THE EFFECT OF AGE ON EDUCATIONAL PERFORMANCES AT THE END OF PRIMARY SCHOOL - A CROSS-SECTIONAL STUDY FROM REUNION ISLAND

JEL classification : I21, I24

Daniel RAKOTOMALALA

Abstract

This work focuses on the quantitative measure of the causal relation between age and school results of pupils at the end of primary school in Reunion Island. The effect of age is composed of at least three distinct ones : (1) age at entry effect, (2) age at test effect (called absolute age effect) or (3) relative age effect (meaning that regardless of the chronologic age, a pupil remains being the oldest or the youngest compared to their classroom peers). According to the empirical literature which could disentangle these different effects, the age at test is revelealed to be the most important component. Indeed, it would reflects the absolute intellectual maturity of a pupil (hence the « absolute » age appelation of the age at test), which could explain that within a given grade, the oldest ones get average higher scholar results than the younger ones. Thus, in order to extend the knowledge about the determinants of educational success, especially about the impact of age on scholar results and then help policy makers in their decisions about optimal policies in the education field by providing informative results ; this paper, using compiled cross-sectional data sets, exploits an exogeneous variation of the age at test within a grade induced by month of births to measure the causal impact of age at test on the national achievement assessment scores in grade 5 in Reunion Island. The principal findings are that the age at test have a substantial positive effect on test scores and this effect is heterogeneous across sex and socio-professional category subgroups. These findings would suggest at best that, in order to improve the educational results of pupils in Reunion Island, policy makers could first increase the minimum age of school entry. Second, they could regulate classroom compositions such that the age distribution within a classroom does not disperse too much. Third, they could normalize national achievement assessment scores by age or making pupils with different ages within a grade pass the national assessment at different times such that they have sufficiently close ages test to not significantly impact their results. The latter enables at the same thime to correct the inequality of having a different month of birth (unchosen by the pupils) which is likely to lead towards different educational outcomes. Pupils would be indeed assessed at their full potential.

Key words : Age at test, Relative age, Month of birth, Educational performances

---

Name : RAKOTOMALALA Daniel Herinjaka
Phone number : +262 692 05 25 52
Mail adress : daniel.rakotomalala@univ-reunion.fr

1 Introduction

In the french (primary) education system, there is a unique cutoff date of eligibility for entering school : children must have turned six by the December, 31st of the calendar year containing the school entry date (August, 18th or 19th).[1] This leads to the existence of a continuum of ages within a grade because children born in different months of birth find themselves being in the same grade. For example, for the 2005-2006 school year, because the school entry date is the August, 18th and because of the rule mentioned above, all chidren (regardless of their months of birth) are eligible to enter school for the 2005-2006 school year. Hence, by the end of the 2005 year (within the grade 1 of the 2005-2006 school year), there are children aged between exactly 6 year (born in the December, 31st of 1999) and almost 7 year (born in the January, 1st of 1999). It is known that generally, this difference in ages causes differences in educational outcomes (test scores at several school grades, cognitive abilities and even later life outcomes). In addition, the youngest ones within a grade have generally some disadvantages compared to their older peers.

Several mixed potential explanations underly this evidence. E. Cascio and others (2008) gives a summary of these concerns.[2]

First, the difference in educational outcomes between the young ones and the older ones can be due to a relative age effect. In other words, there is peer effects such that, because of their younger peers, the older ones tend to do better.[3]

Second, the observed age effect can cover an absolute maturity effect. This is of special concern if the studied outcomes are of educational performance types (for example test scores or higher education participation) because the absolute maturity can be reflected in the age of which the individual sits the examination (thus there is the « absolute age » appelation in some papers in the literature). More precisely, this second effect has to be distinguished from the first in the sens that on average, older an indiviual takes an examination, regardless of his rank into the age distribution of his grade, better his results will be. It means that within a grade, if the older ones and the younger ones within a grade take their respective exams at different dates such that they are equally aged at the moment of these exams, on average, they would have the same performance (if the age effect is solely composed of an absolute age effect). As matter of intepreting this second effect, borrowing Kaila (2017) terms, « older students do not learn at a faster rate, but they do better in exams just because they have had more time to accumulate knowledge ».

[1] A more detailed presentation of the french primary education system is provided in Section 3.

[2] Although the paper refers to entry-age differentials, the reasoning can be reported into a within grade age differentials based reasoning.

[3] An hypothetical example is that the olders gain self-confidence and motivation due to their consciousness of being more aged compared to their classmates.

Third, the age-based gap in educational outcomes can be caused by a pure age of school entry effect. This feature is partially motivated by a child developmentalists concern : the readiness of a child to enters school (Fredriksson and Öckert 2006). In fact, since being among the youngers within a cohort implies having an inferior age at school entry, younger children could perform worse because they were not sufficiently mature when they entered school, which negatively troubled their learning skills in the following school years. Moreover, to clearly distinguish this effect from the two previous ones, consider children such that some are relatively older compared to their peers and at the same time older at test dates. Then, if there were no existence of relative age effects nor age at test effects and all children started school at the same age, that would lead to zero difference in educational outcomes at all between the olders and the youngers. Fourth, there is potentially a length of schooling effect. This can be interpreted as follows : children differs in some educational outcomes because some of them spent more time in school[4] (hence had accumulated more knowledge).

As this is all theory, and no measure of such effects can be done theoretically, measuring the causal impact of age within a grade is reported to an empirical problem (Black, Devereux, and Salvanes 2011, Fredriksson and Öckert (2006), Robertson (2011)). Then, data is necessary to perform measures of age effects on education outcomes. Nevertheless, isolating such effects is empirically difficult because of the perfect collinearity between age at test, age of school entry and length of schooling. Indeed, the age at which a pupil takes a test equals to the sum of his age at entry school and his length of schooling.

Consequently, to isolate an, age at test effect for example, there is the necessity to find a framework and data in which age at test varies independently from age at school entry and length of schooling (Crawford, Dearden, and Meghir 2007). It means that the framework allows the age at test to vary for the same age at school entry and length of schooling. Given that such conditions are hard to fulfill, only few studies managed to separate some age effects from another : Black, Devereux, and Salvanes (2011), Crawford, Dearden, and Meghir (2007), E. U. Cascio and Lewis (2006) (see the literature review for more details). Otherwise, most of the studies present estimates that contain more than one of these effects.

On the other hand, in the within grade comparison of educational outcomes by age, the age variable (age at school entry, age at test) is endogeneous. In fact, given a grade, there are some children that are aged at least one year more than their theoretical age (the age they would have if they entered at the first time they were eligible and if they did not repeat or skip a grade)[5] and there are some aged at least one year less. The latter phenomenon is not a random one, impying that the comparison mentioned above suffers from bias

[4] As explained in Grenet (2009), this is caused by the existence of multiple entry cutoff for the same school year and compulsory laws in some states.

[5] For instance, in France, the theoretical age for being in grade 5 is the age of 10.

selection. In fact, those who have one year of delay compared to their theoretical age can be scinded into two types : those who went through grade retention and those whose parents intentionally delayed their age at school entry (this practice is called « redshirting »). The second one has two potential reasons : either the parents are awared about the advantage of the olders within a cohort and then delay their child school entry to make him one of the oldests in his grade ; either the parents observed that their child have learning disabilities compared to normal children and then decide to delay their school entry in order to give time to the child to be more ready for school environments. There is then a positive correlation between the age and grade retention (which is not accounted for in the comparison of educational outcomes by age). Following this reasoning, while grade retention is negatively correlated to educational outcomes because repeaters are supposed to have poorer abilities then poorer results, redshirting is positively correlated to educational outcomes. What is the sign of the bias then ? To answer this question, consider that since repeaters are generally more numerous than redshirers, the age effect estimation suffers from a downard bias (Bedard and Dhuey 2006, Grenet (2009), Hámori (2007), Hámori and Köllő (2012)). The same logic applies to those who are in advance of one year because it is most likely that children with higher abilities are enrolled (by decision of their parents) earlier in school, thus early enrollement is negatively correlated to age and positively correlated to educational outcomes. Though, the amount of children in advance is generally very small so this last consideration is not of much concern.

This paper takes into account such endogeneity by instrumenting the age at test by the assigned relative age[6] (the age position at which a pupil should be if he entered school by the first time he were eligible and if he did not repeat or skip a grade) to estimate the effect of age on standardized national achievement assessment scores in the end of primary school in Reunion Island using cross-sectional data set. I indeed exploit the exogeneity of month of birth and its strong prediction power of the actual age at test to estimate by an instrumental variable and a reduced form framework the effect of age on national test scores. According to my knowledge of the existing studies about this issue or a comparable one, this is the first conducted with Reunion Island data. Also, this study is of education policy matter because of its informative causal findings combined with the possibility for policy makers to influence one variable of interest determining educational outcomes : the age. In fact, I find a substantial and positive effect of being one year older on total test scores that varies from +0.2 to +0.3 of a standard deviation. I also find heterogeneous effects by group of sex with the female pupils generally gaining more than males pupils from being a year older.

The remainder of the paper is organized as follows : Section 2 presents a literature review, Section, 3 describes the data I used, the Reunion Island education system and the econometric framework. The two next sections

[6]It is computed as the time distance between the December, 31st of the pupil's year of birth and his month of birth. And in this paper, it is reported to year unit.

: 4 and 5 respectively exposes different regression results (with discussions) and concludes.

2 Relevant literature

Existing literature

The relationship between age (in different forms) and educational outcomes in general is a well documented topic. The literature body was already large until 10 years earlier. In the end of the 1980's for example, Cahan and Cohen (1989) investigate the effect of age and schooling (to be distinguished) on scores obtained from ability tests in grades 5 and 6 using Jerusalem's data. Another example in the beginning of the 1990's is the study performed by Bell and Daniels (1990) where they compare within a grade automn-born children and summer-born children at their 11^{st}, 13^{th} and 15^{th} years, on their APU (*Assessment of Performance Unit's Science Project*) Science Syrvey tests to assess the effect of being one year older (birthday effect) than their classmates in education.[7] More studies using number of outcomes types appear to adress the relationship between age and education in the 2000's (and before 2010) than in the 1990's. Taking an example fo them, consider Graue and DiPerna (2000). This paper performs a statistical analysis of achievement gap (promotion to the next grade) between those who delay their entry in kindergarten or are retained[8] and those who enter school as soon as they are eligible in Wisconsin. Using another type of variable of interest, Leuven et al. (2004) work on the impact of expanding school enrollment opportunities on achievement (language or math test scores) in Netherlands. In other point of view, this study estimate the effect of schooling (measured by potential months enrolled in school) on the test scores of interest using datas from the PRIMA (Primary Education and Special Education Cohort Studies) survey. Also, almost all of these papers are interested in the effect of school starting age and they come from considerable variety of states. For instance, Strøm (2004) estimates the effect of school starting age on reading test scores for 15-16 years old pupils in Norway using PISA (*Programme for International Student Assessment*) data. In addition, the same outcomes are of interest in Germany, United States, Sweden, England, Hungary and France respectively in P. Puhani and

[7] For other references, see Langer, Kalk, and Searls (1984) ; Cahan and Davis (1987) or Mayer, Knutson, and others (1997). The first attemps to highlight the causal effect of school entry age and relative age on mathematics, science and reading test scores using National Assessment of Educational Progress (NAEP) data . The second in its case adress the measure of the effect of schooling on achivement (measured by nationa test scores) whilst the third exploits the National Longitudinal Survey of Youth (NLSY) to estimate the school entry age effect on cognitive and non-cognitive (behaviour problems) developement.

[8] In kindergarten or grade 1-3

Weber (2005)[9], Datar (2006)[10], Fredriksson and Öckert (2006)[11], Crawford, Dearden, and Meghir (2007)[12], Hámori (2007)[13] and Grenet (2009). For studies interested in the effect of age on another outcomes, see for example Fertig and Kluve (2005) which use obtained degree and probability of retention as outcome in Germany ; or Dhuey and Lipscomb (2008) which adress the effect of relative age on high school leadership activities (being the president of a club or the captain of a team).

Even from 2010 till the present moment, the literature body continues to extend. For example, an interesting study with noticeable outcomes of study is Black, Devereux, and Salvanes (2011). This paper investigates the long-run impact of entry age on IQ scores at age 18 and teenage pregnancy (in addition of usual educational attainment and earings.). Additionally, in the recent years and from many states, there seems to be a set of studies that are interested in the effect of age on outcomes besides test scores type. In fact, while Mühlenweg and Puhani (2010) ; Jürges and Schneider (2011) or Schneeweis and Zweimüller (2014) are into outcomes of track choice types in, respectively Germany (for the two first) and Austria, studies like Suziedelyte and Zhu (2015) and Dhuey et al. (2017) investigate in outcomes of cognitive and non-cognitive development types in, respectively Australia and Florida.

Conceptual considerations

Almost all of these papers mentioned above are awared of the potential explanations underlying age effects (relative age effect, absolute age effect, school starting age effect and length of schooling effect). Note that in the french context, since there is generally a single school entry cutoff date considered, it makes sens to discuss issues about school entry age and length of schooling that are massively developped in the literature body, especially those of school starting age.

Consider the following questions that needs involvement of conceptual considerations. Is it better to delay school entry ? What is the optimal age at which a child should start school and what policies or interventions[14] could be made about ? are necessary to improve outcomes of interest without harming other children ?

To begin with, several theoretical considerations about how age would affect educational and later was made over years. The ones presented here is surely not exhaustive[15], but appears to have their importances.

[9]Another studies in Germany as Fertig and Kluve (2005) can be retrieved who are interested in the impact of school starting age on schooling and retention ; in Mühlenweg and Puhani (2010) which consider track attendance (academic versus non academic track) as outcome.

[10]See Aliprantis (2014) and Fletcher and Kim (2016) for other studies in the United States. The former estimates separately pure entry age effect and relative entry age effect on mathematics and reading item response theory test scores while the latter investigates the impact of entry cutoff changes on National Assessment of Educational Progress (NAEP) test scores.

[11]The same authors published another study from Sweden in 2013 (Fredriksson and Öckert 2013) in which they put interest in educational attainment and 25-54 years old earnings.

[12]A recently published study from England (Wales) is Hart and Moro (2017) in which the authors study the impact of quarter of birth on the probability of gaining selective school entry.

[13]See also Hámori and Köllő (2012).

[14]Beside action from policy makers, parents could decide of their children's own scholl entry age.

[15]This is because theoretical concepts are often given very implicitely in the economic literature. Also, these are more of child

Councerning the debate about delaying or not school entry, this question was largely adressed in the United States around the 2000's[16].

One key concept of interest behind the age of school entry of a children is the « readiness ». Under this latter, there are hypothetical factors that are of cognitive, as non-cognitive dimensions (Stipek 2002).[17] More concise, DeCos (1997) provides a clear classification of theories behind « readiness for kindergarten » : *maturationist*, *behaviorist*, *envrionmentalist* or *interactionist/constructivist* theories.

