the previous ones.

Regarding the impact in national exams' scores, it is difficult to draw any conclusions. Some coefficients of the Portuguese exam seem to indicate that younger entrants tend to perform worse, which would be aligned with the previous conclusion of this report. However, when analysing the Math exams' information, many coefficients are not significant even at 10%. Even when the coefficients for Portuguese exams are significant, their value is very small, which may imply that the benefits from enrolling one year later are not that great.

An explanation for these contradicting results is that the results in table 8 may be biased due to endogeneity/selection problems since early enrolment is most likely not exogenous. It is reasonable to assume that parents decide on early or late entry by taking into account unobservable characteristics, namely development of cognitive skills and maturity, which in turn are also related to future achievement.

CONCLUSION AND DIRECTION FOR FUTURE RESEARCH

The main goal of this report was to use the Portuguese data to discuss one of the most important issues of the Economics of Education field: the problem of the school starting age. Plenty of studies were developed for several countries, and the results are still ambiguous.

Complementary to this study, an introductory question was also studied with the purpose to infer the characteristics that may influence a child's early enrolment. Such question is relevant because Portuguese cut-off rules are not strictly imposed and allow for parents to determine, in some cases, whether children should start at five or six years old.

To answer the research questions, an Ordinary Least Squares approach was developed and the main findings were that students that start at an older age tend to perform better than their younger counterparts and are less likely to be retained at least one time during primary school. Such conclusions are according to most literature review, as there seems to be a consensus of a positive relation between school starting age and primary school performance.

However, when addressing the deferment issue, results are not so clear. When restricting the analysis to those children for whom parents can decide to enrol them with five years old or wait for another year (those children born between September 16th and December 31st), then what is obtained is that there is some evidence that students that start at a younger age in the sample may perform worse in exams (despite the low effects), however they are less likely to be retained. This raises one important outcome for public policy, related with the cut-off rule: allowing for parents to decide the enrolment of five-year-olds is positive, as parents are more aware of the ability of their children. The results of the estimation suggest that students are deferred most likely because they were not ready to start school early, while the ones who start school with five years old have maturity enough to keep up with their class.

Regardless of the cut-off rule stipulated, there will always be relative age differences. Therefore, the state could help by providing some assistance and developing programs in schools, so that the relative age difference can dissipate in the first years of schooling. The opinion of the kindergarten teacher could also be an important contribute to perceive child's maturity and readiness to enrol.

Educational field is of upmost importance for the Government, however there are plenty of costs associated with it. Both retentions and deferment of school entry are not desirable, so the state must balance both phenomena to obtain better educational and economic outcomes.

In this report an OLS approach was used, using three different approaches. However, for future research, in order to address the question of exogeneity in a better way, an instrumental variable can be explored. Further research could also expand the analysis of this report by including private school information, analysing the impact of school starting age in a medium and long term perspective and even including different control variables.

REFERENCES

Angrist, J. and Krueger, A. 1991. “Does Compulsory School Attendance Affect Schooling and Earnings?”. *Quarterly Journal of Economics*, 106, 979-1014.

Attar, I. and Cohan-Zada, D. 2017. “The Effect of School Entrance Age on Educational Outcomes: Evidence Using Multiple Cutoff Dates and Exact Date of Birth”. IZA DP No. 10568

Barua, R., and Lang, K. 2008. “School entry, educational attainment and quarter of birth: A cautionary tale of LATE”. NBER Working Paper No. 15236.

Becker, G. 1964. “Human Capital: A Theoretical and Empirical Analysis, with Special Reference to Education”. *National Bureau of Economic Research*.

Bedard, K. and Dhuey, E. 2006. “The persistence of early childhood maturity: Inter- national evidence of long-run age effects” *The Quarterly Journal of Economics*, 121 (4), 1437–1472

Black, S., Devereux, P., and Salvanes, G. 2011 “Too young to leave the nest? The effect of school starting age”. *The Review of Economics and Statistics*, 93 (2), 455–467.

Coutinho, M., Reis, H. 2014. “Grade retention during basic education in Portugal: determinants and impact on student achievement”. Economic Bulletin, Banco de Portugal, 61-83.

Crawford, C., Dearden, L., and Meghir, C. 2010. “When you are born matters: The im- pact of date of birth on educational outcomes in England” Centre for the Economics of Education Working Paper No. 10-09.

Cunha, C., Heckman, J., Lochner, L., Masterov, D. 2006. “Interpreting the evidence on life cycle skill formation” in *Handbook of the Economics of Education, Volume 1*, ed. Eric Hanushek and Finis Welch, 698-747.

Dobkin, C., and Ferreira, F. 2010. “Do school entry laws affect educational attainment and labor market outcomes”. *Economics of Education Review*, 29 (1), 40–54.

Elder, T., and Lubotsky, D. 2009. “Kindergarten entrance age and children’s achievement: Impacts of state policies, family background, and peers”. *The Journal of Human Resources*, 44 (3), 641–683.

Fertig, M. and Kluve, J. 2005. “The Effect of Age at School Entry on Educational

Attainment in Germany”. IZA Discussion Paper 1507.

Fredriksson, P., and Öckert, B. 2004, "Is Early Learning Really More Productive? The

Effect of School Starting Age on School and Labor Market Performance", IZA DP No. 1659

Kawaguchi, D. 2011. “Actual age at school entry, educational outcomes, and earnings”. *Journal of the Japanese and International Economies, 25* (2), 64–80.

Mühlenweg, A., and Puhani, P. 2010. "The evolution of the school-entry age effect in a school tracking system" *The Journal of Human Resources, 45* (2), 407–438.

Puhani, P., and Weber, A. 2007. "Does the early bird catch the worm? Instrumental variable estimates of early educational effects of age of school entry in Germany". *Empirical Economics, 32* (2–3), 359–386.

Peña, P. 2016, "Creating winners and losers: Date of birth, relative age in school, and outcomes in childhood and adulthood". *Economics of Education Review*, 56 (2017) 152–176

Ponzo, M. and Scoppa, V. 2014. "The long-lasting effects of school entry age: Evidence from Italian students". *Journal of Policy Modeling,* 36 (2014) 578–599