For the maturationists, the readiness is defined solely within the child and depend only on his biological age. Thus, education has just to provide optimal environment for the child's maturation. From this point of view , children who are suspected to be not mature enough at a certain point are given a « gift of time » (by redshirting, retention or transitional classes). On the other hand, for the behaviorist and the environmentalist, the knowledge of a child is external and skill is considered as a puzzle such that its peices are supposed to be identified and assembled by the education. Last, an interactionist/constructivist have a combined idea between the maturationist's and the behaviorist's. About the optimal age of school entry, it appears to vary from 4 to 7 years old for several countries cited in DeCos (1997). Since there appears to be no theoretical conception of this issue, empirical studies adressing relationship between age at school entry and cognitive and non-cognitive development can help for clarifications. Refering to Aliprantis (2014) which provides a stress on the importance of investing in earcly childhood education, it is important because early skill accumulation is complementary with later one then can positively later life outcomes. Hence, this is of policy matter because policy makers can act on the age of kindergarten entry. Indeed, in the United States, there was a massive rise in minimum school entry age in the last decades (Datar 2006, Elder and Lubotsky (2009)) in response to these considerations. The most chosen kindergarten entry age appears to be around 5 years old (Elder and Lubotsky 2009). The other maneers to affect the age at which a child enters formal education are redshirting, pre-school interventions, transitional classes or grade retention. How did studies manage to highlight age effects on educational or later-life outcomes ? As Stipek (2002) managed to survey 36 papers on age effect of his time, the approaches used by the authors can generally be classified in three categories : first, a comparison in outcomes between pupils who delay their school entry and those who do not ; second, a comparison within a grade of pupils with different birth dates ; and third, the combination of the previous two.

Existing evidences

While doing a within grade comparison of children's outcomes, it is largely demonstrated (with rare exceptions)

development concerns rather than economic concerns. Finally, several desagreements seems to persists about the theory.

[16] Some empirical examples are cited above

[17] Some consider wellbeing, social and language development for example.

that the older children perform better than their younger peers. See Dhuey et al. (2017) for an extensive citation of this evidence. In other words, the age have a positive causal effect on educational outcomes. This relative advantage of the olders is more pronounced in early ages and fades over time (Bedard and Dhuey 2006, Elder and Lubotsky (2009) or Grenet (2009)). Bedard and Dhuey (2006) which investigated the effect of school starting age on mathematics and science internationally standardized test scores in OECD countries found that the oldest children within grade 4 score from about +0.2 to +0.4 of a standard deviation. The remaining effect at grade 8 is from about +0.1 to 0.4 for mathematics and science test scores again. Similarly, Elder and Lubotsky (2009) found an advantage of +0.16 of a standard deviation in mathematics test scores at grade 8. Also, for France, Grenet (2009) found that being a month older compared to the youngest children inscrease mathematics and test scores by approximatively +0.02 of a standard deviation. When reported to a year scale, it is equivalent to an advantage of +0.24 of a standard deviation, which joins the precedent results. This effect is attenuated in the future (grade 9) as mentionned earlier by shrinking to an advantage of just +0.02 and +0.13 in, respectively mathematics scores and french scores for those who are a year older.[18] A recent study, yet using alternative method (a regression discontinuity) but leading to similar results[19] is Peña (2017) which estimates the effect of relative age in test scores in Mexico for grade 3-9 pupils. The author found differences between means of mathematics test scores on one side and on another the cutoff (a causal effect in this case) from +0.3 to +0.36 of a standard deviation. Besides, note that the comparison between test score age effect on test scores reported in unit of a standard deviation only makes sens when recalling that over a year, an average student can gain from one quarter to one third of a standard deviation in test scores (Woessmann 2016).

On another side, few studies find small or negative age effects. Concerning the first case, probably the most known example is the Black, Devereux, and Salvanes (2011)'s paper in which they find, in Norway, small effect fo school entry age on IQ scores.[20] An example for the second case is Fertig and Kluve (2005) in which they found no effect on entry age on school attainment (schooling degree and probability of retention) in Germany. Even, Mayer, Knutson, and others (1997) assessed at that period that entering school at an younger age give an advantage in cognitive and non-cognitive development. Going with, Dobkin and Ferreira (2010) found that younger school enrollment increase the education attainment level in Texas and California. These exceptional evidences make somewhat informations about age effects on educational outcomes related variables slightly inconclusive.

In the other hand, the advantage of older children compared to their same grade peers is sometimes revealed

[18]In the Table 1 of Grenet (2009), they are expressed in month scales : +0.002 and +0.011 respectively.

[19]Though higher estimates

[20]Eventhough they reported a strong effect of age at test, the advantage of entering school a year later in Germany was founded to be also +0.06 *of a stanine* (about +0.04 of a standard error).

to be heterogenous with subgroups (generally with sex subgroups or social category or similar type subgroups). For instance, Bedard and Dhuey (2006) found found that the relative age of school entry effect to be higher for children at risk over OECD countries. This is as well what Grenet (2009) highlighted in France : the age effect is more important in early years for disadvantaged (refering to the child's household's occupation) pupils.[21] It appears to be also the case with Schneeweis and Zweimüller (2014) which studied the causal effect of relative age on probability of getting higher school tracking using PISA data. Recently, Aliprantis (2014) likewise found highly heterogeneous effects on math and reading item response theory test scores by home environment in the United States. These heterogeneity evidences are relevant in policy perspectives in the sens that they sugger policy makers to point interventions towards the concerned subgroups instead of towards all types of individuals. More recently and going with the set of papers that highlighted heterogeneity of age effects, Kaila (2017), in Finland, found a greater effect[22] of females than for males.

Overall, these existing evidences should be analyzed and interpreted carefully because many mechanisms besides the considered age of study can change the resulting effect (Kaila 2017, Aliprantis (2014)). One of these mechanism is the plurality in features of school systems all over the world. For example, in Japan, the length of schooling of a pupil is invariant with the month of birth because the Japan school system requires individuals to accomplish a fixed amount education regardless of their month of birth (Kawaguchi 2011). Thus, the author could estimate a pure school entry age effect. Another case when Kaila (2017) compared her results with that of Bedard and Dhuey (2006) within Finland. The magnitude of the estimate of school starting age on test scores reported by Kaila (2017) is smaller than that is reported in Bedard and Dhuey (2006). One interpretation the former's author give is that this can be caused by the simple fact that the test score used as outcome in her study, the Grade Point Average (GPA) is based on teachers'personal assessment while in the latter the test scores were internationally standardized ones (Trends in International Mathematics and Science Study or TIMSS).

[21]Though, Grenet (2009) found no significance difference between males and females pupils.
[22]Effect of relative school starting age on test scores and admission in general school.

3 Methods

3.1 Institutional background and data

3.1.1 The Reunion Island educational system

Since the Reunion Island is a region of France, its educational system can be, at least in the interest of this paper, presented as the Reunion Island system Alet, Bonnal, and Favard (2013) provides a good description of primary school in France : primary school is composed of five years (grade 1 to grade 5). Compulsory school begins in grade 1 and the age at which a children must starts school is the age of 6 (at least).

At the end of the grade 5, all children pass a national achievement assessment (*évaluation nationale des acquis*). Resulting from that, children receives a score between 0 and 100, with the government expected score to be 50. A brief description of the national achievement assessment will be presented next. One invariant feature is that for every assessment (every school year), two documents are established : one for the pupil that takes the evaluation (*cahier de l'élève CM2*) and one for teachers who will manage to correct the assessments (*Livret enseignant*). The first is just the document in which there are the assessment questions and materials inteded for the pupils answerings. The second contains precise instructions intended for teachers when the national evaluation takes place. For example, it requests teachers to check before the examination that the children have at their disposal necessary materials such as an eraser or a serviceable pencil. The principal informations within the teachers document are : the identification of the knowledges and skills to be assessed, necessary informations for passing exercises, the time affected to each exercise and all necessary instructions concernint the correction of exercises.

About the content of the assessment itself, there are one hundred questions called « items » (which are ranked from 1 to 100) to which pupils are supposed to give answers. These can be divided and subdivided by topics. Indeed, the first level of classification is such that the items from 1 to 60 are related to french topic exercises and the items from 61 to 100 are related to mathematics topic exercises. Moreover, both the french topic items and the mathematics topic items can be themselves subdivided. Indeed, the subtopics within the french items are reading, writing, vocabulary, grammar and spelling while those within the mathematics items are numbers, calculation, geomtetry, sizes and measures, and organisation and data management.[23] Descriptions of each subtopics are given in the teacher document.[24]

Concerning the cutoff rules, a pupil is first eligible to the $t-t+1$ school year for grade 1, with the entry date

[23]These features are invariant accross school years.

[24]For example, the purposes of the writing items are for pupils to copy without error and with an adapted presentation a text ; and to write several types of texts of at least two paragraphs with coherence and good spelling.

placed at the end of August of the year t, if he turns 6 by the December, 31^{st} of the year t. For instance, as mentioned earlier, in Reunion Island, the school entry date turns around the August, 18^{th}. It is because of the cyclonic period in Reunion Island around the beginning of the calendar year which makes the school vacation longer than the vacation in metropolitan France. This difference is then compensated by the earlier school entry date in Reunion Island. Moreover, it is prohibited that a pupil is retained more than once in the primary education. Empirically, the retention ratesin Reunion Island is about 16%, as illustrated in the third set of statistics in Table 1. This is quite large compared to the Metropolitan France, where it's around 6% (Alet, Bonnal, and Favard 2013). The proportion of pupils who are in advanced compared to their theoretical age in grade 5 is very small : 2% (see again the third set of statistics in Table 1). Hence, the proportion of pupils who are aged as their theoretical age in grade 5 is around 82%.

In other perspectives, the Reunion Island educational system is made of 92% of public schools and 8% of private schools as seen in the fourth set of statistics in Table 1. What can be concluded is that education is quite homogeneous in all angles. For example the average class size is around 23 ± 5 pupils per class and this feature is stable across three school years[25] in which these statistics were observed. In addition, from 1981, the french system introduced the concept of *« zones d'éducation prioritaires »* which are defined as areas where there are identified factors that cause school-based difficulties to pupils living in there. Hence, some school level measures are made in order to overcome these difficulties : the education placed-based policies. In the data, schools are classified in three education placed-based policies categories : *« Hors Éducation Prioritaire (HEP) »*, *« Écoles, Collèges et Lycées pour l'Ambition, l'Innovation et la Réussite (ECLAIR) »* and *« Réseau de Réussite Scolaire (RRS) »*. The first cateogry designates schools that do not benefit from these measures. While schools classified in the second type benefit from policies that are rather focused on elements concerning the education personnel, the RRS category is rather focused on social criterions. According to informations displayed in the fifth set of variables in Table 1, the proportions are stable across school years : around 52% of HEP schools, 25% of ECLAIR schools and 23% of RRS ones.

3.1.2 Generic description of the data

I use three cross-sectional administrative data sets to perform the estimations of the impact of age on test scores in grade 5. Each data set is a micro-level data within a school year.[26] They are an administrative data sets directly obtained from the rectorship of Reunion Island. In addition, each set contains personal informations about the pupils (exact date of birth, sex and socio-professionnal category of the legally first

[25] 2009-2010, 2010-2011 and 2011-2012, refer to the next subsection for more details.
[26] The school years are 2009-2010, 2010-2011 and 2011-2012.

responsible for the pupils), their educational achievements (5^{th} grade national assessment scores in details[27]) and their schools (townships, either they are public or private schools and the type of education placed-based policy which is applied to the school). Further details about the data variables will be given afterwards. The pupils are born between 1998 and 2002[28] and are aged between 9 and 12. Since the normal (without an advance or a delay of a year relative to the class) age to be in grade 5 is 10 year old and the test were taken in January and May of the civil year after the December of the school year entry, the maximum age of 12 (instead of 11) corresponds to the pupils born in January and had repeated a class.[29] Moreover, considering the three cohorts successively, the number of observations were in order of 14000 in the three cohorts.[30] This approximation represents the whole population of Reunion Island grade 5 pupils in year schools 2009-2010, 2010-2011 and 2011-2012. Since there were outliers within the data in the sens that there were pupils being aged more than or less than one year compared to the normal age at which they should be in grade 5, they were removed from. Note that even their proportions (in the three cohorts) are very small[31], they can create serious bias because some individuals were indicated to be less than 1 year old in grade 5 for example.[32]

3.1.3 Variables and summary statistics

The dependent variable is the 5^{th} grade national assessment scores. It was originally a 100 scale integer score since it is defined as a sum of one hundred items. An item (as presented earlier) is a binary variable related to a specific question within the assessment. It takes the value of 1 if the pupil had the correct answer, 0 if not. For the purpose of comparing my estimates with other works results, I normalized this test score to the mean of 0 and a standard deviation of 1. Furthermore, the one hundred items can be grouped by two main topics : french and mathematics, leading to the french topic and mathematics topic test scores. The french topic and mathematics topic test scores variables had originally, respectively, a scale of 60 and 40. In the same manner as the total score, I computed normalized versions of these two variables for comparability purpose.

A noticeable feature of test scores is their rising across cohorts, as seen in the first set of rows of Table 3. Indeed, the total test score mean varies from 47.50 to 52.89 (an increase of about 11%) between the school years 2009-2010 and 2011-2012. A similar variation in proportion is observed for the median and the third

[27] Total score, french topic score and mathematics topic score

[28] Those born in 1998 are those who were enrolled 1 year later than the year in which they turn 6 or those who repeated a class during primary school in the 2009-2010 cohort. Those born in 2002 are those who were enrolled 1 year earlier than the ear in which they turn 6 or those who skipped a class during primary school in the 2011-2012 cohort.

[29] For example, in the 2009-2010 cohort, a repeater who was born in January, 1^{st}, 1998 took the national test on January 2010 ; hence he is aged at least 12 at the moment of the test.

[30] 13630, 14708 and 13786 individuals within respectively the three cohorts

[31] Resepctively 0.5%, 0.6% and 0.4%

[32] The remaining data thus have respectively 13561, 14622 and 13733 observations.

Table 1: Summary statistics

	Cohort		
	2010	2011	2012
Age at test			
Min	9.050	9.070	9.400
Mean	10.710	10.690	11.030
Standard deviation	0.470	0.470	0.450
Max	12.050	12.050	12.400
Sex			
Females (%)	0.492	0.501	0.400
Males (%)	0.508	0.499	0.386
Missings values (%)	0.000	0.000	0.214
Position			
Repeaters/redshirters (%)	0.170	0.160	0.150
On time (%)	0.810	0.820	0.830
Advanced (%)	0.020	0.020	0.020
School status			
Privates (%)	0.080	0.080	0.084
Publics (%)	0.920	0.920	0.916
Education placed-based policies			
HEP (%)	0.524	0.527	0.517
ECLAIR (%)	0.251	0.246	0.264
RRS (%)	0.226	0.227	0.219
Class size			
Mean	22.835	23.520	23.366
Standard deviation	5.357	5.271	5.329

Table 2: Mean of total, french and mathematics scores by institutional features

	2010			2011			2012		
	Total	French	Mathematics	Total	French	Mathematics	Total	French	Mathematics
Sex									
Females	50.30	34.02	16.28	53.72	34.06	19.66	57.62	35.29	22.33
Males	44.79	28.93	15.86	47.84	28.87	18.97	52.56	30.67	21.90
SPC									
Farmers	48.62	32.00	16.62	54.45	33.55	20.90	55.71	32.97	22.74
Entrepreneurs	54.37	35.55	18.81	59.61	36.55	23.06	59.47	35.68	23.79
Executives	63.50	41.00	22.50	69.22	42.15	27.07	69.82	42.16	27.66
Intermediates	56.32	37.03	19.29	61.62	37.89	23.73	62.24	37.36	24.89
Employees	50.88	33.68	17.21	55.85	34.59	21.26	57.11	34.41	22.70
Workers	47.36	31.38	15.97	50.91	31.53	19.38	52.46	31.26	21.20
Retired	56.25	36.76	19.48	63.04	39.26	23.78	64.06	38.06	26.00
Unemployed	43.14	28.84	14.31	46.99	29.36	17.63	47.41	28.23	19.18
Others	40.51	26.89	13.62	42.62	26.41	16.21	45.39	26.95	18.44
School status									
Privates	56.43	37.24	19.18	62.12	38.33	23.79	62.68	37.89	24.79
Publics	46.72	30.92	15.80	49.80	30.87	18.93	51.99	31.05	20.94
Education placed-based policies									
HEP	50.73	33.50	17.23	53.73	33.25	20.47	55.67	33.44	22.24
ECLAIR	43.54	28.91	14.63	46.31	28.78	17.53	50.14	29.69	20.45
RRS	44.39	29.44	14.95	48.82	30.24	18.58	49.61	29.67	19.95

quartile.[33] The underlying pattern that explains this variation in total test scores is the significative variation of the mathematics test score accross the three cohorts (see the third set of rows in Table 3) and not of the french test score (see the second set of rows in Table 3 as the mean, median and quantiles are stable).

Furthermore, this difference in results in mathematics is mainly caused by the changing structure and contents of the three assessments across cohorts. Indeed, the items are not assinged to the same exact question types through the 2009-2010, 2010-2011 and 2011-2012 school year national assessments.

The independent variable of interest is the age at test of pupils. It is measured in years, taking in account month, exact date of birth and test date.[34] As completion of the information given above on the age at test, see the first set of statistics in Table 1. As can be observed, the mean age at test within the 2009-2010 and the 2010-2011 cohorts is about 10.7, whereas it is about 11 in the 2011-2012 cohort. Two underliying ideas can be illustrated here. First, the non negligeable amount of retention or redshirting, drives unsurprisingly the mean age to be above $\frac{9+12}{2} = 10.5$. Second, the mean age within the 2011-2012 cohorts is clearly above

[33]The corresponding variation for the first quartile is about 16%.

[34]For example, a pupil in the 2009-2010 cohort with an age at test of 10.08 is aged 10 years and one month ($\frac{1}{12} \approx 0.08$) at the date of January, 20th 2010

Table 3: Summary statistics of the total, french and mathematics test score variables

Cohort	Min	First quartile	Mean	Standard deviation	Median	Third quartile	Max
Total score							
2010	0	31	47.50	21.55	47	64	100
2011	0	34	50.79	21.79	51	67	100
2012	0	36	52.89	22.32	53	71	100
French score							
2010	0	21	31.43	13.64	32	42	60
2011	0	21	31.47	13.43	32	42	60
2012	0	21	31.62	14.03	32	43	60
Mathematics score							
2010	0	9	16.07	9.04	15	22	40
2011	0	12	19.32	9.39	19	26	40
2012	0	14	21.26	9.38	21	29	40

10.7 because of the date of the assessment in 2012.[35] Since the age at test variable is directly related to date of birth, we can describe the latter next. In a month of birth perspective, their proportions are illustrated by Figure 1 and generally, the month of birth appears to be uniformly attributed to pupils since there seems to be no overrepresented or underrepresented month of birth. Also, note that in the set of statistics presented in Table 1, for the three cohorts respectively, repeaters/redshirters are born in 1998, 1999 and 2000, those who are on time are born in 1999, 2000 and 2001 and those who are in advance are born in 2000, 2001 and 2002.

The additional independent variables are the sex and socio-professional category of the parents (here this is referring either to the father, the mother or the person that is legally first responsible of the pupil, not both) of the pupils. Refer to the first set of statistics in Table 1 for informations about proportions of sex accross cohort taking into account missing values within the data. What can be concluded is that the sex proportion appears to be fair and this is invariant across cohorts eventhough there is a considerable amount of missing values in the 2012 cohort. Concerning the socio-professional category variable, the proportions by cohort are illustrated by Figure 2. The pattern of the distribution of this variable appears to be stable through the three cohorts. More precisely, the unemployed are those who occupy the most of the part in each cohort (35% in the 2010 cohort, 36% in the 2011 cohort and 37% in the 2012 cohort) whereas the retired are the least numerous of the categories (0.8% in the 2010 cohort and 0.7% in the 2011 and 2012 cohort). Besides, the farmers take only around 1% of proportion in each cohort.

The logical continuation of the description is to check if some of these institutional features is likely to be highly correlated with the level of pupils. In order to highlight this concern, refer to Table 2.

First, concerning the sex variables, males perform worst than females in total scores. Moreover, it can be

[35] It was taken on May, 25th 2012 while the two other assessments were taken on January, 20th.

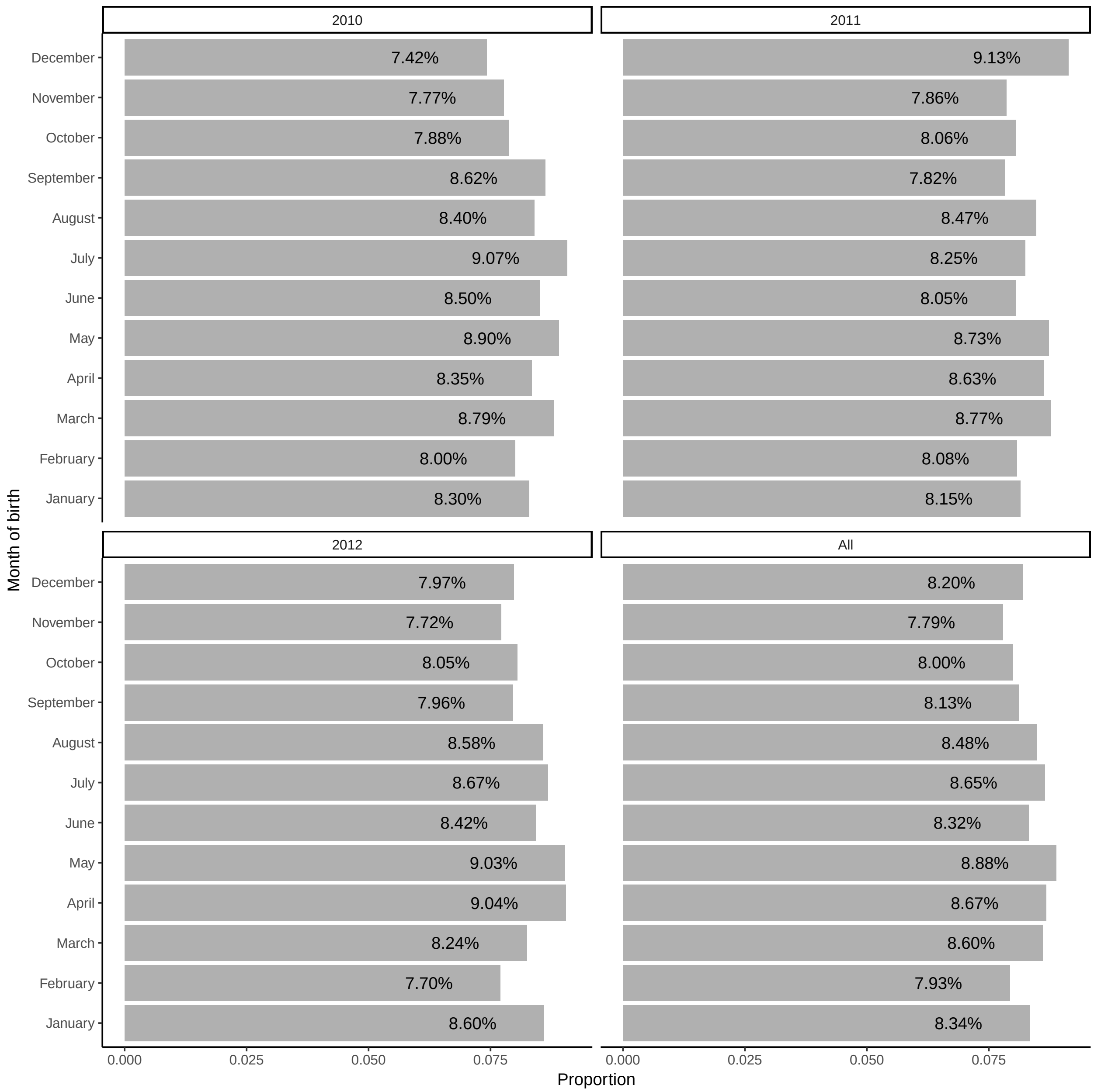

Figure 1: Month of birth proportions within each cohort

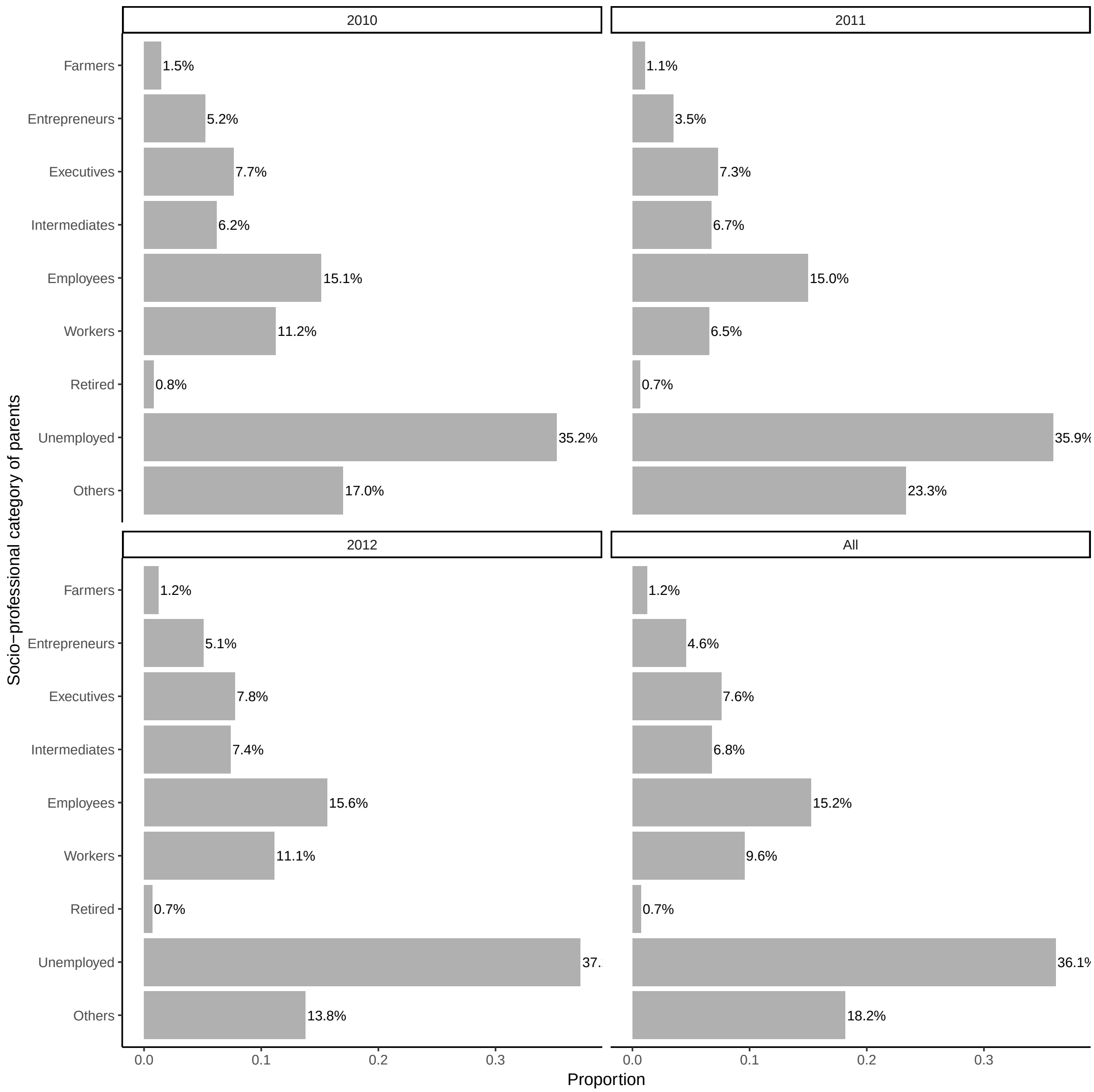

Figure 2: Proportions of socio-professional category of parents by cohort

drawn that this gap is due to the difference in french scores since it is greater than the mathematics scores.[36] This pattern is observed for the three school years.

When focusing on the socio-professional category of parents, it is revealed that the best performance in attributed to the executives children while the worst is attributed to the unemployed's children.[37] Also, if these eight categories are rearranged by the mean of total test scores in descending order, the resulting ranks are stables over the three cohorts.[38] When considering the french and mathematics scores separately, the rankings are non significatively different.[39]

Concerning the average scores by school status, it can be assessed that private schools perform better than public ones. This is true wether with total, french or mathematics scores, wether in 2010, 2011 or 2012 cohort. The private-public gap in total scores turns around 10 points, caused by a gap about 6 points in french and 4 points in mathematics. Finally, for the average scores by education placed-based policies, schools that perform the best are those which do not benefit from any policy. This is true regardless of wether the total, french or mathematics scores are considered and wether within the 2010, 2011 or 2012 cohort. For the two remaing categories of school (ECLAIR and RRS), their performances seems to be similar with the RRS schools performing, in general, slightly better than the ECLAIR schools.

3.2 Econometric framework

3.2.1 Strucural equation estimation

For practical reasons, let us present next the critical notations. The dependent variable (total test score, french test score or mathematics test score) is denoted Y. The independent variable of interest (age at test) is represented by A and individual-level independent variables (sex and socio-professional category of parents) will be compiled in the notation X.[40] Also, for later necessity, let the set $(1, A, X)$ be denoted by J. The instrumental variable, called *assigned relative age* and computed as $\frac{(12-m)}{12}$ where m represents the rank of the month of birth[41] is denoted Z. The indexes c and i that will be writen with these notations depending on the necessity represent respectively class and pupil.

[36] For example, in 2010, the females-males gap in french scores is $34 - 28.9 = 5.1$ points while in mathematics it is only $16.3 - 15.9 = 0.4$ points.

[37] Excluding the « Others » category since it contains unkown categories as well as missing values

[38] In 2010, the ranking, from the highest total test scores is : executives > intermediates > retired > entrepreneurs > employees > farmers > workers > unemployed.
In 2011 and 2012, the intermediates and retired's rank are inverted : executives > retired > entrepreneurs > employees > farmers > workers > unemployed.

[39] With the french scores, the ranking is exactly the same as with the total scores. With the mathematics scores, the ranking is, for the three cohorts, as with the total scores for 2011 and 2012 cohort, as illustrated earlier.

[40] This simply means $X = ($ sex, spc of parents)

[41] $m = 1$ for January and so on.

For illustration purpose, I first establish a linear regression of the test score on age at test and covariates, as illustrated below :

$$Y_i = \alpha_0 + \alpha_1.A_i + \alpha_2.X_i + \nu_i \tag{1}$$

ν_i is the usual error term.

Note that this equation will be estimated separately for the three cohorts.

The parameters of interest is α_1, which should capture, under the assumption of exogeneity of A and X, the causal effect of age at test[42] on national assessment scores. However, as asserted previously, A is highly expected to be an endogeneous variable, which formally means $E[A_i.\nu_i \mid X_i] \neq 0$.[43] COnsequently, the OLS estimator $\widehat{\alpha_1}(OLS) = (J^T J)^{-1} J\ Y$ is biaised and non-convergent.

The problem is typically an omitted variable bias. For example, consider that since the « ability », which determines test scores and is expected to be correlated to the age variable[44], is unobserved, the modeller is constrained to insert it into the error term. This leads then to a correlation between the error term and the independent variables and results in a biased estimate of the parameter of interest α_1.

On another point of view, the framework in equation (1) suffers from a selection bias problem because of the repeaters and those in advance. Indeed, being born late in the year is likely to rise the probability of repeating a year. Also, pupils who have an advance of one year are likely to be born earlier in the year (See Figure 4). Since the proportions of pupils who have a year of advance are very low within the three cohorts (see again Table 1), the main features that causes the bias selection problem is the presence of repeaters and redshirters. Furthermore, the bias arise with the fact that the age at test variable, within the equation (1) has two effects : a direct effect (α_1) and an effect going through ν_i because retention or redshirting is included in this error term. Note that these two features have distinct correlation with the independent variable – the test score. In fact, retention is negatively correlated with test score as repeaters are very likely to have lower ability and redshirting as positively correlated with test score as redshirtirers have higher maturity compared to their peers. Nethertheless, if repeaters account more than redshirters, which is most likely the case (Bedard and Dhuey (2006), Grenet (2009))[45], $\widehat{\alpha_1}(OLS)$ is downward biased.

As illustration about this issue, see Figure 3. It clearly shows that, either for the three separate cohorts or

[42] Although, the data don't allow to separate the different age effects, hence this still is a mixed effect of them.

[43] Moreover, an endogeneity test of A is performed and presented in the next subsection. That is because before performing the endogeneity test, the establishment of the instrumental variable framework is necessary.

[44] Because of the repeaters / redshirters and the advanced pupils.

[45] I don't have enough data to possess this information, hence I rely on these few papers to make this assumption.

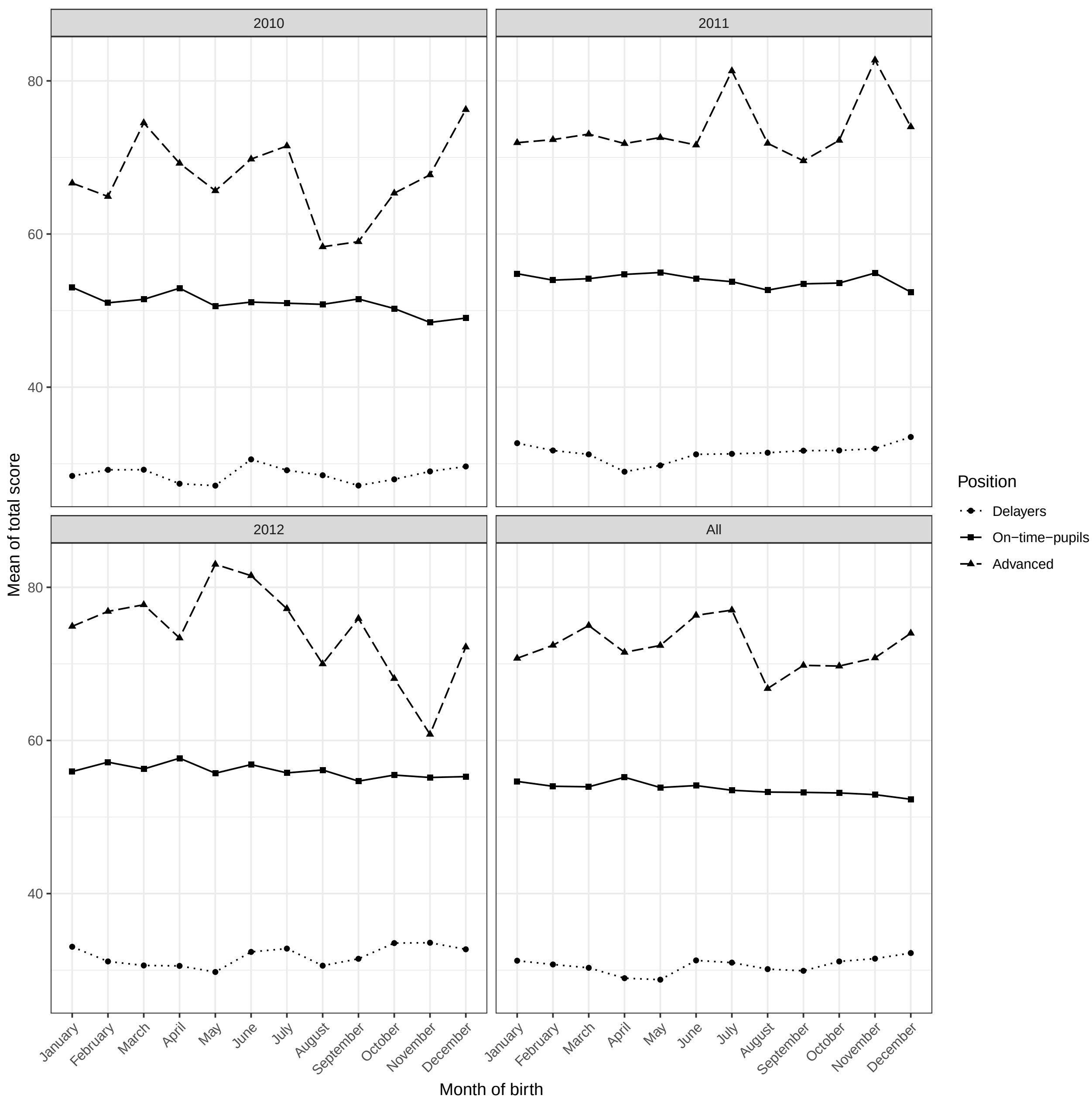

Figure 3: Mean of test score by month of birth and position

for the three cohorts combined, in average, repeaters / redshirters performs poorer than pupils who are on time. Combined with the observation in Figure 4 which shows that this institutional feature (retention / redshirting) is not a random one (because it appears that the proportion of delayers is positively correlated to the month of birth) ; one could expected the average test scores of the oldests to be decreased in a non randomly way (because the delayers are among the oldest ones within a grade), which is the source of the discussed downward bias in the estimation.

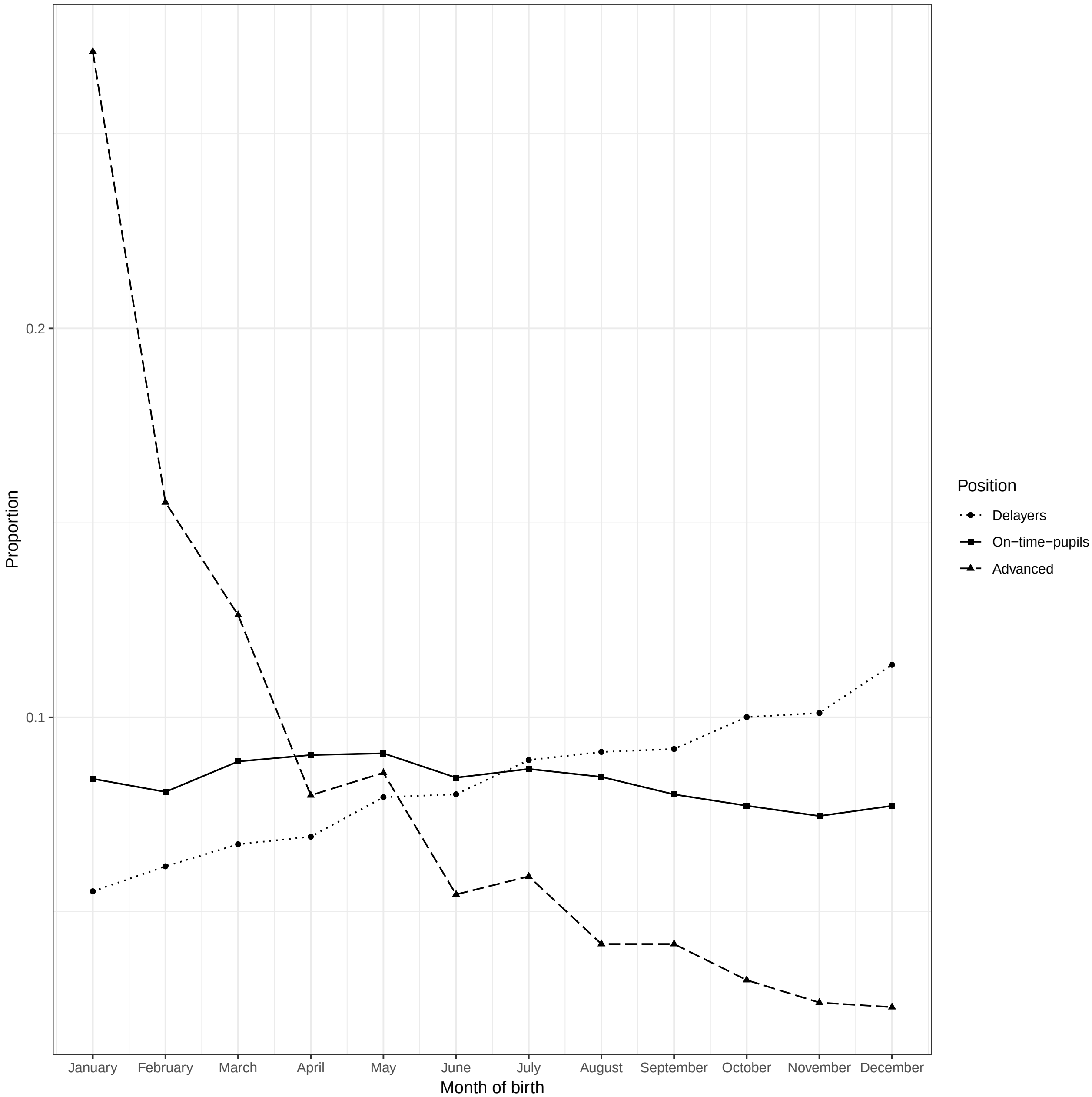

Figure 4: Month of birth, retention/redshirting and advance of pupils in grade 5

This drives me to an instrumental variable approach[46], an usual solution of the omitted variable problem, as presented in the upcoming paragraph.

3.2.2 Simultaneous equations model - Instrumental variable estimation

Suggested by Bedard and Dhuey (2006) and Grenet (2009), the assigned relative age, which reflect the relative[47] age if all the pupils were on time, can be used as an instrument for the endogenous age at test. This leads to the following simultaneous equations model :

$$\begin{cases} Y_i = \alpha_0 + \alpha_1.A_i + \alpha_2.X_i + \nu_i \\ A_i = \gamma_0 + \gamma_1.Z_i + \gamma_2.X_i + \eta_i \end{cases} \tag{2}$$

α_1 still is the parameter of interest in the system above. It is often estimated by 2SLS.[48]

More important, following G. W. Imbens and Angrist (1994), J. D. Angrist and Imbens (1995) and J. D. Angrist, Imbens, and Rubin (1996), there are three conditions to be verified so that the instrument Z is a valid one. First, Z needs to be randomly assigned. Knowing that Z is a linear transformation of the month of birth, this first condition is equivalent to the random assignation of month of birth. A potential pattern that would drive to the incompletion of this condition is the existence of a seasonality in month of births across the socio-professional categories of parents. In other words, higher-category parents may tend to have their children born in a particular quarter of the year whereas lower-category-parents children in another quarter of the year (Buckles and Hungerman 2013).[49]

In order to investiage this question, I performed a χ^2 test of comparison of proportions of months of birth across the different socio-professional categories of parents. For clarification, see Figure 5 where each panel illustrates the proportion of month of births within the corresponding socio-professional category sample. The test is performed to assess whether these proportions (comparing the categories) are equal or not. The former gives credit to the validity of the instrument because it suggests an absence of seasonality of birth

[46]One alternative solution is to use a exogeneous good proxy of the ability as a regressor, as in Pellizzari and Billari (2012). Given that I don't possess such variable but possess precise month of birth instead, using instrumental variable approach seems to be the most logical solution.

[47]Relative in the sens that it is expressed as a difference in month of birth compared to the theoretical youngest (born in December) within a grade.

[48]An alternative is to estimate it by a control function approach. See Hámori (2007) for example. Although, a control function approach is performed at the same time as the endogeneity test mentionned earlier. The results are presented in the appendix within the Table 13.

[49]For France, Grenet (2009) found that the month of birth had a seasonality pattern such as those born in April-May had on average the highest earnings (thus the highest socio-professional categories) parents and those born in August had on average the lowest earnings (thus the lowest socio-professional categories) parents.

by category. Since the null hypothesis is the equality of the proportions, and the p-value here equals to 0.113, at the 10% level, it can be assessed that the proportion of month of births does not change across the socio-professional categories of parents, which give credits to the use of the assigned relative age instrument.

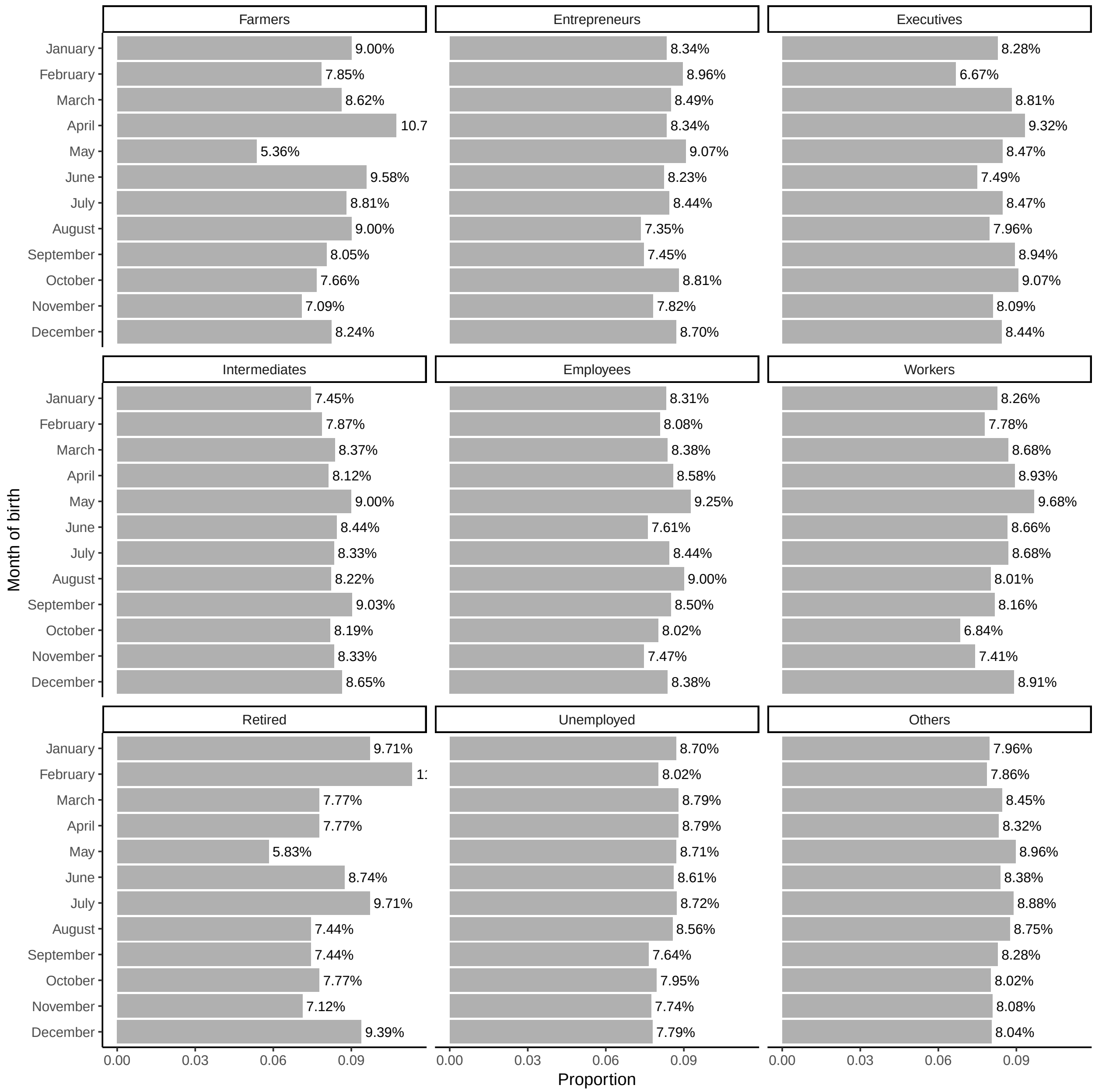

Figure 5: Proportions of month of birth by socio-professional category of parents across cohorts

Second, the instrument Z is required to have a non-zero average effect on the endogeneous variable A. Higher is the effect of the assigned relative age on age at test, stronger is the instrument. This condition can be verified quite straightfully with the estimation of the parameter $\gamma 1$ which is the average causal effect of Z on

A.[50]

Third, the assigned relative age is required to satisfy the monotonicity condition. This condition requires that the effect of the instrument on age at test does not have to change in sign over all the pupils. In our case, the effect is positive, this means that there should be no pupils such that the augmentation of Z (*i.e* being born earlier in the year) would lead to a decrease of its age at test. To evaluate if the monotonicity condition is likely to be verified, I check wether the variation of the sample analog of $E[A \mid m]$[51] is monotone with the variation of m. It makes sens since Z is a function of the month of birth. The Figure 6 shows that the variation of the age at test average with the month of birth is solely decreasing, which give credits the monotonicity assumption. The two straight lines without shaped points by month of birth in this figure represent the theorical straight lines formed by the means of age at test by month of birth if all individuals had their theoretical age. This is of interest to have an idea of the deviation caused by the presence of repeaters / redshirters and advanced pupils in the average age at test by month of birth within a cohort. Beside, a decortication of this Figure 6 is proposed in the appendix with the Figure 7. More precisely, Figure 7 illustrate observed versus theoretical mean ages at test by month of birth separately for the three cohorts (represented by the first three first panels of the figure) and for all cohorts combined (the last panel). In the latter, it is clearly highlighted that the actual mean age differences between individuals within the same grade (that is related to the slope of the lines) are lower than the theoretical mean age differences (the slope of the theoretical straight line is higher in absolute value than of the actual line ones). This observation joins that of (Grenet 2009) in the panel (b) of his Figure 7.[52] In addition, the theoretical straight line for the 2012 cohort have a higher intercept because of the later date of the test in this year. However, this is not an obvious result because it requires that within a cohort, the proportion of repeaters / redshirters by month of birth does not have a great increase from one month to another. Indeed, if it was the case, since these pupils are aged at least one year more than the others, they would increase the mean age by month of birth. According to Figure 4, the rise of repeaters / redshirters (labelled delayers in the figure) by month of birth is worrying for the monotonicity assumption validation. However, it appears that the rise was not large enough to discredite the mentioned assumption.

The IV estmations of the three[53] α_1 should lead to a LATE-type estimate of the impact of age on educational performances.

[50]This is the first stage regression, the results shown in Table 4 demonstrate that the assigned relative age have a strong effect on the age at test. This is explained by the fact that in each cohort, most of the pupils were one time. Indeed, for these individuals, the assigned relative age variable is a linear transformation of the age at test variable.

[51]m here represents the mere month of brith instead of the considering the day of birth.

[52]The explanation is, as expected, the presence of repeaters / redshirters and advanced pupils such that their month of of birth are not distributed uniformly across the year (non-randomness of the retention/redshirting/advancing phenomenon).

[53]Recall that regressions were run separately for the three cohorts.

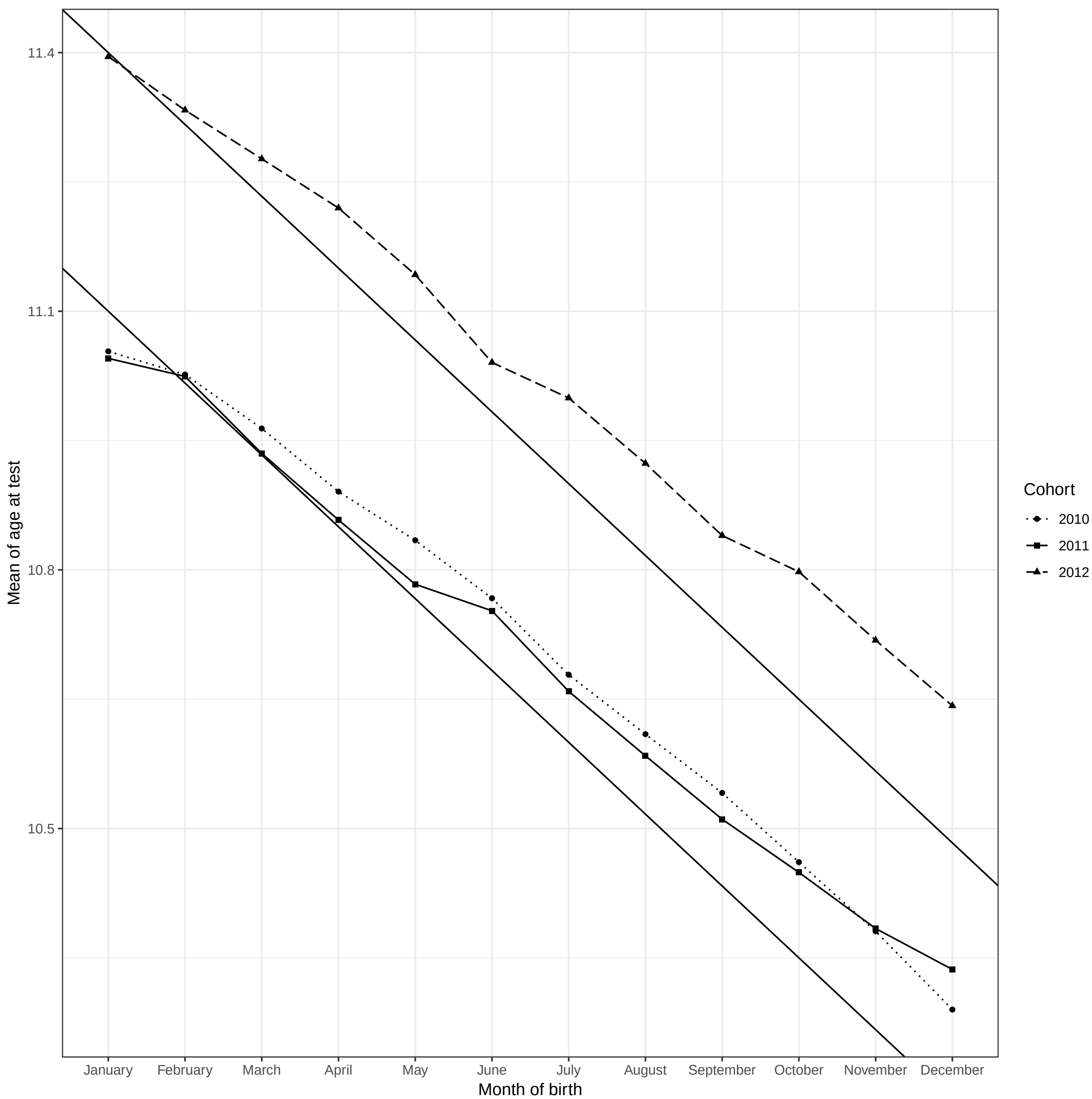

Figure 6: Mean of age at test by month of birth

3.2.3 Reduced form estimation

The reduced form equation is obtained after integring the first stage regression equation (the regression of A on Z and X) into the structural equation (the regression of Y on A and X). Therefore, the reduced-form is a regression of Y on Z and X, as shown in the following :

$$Y_i = \delta_0 + \delta_1.Z_i + \delta_2.X_i + \epsilon_i \tag{3}$$
$$\text{with} \quad \epsilon_i \equiv (\alpha_0 + \alpha_1).\eta_i + \nu_i$$

Note that the resulting error term is composed by the error terms of the first stage (η_i) and the structural equation (ν_i). The equation (3) is used to estimate the *« intention-to-treat-effect »* of the assigned relative age on the test scores : δ_1. Furthermore, since Z is a linear transformation of the month of birth variable, the reduced form equation can be taken for estimating the effect of month of birth on test scores.

The fundamental difference between estimating α_1 and estimating δ_1 is that for the former takes into account the not-on-time pupils[54] but not the *defiers*[55] whereas the latter considers as if all the pupils were on time and the defiers are taken into account. A more palpable interpretation is proposed in Section 4. Moreover, the link between the 2SLS estimation shown in the precedent paragraph and the reduced form estimation is[56],

$$\widehat{\alpha_1}(IV) = \frac{\widehat{\delta_1}}{\widehat{\gamma_1}}$$

3.2.4 Specification issues

In this paragraph, I adress the problem of the exact forms that should have the equations (1), (2) and (3). This question is relevant since the data are cross sectional ones. The main specification issues which would be investigated in the different models are the presence of class effects and the heteroskedasticity of the error terms.

Fixed effects

Since the data are three cross-sectional data and have grouping variables,[57] it is highly suspected that there

[54] The repeaters or those who entered earlier in school, and those who were in advance

[55] In the IV jargon, these are pupils who do not statisfy the monotonicity condition : those who would be younger as the instrument would Z increase in value (which is equivalent to the fact that the pupil would be born earlier in the year).

[56] See Wooldridge (2013), p. 516.

[57] School township identification, school identification and class identification

are effects of one of these variables. Besides, in the french educational context, as Piketty, Valdenaire, and others (2006) found a substantial negative effect of the class size in primary school, I focus my fixed effects analysis on the class identification grouping variable. An F-Test was implemented to detect the presence of class fixed effects in the equations (1), (2) and in the equation (3). With no surprise, the presence of class fixed effects is detected in all four equations.[58] As consequences, the final forms of the equations are presented below :

Structural equation estimation

$$Y_{ic} = \alpha_0 + \alpha_1.A_{ic} + \alpha_2.X_{ic} + \phi_c + \nu_{ic} \tag{4}$$

Simultaneous equations - Instrumental variable estimation

$$\begin{cases} Y_{ic} = \alpha_0 + \alpha_1.A_{ic} + \alpha_2.X_{ic} + \phi_c + \nu_{ic} \\ A_{ic} = \gamma_0 + \gamma_1.Z_{ic} + \gamma_2.X_{ic} + \psi_c + \eta_{ic} \end{cases} \tag{5}$$

Reduced form estimation

$$Y_{ic} = \delta_0 + \delta_1.Z_{ic} + \delta_2.X_{ic} + \omega_c + \epsilon_{ic} \tag{6}$$

ϕ_c, ψ_c and ω_c represent the class fixed effects in, respectively, the structural equation, the first stage equation and the reduced form equation.

Heteroskedasticity robust standard errors

Again, in the presence of cross-sectional datas, the sandard errors are likely to be heteroskedastic. Since heteroskedasticity can take several forms, I compute heteroskedasticity robust standard errors using the Arellano (1987) estimator. It is preferred over the White and others (1980) estimator because of the presence of the class effects.

A test for endogeneity

The test procedure, being at the same time a control function approach (Heckman and Robb 1986) is presented in Wooldridge (2013). It is a regression-based test which consists of : first, within the equation (5), computing the residuals $\widehat{\eta_{ic}}$ then performing the following regression

[58] All of the p-values are equal to zero.

$$Y_{ic} = \alpha_0 + \alpha_1.A_{ic} + \alpha_2.X_{ic} + \alpha_3.\widehat{\eta_{ic}} + error \tag{7}$$

Last of the test consists of testing the significance of $\widehat{\alpha_3}$. The null hypothesis is $\alpha_3 = 0$, which corresponds to the exogeneity of A. The results, presented in the next section, demonstrate the endogeneity of A, justifying the construction of the IV framework in equation (5).

4 Results and discussions

4.1 Endogeneity test of the age at test and first stage regression results

The estimations resulting from equation (7) and the first stage in equation (5) are presented in Table 4. Columns (1) to (3) reports for the three cohorts respectively the endogeneity test results, what is of interest in these are the significance of the estimates. Columns (4) to (6) corresponds to the first stage regression results. Since parameters corresponding to the covariates are not illustrated here, detailed versions are presented in the appendix within the Tables 13 and 14 respectively.

First, As highly expected, the $\widehat{\alpha_3}$ (columns (1) to (3)) are all revealed significant at the 1% level, which means that the age at test is indeed an endogeneous variable within the structural equation.

On the other hand, as discussed above, a way of investigating the validity of the instrument in our case is to measure its the prediction power of the independent varible of interest (the age at test). This is why the firs stage results[59] are of importance. It can be observed that the causal impact of the instrument on the endogeneous variable for the three regressions is very strong, since it is not less than 0.8 with significance at the 1% level.[60] This general result suggests that the instrumental variable − assigned relative age of the pupils − does not suffers from weak prediction power because *ceteris paribus*, being the oldest within a cohort instead of the youngest (a variation of 1 year in assigned relative age) causes on average the age at test to vary nearly about 0.8 year. Moreover, it can be observed that the F-statistics in the first stage regressions are all way above 10, which supports that the instrument is not « weak » (Staiger and James 1997).

[59] With covariates and class fixed effects

[60] +0.845 for the 2010 cohort, +0.806 for the 2011 cohort and +0.843 for the 2012 cohort

Table 4: Endogeneity test of age and First stage regression results

	Endogeneity test (dep.var : Total test score)			First stage regressions (dep.var : Age at test)		
	2010	2011	2012	2010	2011	2012
	(1)	(2)	(3)	(4)	(5)	(6)
Assigned relative age				0.845	0.806	0.843
				(0.012)***	(0.012)***	(0.012)***
$\widehat{\eta_{ict}}$	−1.121	−1.090	−1.019			
	(0.034)***	(0.032)***	(0.035)***			
F-statistic				543.921***	578.541***	552.245***
N	13,561	14,622	10,790	13,561	14,622	10,790
R^2	0.236	0.256	0.201	0.297	0.294	0.353
Adjusted R^2	0.196	0.218	0.148	0.260	0.258	0.310

Notes: ***Significant at the 1 percent level.
**Significant at the 5 percent level.
*Significant at the 10 percent level.

4.2 The impact of age at test and month of birth on test scores

Table 5 presents the results of the estimations of the three equations outlined earlier : (4), (5) and (6). The dependent variable is the total test score. For the three cohorts, columns (1) to (3) reports OLS estimates of the impact of age at test on test scores ; columns (4) to (6) corresponds to the IV estimates and columns (7) to (9) gives the reduced form estimations. Note that the estimates of the parameters of interest ($\widehat{\alpha_1}$ by OLS and IV and $\widehat{\delta_1}$ in the reduced form spectification) are placed within the first two lines of Table 5.

A first general pattern observed in the three tables mentioned above is the confirmation of the downward bias in the direct estimation within equation (4). In fact, the estimates of α_1 in column (1) to (3) all have negative signs[61], suggesting that with all other determinants of the total test score remaining constant, in average, being older at test causes a penalty between 0.4 and 0.5 of a standard deviation in the total test results. In comparison, the IV and reduced form estimates have positive signs and have lower absolute values.[62]
Second, a decrease of the penalty is reported across the three cohorts : between the apparent penalty of beeing older in the 2010 cohort and the 2011 cohort, the 2011 cohort penalty is 0.014 lower. And between the the 2011 and 2012 cohort, the 2012 penalty is 0.085 lower. This could be explained by the change in the form and content in the assessments through the three school years.

[61] −0.509, −0.495 and −0.410 of a standard deviation
[62] The IV and reduced form estimates take values between +0.229 and +0.304 of a standard deviation.

Table 5: Main regressions results

	Dependent variable : Total test score								
	OLS			IV			RF		
	2010	2011	2012	2010	2011	2012	2010	2011	2012
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Age at test	−0.509	−0.495	−0.410	0.304	0.302	0.271			
	(0.015)***	(0.015)***	(0.019)***	(0.033)***	(0.031)***	(0.034)***			
Assigned relative age							0.257	0.244	0.229
							(0.026)***	(0.024)***	(0.028)***
Sex - Male	−0.218	−0.229	−0.215	−0.288	−0.289	−0.255	−0.262	−0.267	−0.241
	(0.015)***	(0.014)***	(0.016)***	(0.017)***	(0.015)***	(0.017)***	(0.016)***	(0.014)***	(0.016)***
SPC - Entrepreneurs	0.112	0.163	0.004	0.157	0.213	0.039	0.137	0.203	0.029
	(0.059)*	(0.077)**	(0.070)	(0.068)**	(0.079)***	(0.073)	(0.063)**	(0.076)***	(0.071)
SPC - Executives	0.436	0.484	0.393	0.550	0.600	0.493	0.508	0.562	0.462
	(0.061)***	(0.073)***	(0.065)***	(0.070)***	(0.076)***	(0.069)***	(0.065)***	(0.072)***	(0.065)***
SPC - Intermediates	0.181	0.295	0.138	0.265	0.356	0.188	0.240	0.337	0.175
	(0.059)***	(0.075)***	(0.065)**	(0.067)***	(0.077)***	(0.068)***	(0.062)***	(0.074)***	(0.065)***
SPC - Employees	−0.003	0.101	−0.059	0.026	0.125	−0.050	0.017	0.119	−0.051
	(0.055)	(0.071)	(0.064)	(0.063)	(0.072)*	(0.067)	(0.058)	(0.069)*	(0.064)
SPC - Workers	−0.104	−0.040	−0.157	−0.084	−0.052	−0.166	−0.090	−0.046	−0.158
	(0.056)*	(0.074)	(0.062)**	(0.065)	(0.078)	(0.066)**	(0.060)	(0.074)	(0.063)**
SPC - Retired	0.218	0.346	0.228	0.235	0.390	0.212	0.226	0.372	0.222
	(0.093)**	(0.104)***	(0.103)**	(0.108)**	(0.116)***	(0.111)*	(0.100)**	(0.108)***	(0.105)**
SPC - Unemployed	−0.231	−0.175	−0.352	−0.282	−0.237	−0.417	−0.264	−0.213	−0.387
	(0.054)***	(0.070)**	(0.063)***	(0.062)***	(0.072)***	(0.066)***	(0.058)***	(0.068)***	(0.063)***
SPC - Others	−0.351	−0.373	−0.286	−0.430	−0.461	−0.360	−0.397	−0.426	−0.315
	(0.056)***	(0.071)***	(0.121)**	(0.066)***	(0.074)***	(0.125)***	(0.061)***	(0.070)***	(0.120)***
N	13,561	14,622	10,790	13,561	14,622	10,790	13,561	14,622	10,790
R^2	0.168	0.194	0.144	0.026	0.051	0.047	0.105	0.135	0.109
Adjusted R^2	0.125	0.153	0.088	−0.026	0.003	−0.016	0.058	0.091	0.050

Notes:

***Significant at the 1 percent level.
**Significant at the 5 percent level.
*Significant at the 10 percent level.

Table 6: Reduced form, first stage and IV estimates relationship

Estimates of parameters of interest	2010 cohort	2011 cohort	2012 cohort
Reduced form	0.257	0.244	0.229
First stage	0.845	0.806	0.843
IV	0.304	0.302	0.271

[a] For example, in the first column, 0.257/0.845 = 0.304

In the other hand, the relationship between the reduced form, the first stage and the IV estimate is verified.[63] As it may be confusing to check it through the different regression tables, the reduced form, first stage and IV estimates, for the three cohorts ar reported in Table 6.

Next, when focusing on the other independent variables, first the sex variable, the difference in total test score caused by the fact of being a male pupil instead of being a female is reported in the third line of the Tables 5. Across cohorts, male pupils suffers from a test score penalty around 0.2 of a standard deviation. Also, the IV estimates in the simultaneous equation model give the highest magnitude of such a penalty between the three models for each cohort.[64]

Following the same logic, about the socio-professional category of parents variable, the reference value corresponds to the farmers. The estimates are reported in Table 5, from line 4 to line 11. Additionally, as the observations that take the value of « Others » basically means unkown values, this is not very interesting to check. The pattern of significance of the estimates of the impacts of socio-professional categories of parents compared to being a farmer's child is heterogeneous. Thus, no generic observation can be made about it.[65] This would potentially corresponds to a heterogeneity of the impact of age at test on test scores for different categories. However, the concern about significance can be hard one to draw interesting conclusion from because it can dramatically change with the refrence value of the parent's socio-professional category variable.[66]

Still, if the focus is made on the estimates, some patterns of the estimates can be noticed. As a practical matter, refer to Figure 8 in the Appendix which illustrates with barplots the magnitudes of the estimates of the impact of socio-professional category of parents on total test scores. First, as could be expected,

[63] Recall that, δ_1 being the causal effect of assigned relative age on test scores (reduced form) and γ_1 the impact of the assigned relative age on the age at test, $\widehat{\alpha_1}(IV) = \frac{\widehat{\delta_1}}{\widehat{\gamma_1}}$

[64] For example, in the 2012 cohort, the IV estimate of the impact of being a male instead of a female on test score is equal to −0.255 of a standard deviation while the structural equation and reduced form estimates of the same causal impact are respectively equal to −0.215 and −0.241.

[65] For example, the farmers being the reference value, the estimates who have significance at 1% level in the three models within the 2010 cohorts corresponds to the executives, the intermediates and the unemployed. This pattern does not reoccur in the 2011 or 2012 cohorts.

[66] For informative purspose, a list of the parent's socio-professional category's impact on total test score with no significance above the 10% level is reported in the appendix within Table 7.

Table 7: Estimates of parents socio-professional category on total test score with no significance

spc	IV	OLS	RF
2010			
Employees	0.026	-0.003	0.017
Workers	-0.084	NA	-0.090
2011			
Employees	NA	0.101	NA
Workers	-0.052	-0.040	-0.046
2012			
Employees	-0.050	-0.059	-0.051
Entrepreneurs	0.039	0.004	0.029

[a] NA : there is no estimate that is reported to be non significant.

being an executive child instead of a farmer's child leads to the most test score advantage compared to the other categories. This advantage varies between +0.393 and +0.631 of a standard deviation depending on the estimation model and the cohort. Second, being an unemployed's child leads to a penalty in test scores compared to being a farmer's child, which is a quite unsurprising result because in terms of life quality[67], a farmer is very likely more apt to provide better of that than an unemployed do. The same observation can be made for the worker's children.These two patterns remain the same regardless of the model or the cohort. The penalty of being an averge unemployed's child instead of being a farmer child varies between -0.417 and 0.175 of a standard deviation. Third, the intermediates children have an advantage compared to the farmers children but this advantage is lower than the advantage of the executives children. Again, this latter pattern remains the same regardless of the model or the cohort. Its value varies from +0.138 to +0.364.

From another point of view of comparison between models, it can be assessed that the IV model give the highest estimates in absolute value compared to the structural equation or the reduced form. Also, the structural equation estimates yields the lowest estimates. This pattern is invariant either for the executives, intermediates or unemployed and invariant across cohorts.

Distinguishing between french and mathematics test scores as depedent variables in results

Tables 8 and 9 present estimates of the effects of ages, respectively, french test scores and mathematics test scores. A clear evidence from these results is that the effect on french scores are about 0.01 of a standard deviation higher. This is not a significative difference in terms of standard deviation. A policy matter implication would be that it is not interesting to distinguish between french or mathematics topics in policy making.

[67]Which is known to be positively correlated with educational performances

Table 8: Regression results with french scores as dependent variable

	Dependent variable : french test scores					
	IV			RF		
	2010	2011	2012	2010	2011	2012
	(1)	(2)	(3)	(4)	(5)	(6)
Age at test	0.285 (0.033)***	0.280 (0.031)***	0.256 (0.035)***			
Assigned relative age				0.241 (0.027)***	0.226 (0.024)***	0.215 (0.028)***
N	13,561	14,622	10,790	13,561	14,622	10,790
R^2	0.040	0.070	0.064	0.119	0.149	0.124
Adjusted R^2	−0.010	0.023	0.002	0.073	0.106	0.066

Notes: ***Significant at the 1 percent level.
**Significant at the 5 percent level.
*Significant at the 10 percent level.

Table 9: Regression results with mathematics scores as dependent variables

	Dependent variable : mathematics test scores					
	IV			RF		
	2010	2011	2012	2010	2011	2012
	(1)	(2)	(3)	(4)	(5)	(6)
Age at test	0.295 (0.032)***	0.301 (0.032)***	0.264 (0.035)***			
Assigned relative age				0.250 (0.026)***	0.243 (0.024)***	0.222 (0.028)***
N	13,561	14,622	10,790	13,561	14,622	10,790
R^2	0.013	0.028	0.026	0.075	0.100	0.075
Adjusted R^2	−0.039	−0.021	−0.038	0.026	0.054	0.014

Notes: ***Significant at the 1 percent level.
**Significant at the 5 percent level.
*Significant at the 10 percent level.

4.3 Subgroup analysis

The two relevant grouping variables on which I reestimate the equations (4), (5) and (6) (with necessarly removing the grouping variable from the regression covariates) are the sex and the socio-professional category of parents. This is performed in order to evaluate the robustness of the previous subsection principal results.

Sex subgroup estimations

Table 15 in the appendix reports the estimates of age at test and relative age impact on test scores by sex for the 2009-2010 cohort, Table 16 and Table17 correspond respectively to the 2010-2011 and 2011-2012 cohorts. In these three tables mentionned above, columns (1) and (2) give the IV estimates for, respectively females and males ; while columns (3) and (4) provide the reduced form estimates.

First, the signs of the estimates remains positive in the subgroup estimations as when the estimations are performed with the whole population. Also, they are all significant at the 1% level. An another noticeable feature is that the female pupils benefit more in total test scores than male pupils do : the estimates, whether from the IV models or the reduced form models are higher for females in the 2009-2010 and 2010-2011[68] For the 2011-2012 cohort, the impact of age at test on total test scores does not differ significantly. This is in line with the results of Grenet (2009) in his Table 2. This latter pattern is probably due to the lack of observations about the sex variable in the 2011-2012 cohort. These are highlighted with comparison between column (1) and column (2) for the IV model, and between column (3) and column (4) in the Tables 15, 16, 17.

Next, one can compute the differences between the estimates from whole population and from subgroup estimations to assess the heterogeneity of the age at test effect through sex subgroups. This is what the Table 10 illustrates.

Except for the 2011-2012 cohort, there is a difference in magnitude of at least 0.04 and at most 0.07 of a standard error. In fact, for the 2011-2012 cohort, it seems that there is no significant difference. These differences (in the last two lines of Table 10) can be dropped out of the next analysis. In addition, while the female sex sample estimates are higher (±0.05 of a standard error higher) compared to the whole sample estimates, the male sex sample estimates are lower (±0.06 of a standard error lower). This suggests a heterogeneity of the impact of age at test across sex subgroups in Reunion Island, which implicates that age policies should potentially be conducted with sex distinction considerations.

Socio-professional category of parents subgroup estimations

[68] The females benefit about 0.1 of a standard error more than the males.

Table 10: Deviations of sex subgroup estimates from whole sample estimates

Models	Females	Males
2010		
IV	0.06	-0.07
RF	0.05	-0.05
2011		
IV	0.05	-0.07
RF	0.04	-0.06
2012		
IV	0.01	0.00
RF	0.01	0.00

[a] For example, in the 2010 cohort, the difference between the female sex sample IV estimate of the age at test impact on test score (+0.365) and the whole sample IV estimate (+0.304) equals to +0.06 of a standard error.

In the appendix, Tables 18 and 19 give for the 2009-2010 cohort, respectively the IV and reduced form estimates of the impact of age at test on total test scores by socio-professional category of parents. Tables 20 and 21 do the same for 2010-2011 cohort. In the same maneer, Tables 22 and 23 give these estimates for the 2011-2012 cohort.

There seems to be no interesting significance at 1% level pattern of the estimates across the models (IV or reduced form) and cohorts. Indeed, the among the eight subgroups (Farmers, entrepreneurs, executives, intermediates, employees, workers, retired and unemployed) estimates of age at test or relative age on total test scores that always yield significant estimates at the 1% level regardless of the model (IV or reduced form) or the cohort are the employees children group and the unemployed's children group. This pattern remains verified even when considering significance at least at 5% level. The corresponding estimates are given in columns (5) for the employees children group and in column (8) for the unemployed's children group. Compared to each other, except in the 2009-2010 cohort, the employees children appear to benefit more in test score from being older (between +0.24 and +0.29 of a standard deviation) than the unemployed do (between +0.18 and +0.27 of a standard deviation).

For illustration purpose, since checking significances at least at the 5% level across all cohorts does not seem to yield to interesting informations, reviewing it cohort by cohort and at least at the 10% level may uncover more interesting patterns of the estimates. In fact, within the 2010 cohort, for the two models (IV and reduced form), the subgroup estimates that are reported to be significant are the farmers, executives, employees, workers and unemployed's children. Concerning the 2011 cohort, the subgroups verifying the same condition are children whose parents are executives, intermediates, employees and unemployed. Last, in the 2012 cohort, the subgroups are the entrepreneurs, intermediates, employees, workers and unemployed's children.

In 2010, among the children whose parents are farmers (column (1)), executives (column (3)), employees (column (5)), workers (column (6)) or unemployed (column (8)) ; the impact of age at test on total test scores is reported, in general to be positive and very substantial. Indeed, for the IV model, the estimates vary from +0.28 to +0.94 of a standard error while from the reduced form estimation, these vary from +0.22 to +0.95 of a standard error. The group that seems to benefit the most from being one year older (IV) or being the oldest ones compared to the youngest ones (reduced form) are farmers children in which the effect is almost a standard deviation. This is equivalent to almost 3 years of learning. Since this result appears to be anormally high, it is careful leave this to discussion because of the number of observations (198 in 2010, see the third lines of Tables 18) which is considerably lower than the other subgroups (all more than 700) and the very low R-squared. This pattern is likely to be retrieved in the IV subgroup estimates across the 2011 and 2012 cohorts, which is problematic for interpreting and discussing the results. Hence, for the next reporting, I suggest to limitate the analysis to the reduced form estimates, excluding the farmers children (Tables 19, 21 and 23), from which the R-squared are generally better.

First, in the 2010 cohort (Table 19), among the children whose parents are executives (column (3)), employees (column (5)), workers (column (6)) or unemployed (column (8)), the estimates magnitudes vary from +0.2 to +0.3 of a standard error. the former corresponds to the executives children where the latter corresponds to the workers children. Difference in magnitude between the employees children estimate and the unemployed's one does not appear to be substantial (about 0.02 of a standard error).
Following, within the 2011 cohort (Table 21), among the executives (column (3)), intermediates (column (4)), employees (column (5)) or unemployed's (column (8)) children, those who benefits the most from being the oldest ones within the grade compared to the youngest ones[69] are the pupils whose parents are intermediates (with +0.3 of a standard error estimate's value) and tho who benefit the less are the unemployed's children (with +0.2 of a standard error estimate's value). Note that if the workers children estimate was significant, it would correspond to the lowest magnitude. The reason behind the fact that it was reported non significant is very likely the lack of observations of workers children in 2011, since in 2010, they account for 1523 obsrevations where in 2011, the corresponding number is only 957 (a difference of 37% from 2010 to 2011). This problem does not occur in the 2012 cohort in which workers children account for 1412 observations.

[69] The highest magnitude among the corresponding estimates

Similarly as stated for the 2010 cohort, the difference in magnitude between the employees children estimate and the executives'one is about 0.05 of a standard error, which is not considerable.

Last, in 2012, among the children whose parents are entrepreneurs (column (2)), intermediates (column (4)), employees (column (5)), workers (column (6)) or unemployed (column (8)), those who benefit the most from being the oldest ones within the 2012 cohort compared to their youngest peers are the entrepreneurs children. On the opposite, those who benefit the less in test scores are the unemployed's children[70]. In addition, the ascending ranking of the remaining subgroup estimates magnitudes are as follow : in the first place there are the workers children, then the employees children and the third place is attributed to the intermediates children. These are outstandingly the most expected results since this ranking is very likely to match a ranking of « quality of life » (which is known to impact eductional performances).[71]

In another point of view, similarly to the Table 10, the Table 11 reports the difference of subgroup estimates from whole sample estimates. A positive value in this table means that the subgroup estimate is higher than the whole sample estimate. A first check to be performed is, within the 2010 cohort and the reduced form line in Table 11, among the executives, employees, workers or unemployed's children, the values are all positives except for the executives children. Also, the highest difference is attributed to the workers with the value of +0.05 of a standard error. On the opposite, the employees reduced form estimate deviates from the whole sample estimate only by +0.01 of a standard error.

Then, in 2011, among the individuals whose parents are executives, intermediates, employees or unemployed, the only negative value is of the unemployed children group. In addition, the employees children deviation is reported to be zero[72], meaning that the effect of being relatively the oldest is the homogeneous between the whole sample and the employees subsample. The highest deviation is attributed to the intermediates children.

Last, within the 2012 cohort, among the entrepreneurs, intermediates, employees, workers or unemployed's children, an outstanding value is of the entepreneurs children group whereas the deviation from the whole sample estimate equals to 0.16. In addition, two negative deviations are reported : those of the workers and the unemployed's children. The lowest positive value is of the children having employees parents. Having in mind that within the 2012 cohort, there is a considerable amount of missing observations about the sex variable, this could be the explanation of the fact that among the three cohorts, only the entrepreneurs (reduced form) estimate was yielded significant.[73]

[70] As stated eralier for the 2011 cohort

[71] In terms of life quality (mostly because of their incomes) : entrepreneurs > intermediates > employees > workers > unemployed.

[72] More precisely, the difference, , is $0.241 - 0.243 = -0.002$, which is neglectable on a unit of a standard error.

[73] If we take a relook on the reduced form deviations for the entrepreneurs children, their magnitude are also outstanding, but they have negative sign. Though, recall that they are excluded from the present analysis due the absence of significance in the estimates of relative age effect on test score.

Table 11: Deviations of parents socio-professional categories estimates from whole sample estimates

	Farmers	Entrepreneurs	Executives	Intermediates	Employees	Workers	Retired	Unemployed
2010								
IV	0.64	-0.19	-0.02	-0.21	-0.02	0.10	2.56	0.05
RF	0.70	-0.16	-0.03	-0.18	0.01	0.05	0.58	0.03
2011								
IV	-0.45	-0.28	0.09	0.05	-0.01	-0.06	4.99	-0.04
RF	-0.34	-0.22	0.05	0.06	0.00	-0.05	1.06	-0.04
2012								
IV	0.67	0.15	-0.09	0.03	0.02	-0.04	2.17	-0.05
RF	0.58	0.16	-0.08	0.04	0.01	-0.03	0.55	-0.05

[a] For example, in the 2011 cohort, the difference between the unemployed's children sample reduced form estimate of the age at test impact on test score (+ 0.206) and the whole sample reduced form estimate (+ 0. 243) equals to - 0.04 of a standard error.

These results suggest that the effect of relative age on test score (corresponding to the reduced form results analysis) is slighly heterogeneous across parents socio-professional category of the individuals while the effect of being absolutely older (corresponding to the IV results analysis) is hard to draw a conclusion in terms of heterogeneity across socio-professional category of parents subgroups. Hence, under reserve of further studies, it may be not interesting to draw policy makings considering the socio-professional of parents. This seems to be an unusual conclusion. The potential explanation is that there were some « unkown »[74] (socio-professional) categories among the data. These values were agregated with missing values in the « Others » label (Figure 2) which could lead to considerable change in the results.

4.4 Discussions

In order to make this subsection appear more coherent with what would be adressed in, let us sum up the findings in the previous subsections to begin with. First, the age at test is with no surprise, diagnosed being endegenous within the structural equation. The reasons standing behing this fact was raised several times : the presence of repeaters and redshirters.

Then, with the assigned relative age measured reported in years unit used to instrument the observed age at test, the former fulfill the condition requiring the instrument to have a non-zero average causal effect on the latter. In addition, the instrumental variable have a strong prediction power on the observed age at test of grade 5 pupils in Reunion Island for the 2009-2010, 2010-2011 and 2011-2012 cohort. This was highlighted by performing the first stage estimations.

Thereafter, main regression results are reported : the OLS estimations all yielded to negative estimates of

[74] Different from missing values

the age at test effect on total test scores as expected, which reflect the downward biais discussed earlier ; the 2SLS estimates was reported to be all positive remarkably substantial in magnitude ; the reduced form estimates are also all revealed positive but with lower magnitudes than from 2SLS estimates in general. Following, the age effects are revealed to be clearly heterogenous across sex subgroups with females pupils benefiting in total test scores considerably more from being older than males pupils do. Assessing essential heterogeneity across socio-professional category of parents subgroups is much more inconlusive, despite of the evidence of a slight heterogeneity of the age effects of total test scores.

What is measured ? : Discussions about the real sens of the different estimates

There is a flow of papers that are concerned about theory behind instrumental variables, and in which the very basic concept is the notion of *« counterfactual outcomes »* (Rubin 1974). One of the most contributive in theorical identification with instrumental variable is G. W. Imbens and Angrist (1994). It seems that from this paper was derived the two following : J. D. Angrist and Imbens (1995) and J. D. Angrist, Imbens, and Rubin (1996).[75] Refering to these, we could define our first stage parameter of interest (γ_1) as the proportion of *compliers*.[76] Recall that a complier is a counterfactual-type vocabulary which is defined as an individual such that if the instrument would take another value ; in a counterfactual world, this individual's treatment would take the corresponding value induced by the instrument. If this sentence is translated into the framework of the present study, it would simply mean that a complier is an individual such that if his assigned relative age would decrease (or equivalently the month of birth would increase[77]), meaning that if the individual *would be* born later in the year, he *would have* lower age at test. When taking this definition into a proportion perspective, we could make a liaison between this and the position (timing in age) within a given grade : the proportion of compliers could be strongly correlated with (but does not exactly equals) the proportion of on-time pupils. In fact, solely for this type of position, being born later in the year intrinsically means being younger when sitting the tests. This idea is supported by the apparent closeness between $\widehat{\gamma_1}$ (Table 4) and the proportions of on-time individuals (second set of statistics in Table 1). In Hámori (2007), it is stated that the proportion of on-time individuals equals the sum in proportion of compliers and *always-takers*. This latter means that being born earlier in a counterfactual world would lead to a lower age at test anyway. This is in line with the feature of pupils that are in advance. Hence, it makes sens to think that the proportion of compliers and always takers is close to the proportion of on-time and

[75]The first produces an instrumental variable identification with both of the treatment variable and instrument being binary variables ; the second completed this with providing instrumental variable identification when the instrument is a multivaluted variable and the treatment is a continuous one.

[76]Note that unfortunately, equivalent papers of J. D. Angrist and Imbens (1995) but with both the instrument and the treatment variable being continuous could not be retrieved yet, if there is.

[77]This can be more easily visualized when one remember that $Z = \frac{12-m}{12}$.

advanced individuals. Better understanding of this potential link is aim for future researches. Consequently, the estimates of the parameter of interest, $\widehat{\alpha_1}$ would indeed reflect a LATE-type estimand.[78] The interest of discussing all of these is the following questioning : if the highlighted quantities (2SLS estimates) are solely valable for a certain subpopulation (logically, in the present reasoning, this subpopulation corresponds to the one-time pupils), how policy maker should react to that information regarding of their objectives ? A potential answer is the suppression of grade retention in primary schools, but this is subject to a very large debate. See Alet, Bonnal, and Favard (2013) for a review of the debate which opposes pros and cons of grade retention.

Let us now focus on the sens of the parameter of interest in the reduced form framework : δ_1. As mentioned earlier, the reduced-form equation (equation (6)) is obtained from regressing the dependent variable on the instrument and covariates. $\widehat{\delta_1}$ designates then the causal impact of the theorical age position compared to the youngest within a grade, net of repetition and redshirting (Bedard and Dhuey 2006). This is intuitive since the computation of the instrument rules out the year of birth which determines the position of a pupil within a given grade.[79] Always following Bedard and Dhuey (2006), note that in the present framework, an estimate of δ_1 captures inevitably potential season of birth fixed effects on total test scores. This is because the data covers a unique country with a unique school entry rule (Reunion Island). Indeed, in their framework, the authors have at their disposal multiple OECD countries (then with multiple school entry rules). Instead, as the data of the present study provides school and classrooms identifications, the class fixed effects are taken into account. In a LATE framework perspective, this would be utlimately an *itention to treat*, transposed into a system where the instrument as well as the treatment variable are continuous.

From another perspective, recall that the age effect measured by 2SLS is a mix of age at test and age of entry effects since these two are perfectly colinear and the data doesn't provide, within a cohort, any random variation of the age at test given a fixed age of entry.

Some comparisons with other studies estimates

First, let us provide necessary and minimal informations that are of interest in this paragraph about these three papers of comparison : P. Puhani and Weber (2005), Bedard and Dhuey (2006) and Grenet (2009).
P. Puhani and Weber (2005), using Progress in International Reading Literacy Study (PIRLS) 2001 dated data in Germany to adress, at least at the limit of what is of interest in this paragraph, the effect of relative age on PIRLS test scores in grade 4 using assigned relative age as an instrument. On their side, Bedard and

[78] A formal writing of this estimand is for the moment hard to provide because of the potential lack of methodological papers adressing this, as I already stated earlier.
[79] For the present study, it corresponds each grade 5 for the three school years.

Dhuey (2006) used TIMSS 1999 dated, supplemented with Early Childhood Longitudinal Study (ECLS) and National Education Longitudinal Study (NELS) data to measure the causal effect of relative age on mathematics and science test scores in grade 4. As well, they make use of the assigned relative age as an instrument. Their results are free from season of birth effects as stated above. Last, Grenet (2009) exploited the *Panel Secondaire de l'Éducation Nationale* (PSEN) 1995 dated french data to highlight the effect of age on global, mathematics and french test scores in grade 6. Note that within all of these, the test scores are standardized ones, making them comparable with test scores within the data of the present study. Also, all of these use assigned relative age as instrument of age to overcome the endogeneity of the latter. In addition, the dating of the data used within these three studies are quite close each other (respectively 2001, 1999 and 1995). Thus, the selection made for comparison is likely to be well justified. Refer to the Table 10 - columns (1a) and (2a) - line « Exo3 » of P. Puhani and Weber (2005) ; Table 3 - columns (2) and (4) of Bedard and Dhuey (2006) and Table 1 - columns (4) and (6) - lines « Year 6 : Maths » and « Year 6 : French » to the upcoming discussion about the estimates.

A striking feature after giving a first look at the estimates of interest is that they all have positive sign, regardless of the dependent variable specification (reading / french, mathematics or global) or model specification (IV or reduced form). Moreover, the IV estimates magnitudes have in general higher values than the reduced form estimates. Also, they are all significant at the 5% level or better. Even if this is probably not of much relevance, my results are in line with this feature. What would be a first interesting subject of discussion is the magnitudes of the different estimates.

Concerning the IV results, one can compare these of the present study for the mathematics score as dependent variable with the results of Bedard and Dhuey (2006) and Grenet (2009). Beside England, Iceland, Japan, Norway (characterized by a very low retention rates), New Zealand (characterized by a high proportion of advanced pupils) and United States (with their higher proportions of redshirters than of repeaters), the IV estimates of Bedard and Dhuey (2006) vary from +0.19 (Czech Republic) to +0.258 (Portugal) of a standard error while the corresponding estimate in Grenet (2009), when the age at test is reported in years equals to +0.276 of a standard error. These magnitudes appears generally to be lower than mines (Table 9). However, the maximum difference with my estimates among these of the other studies is reported to be around one tenth of a standard error and besides these differences are smaller. Hence, one could conclude that despite of differences in institutional characteristics across studies, shaping the effect of being one year older at the moment of sitting the test on mathematics test scores, the mentioned effect is stable. Concerning the case when the french score is specified as the dependent variable, my IV estimates are reported to be slightly lower than these of Grenet (2009). Again, the difference's value is about 6 percent of a standard deviation, which

Table 12: Summary of subgroup estimates by socio-professional category of parents

	« Privileged » categories			« Underprivileged » categories				
	Executives	Entrepreneurs	Intermediates	Farmers	Employees	Workers	Retired	Unemployed
2010								
IV	0.283	NA	NA	0.948	0.286	0.402	NA	0.356
RF	0.227	NA	NA	0.952	0.269	0.311	NA	0.287
2011								
IV	0.390	NA	0.351	NA	0.290	NA	5.289	0.263
RF	0.293	NA	0.305	NA	0.241	NA	1.302	0.206
2012								
IV	NA	0.424	0.306	0.940	0.288	0.235	NA	0.217
RF	NA	0.387	0.273	0.807	0.240	0.203	NA	0.181

[a] NA : not significant.
[b] The dependent variable is the total test score.
[c] The privileged-underprivileged classification is the same as in Grenet (2009).

is likely to be unimportant. On the opposite, the IV estimates of P. Puhani and Weber (2005) are higher with a greater difference (having total test scores as dependent variable).[80] One potential explanation of why Grenet (2009)'s estimates are lower and P. Puhani and Weber (2005) higher is the decreasing pattern of age effects over time. In fact, these of Grenet (2009) was based on grade 6 pupils while P. Puhani and Weber (2005) on grade 4 (mines are concerning grade 5 pupils).

Within sex subgroup analysis, my IV estimates are also consistent with Grenet (2009)'s and P. Puhani and Weber (2005)'s findings : female pupils benefits more from being one year older at test than male pupils do. Again, the magnitudes are revealed higher in the present study probably because of the difference in the result's timing combined with the decreasing pattern of age effects over time. Concerning the parent's socio-professional category subgroup estimates, as a matter of comfort, refer to Table 12 for a summarised visualization. The higher magnitudes of the IV estimates of the present study compared to these of Grenet (2009) is retrieved one more time. More interesting, the individuals whose parents are classed privileged appear to benefit more than those whose parents are labelled unprivileged, except for the 2010 cohort. In fact, the only IV estimate that was yielded significant is the one corresponding to the executives category of parents in 2010. Compared to the employees or unemployed's children groups, the latter appears to benefits slightly higher. Results in grade 6 within Grenet (2009) appears to be quite inconclusive as well.[81]

Last, if the focus is made on reduced form results, an important feature remains observed : higher estimates for females than for males. A contrast between my reduced form results and these of Grenet (2009) is that while within the former, higher estimates are reported when specifying the mathematics score as dependent

[80] About +0.4 of a standard error
[81] Because the advandage is switched depending on the specification of the dependent variable.

variable, the latter reports the opposite. Given that the educational systems in the two studies are the same, this difference should reflect a difference in pupils background (between Reunion Island pupils and whole France pupils) instead of a difference in institutional background. What would reflect institutional background difference is with noticing that the reduced form estimates within Bedard and Dhuey (2006) are ultimately lower in magnitudes.

5 Conclusion

In conclusion, the causal relationship between age and national test scores in grade 5 is positive and important for Reunion Island even though the data doesn't allow the identification of which component is the most important. Fortunately, based on previous works, we can give more credit to the hypothesis that the age at test is likely to be the principal driver of the age effect. In addition, the effect of age on test scores is heterogeneous accross the sex, such as female pupils benefit more from being older than do male pupils. Also, the results suggest that the effect is slightly heterogeneous with the socio-professionnal category of the parents. Last, the effect does not seem to be importantly heterogeneous when one distinguish between french test scores and mathematics test scores since the difference is about one hundred of a standard deviation.

Based on these results, some policy suggestions can be made. Given the aim of increasing the grade 5 pupils test scores (which is important for later educational outcomes, according to the literature), policy makers can influence on few possibilities. First, since the age at test is likely to be the most important driver of the age effect on educational outcomes, they could either normalize the test scores by month of birth (correcting for the inequality in test scores solely caused by the date of birth of the child, which is clearly unchosen by pupils). They could also make pupils with different ages pass the national test at different times such as everyone is taking the test at the exact same age. Among these two possibilities, the first is likely to be preferred because the second would lead to higher and even aberrant costs since it equals to multiply the national tests by the number of possibilites of ages of the children within a cohort. Second, policy makers could act on the age at school entry of pupils by increasing the minimum age at which a pupil is eligible for school. It equals to changing the school entry laws. For example, by setting that pupils are eligible to grade 1 as long as they turned six by the 18^{th} August instead of the 31^{st} December, it will increase the mean age within a grade because those who are born between the 18^{th} August and the 31^{st} December would be constrained to wait till the next school year (then till being 8 months older). Hence, by being more « ready » for school than they would do if they entered earlier, children are likely to perform better in school. Third, since there is potentially a relative age causal effect, policy makers could influence on the distribution of ages within a grade such as age differentials are in their minimum. This can be done by regulating school classroom compositions aiming to the ideal age distribution within a classroom. For example, duplicating a classroom into two such as the first contains the youngest ones and the second contains the oldest ones (with an age threshold) is a possibility.

Nevertheless, these three suggestions are based on assumptions that there is at least a significant absolute age effect, age at school entry effect or relative age effect (recall that since the length of schooling within a cohort is rigorously the same, there can be no length of schooling effect in this study). Furthermore, one should have an idea in what extent these results are valid. In fact, the present study adressed in Reunion Island measures age effects on test scores in grade 5, for 2009-2010, 2010-2011 and 2011-2012 school years. There is *a priori* no guarantee that the causal effect of interest would be stable over the years (Aliprantis 2014) ; in addition that since 2009, a decade has passed until the present time. Consequently, this study is more informative, but have its interest because it should be a step to begin with, since as far as what my knowledge of the existing literature let me assert, this is the first study to adress a measure of causal relationship between the age and educational performance. What would be logical for future investigations is the concern of persistence of these age effects over time. In fact, a potential evidence of some significant persistence of early age effects in later educational and even adulthood outcomes would increase considerably the importance of having an idea of how and how much age differences lead to difference in outcomes in the early years of education of pupils in Reunion Island. These several considerations are reserved for future researches.

Appendix

Observed age at test versus Theoretical age at test

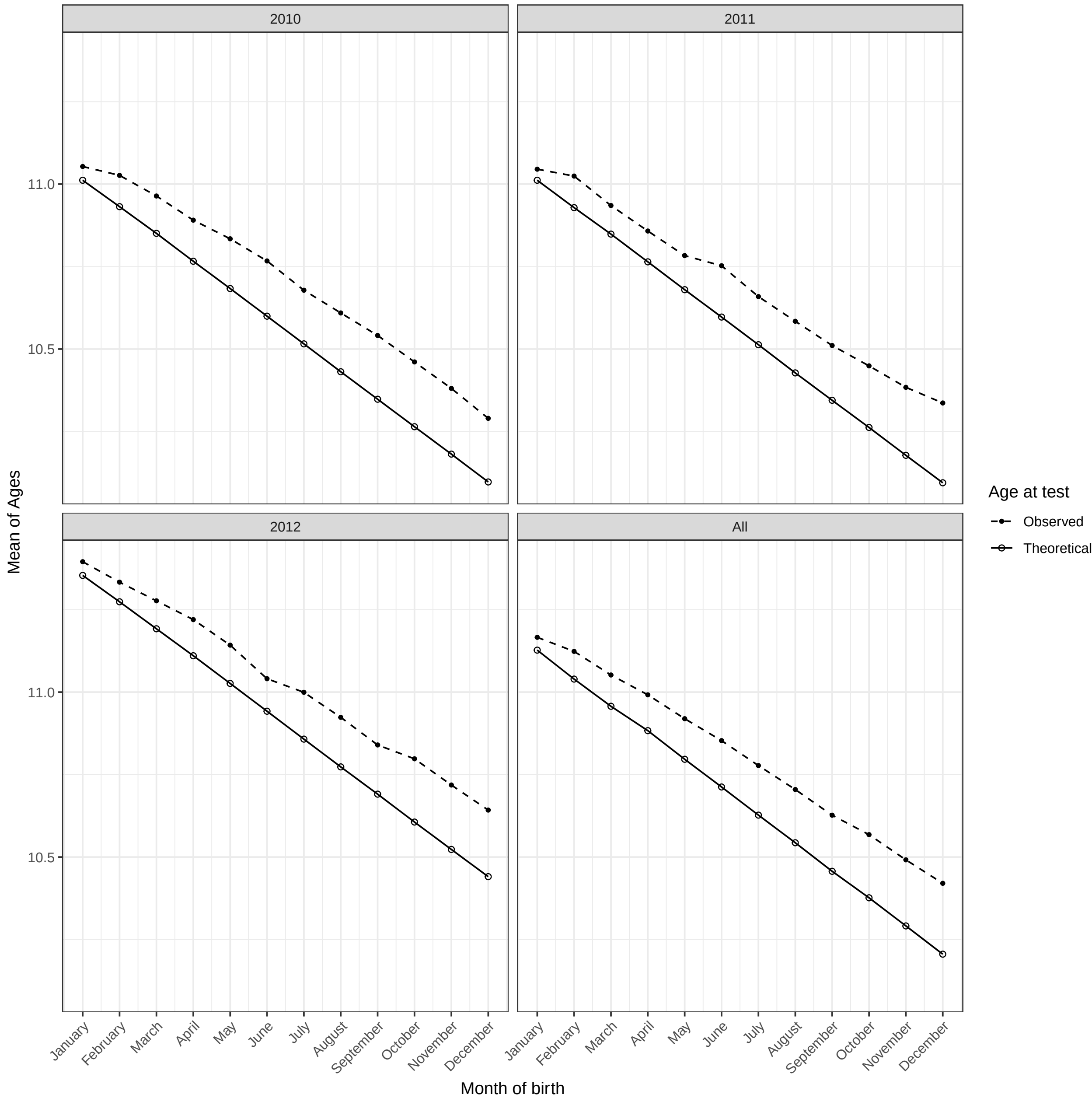

Figure 7: Observed age at test versus Theoretical age at test

Magnitudes of the parents socio-professional category subgroups estimates

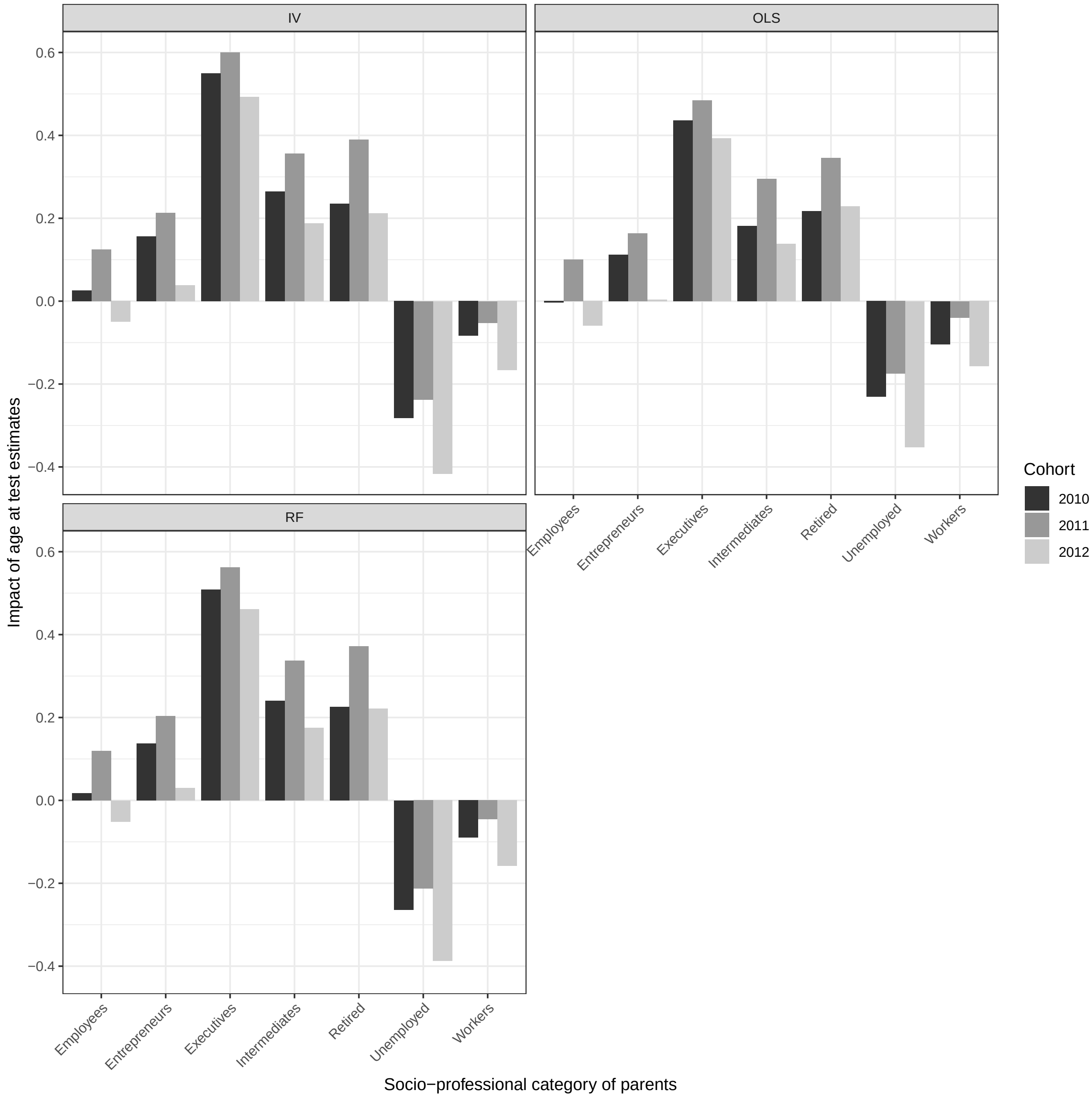

Figure 8: Estimates of the impact of parents socio-professional category on total test score

Detailed results of the endogeneity test and control function approach and of the first stage regressions

Table 13: Endogeneity test for age at test and control function approach

	Dependent variable : Total test score		
	2010	2011	2012
	(1)	(2)	(3)
Age at test	0.304	0.302	0.271
	(0.029)***	(0.028)***	(0.030)***
Sex - Male	−0.288	−0.289	−0.255
	(0.015)***	(0.013)***	(0.015)***
SPC - Entrepreneurs	0.157	0.213	0.039
	(0.057)***	(0.072)***	(0.066)
SPC - Executives	0.550	0.600	0.493
	(0.058)***	(0.068)***	(0.061)***
SPC - Intermediates	0.265	0.356	0.188
	(0.057)***	(0.069)***	(0.061)***
SPC - Employees	0.026	0.125	−0.050
	(0.052)	(0.065)*	(0.060)
SPC - Workers	−0.084	−0.052	−0.166
	(0.054)	(0.069)	(0.059)***
SPC - Retired	0.235	0.390	0.212
	(0.088)***	(0.098)***	(0.097)**
SPC - Unemployed	−0.282	−0.237	−0.417
	(0.052)***	(0.065)***	(0.059)***
SPC - Others	−0.430	−0.461	−0.360
	(0.054)***	(0.066)***	(0.121)***
$\widehat{\eta_{ict}}$	−1.121	−1.090	−1.019
	(0.034)***	(0.032)***	(0.035)***
N	13,561	14,622	10,790
R^2	0.236	0.256	0.201
Adjusted R^2	0.196	0.218	0.148

Notes: ***Significant at the 1 percent level.
**Significant at the 5 percent level.
*Significant at the 10 percent level.

Table 14: First stage regressions results

	Dependent variable : Age at test		
	2010	2011	2012
	(1)	(2)	(3)
Assigned relative age	0.845	0.806	0.843
	(0.012)***	(0.012)***	(0.012)***
Sex - Male	0.086	0.072	0.053
	(0.007)***	(0.006)***	(0.007)***
SPC - Entrepreneurs	−0.064	−0.034	−0.035
	(0.032)**	(0.029)	(0.027)
SPC - Executives	−0.135	−0.124	−0.115
	(0.032)***	(0.028)***	(0.028)***
SPC - Intermediates	−0.081	−0.065	−0.048
	(0.032)**	(0.027)**	(0.028)*
SPC - Employees	−0.029	−0.018	−0.006
	(0.031)	(0.027)	(0.026)
SPC - Workers	−0.021	0.022	0.031
	(0.032)	(0.029)	(0.027)
SPC - Retired	−0.030	−0.057	0.034
	(0.049)	(0.052)	(0.049)
SPC - Unemployed	0.060	0.082	0.108
	(0.031)**	(0.026)***	(0.026)***
SPC - Ohters	0.110	0.117	0.165
	(0.032)***	(0.028)***	(0.071)**
N	13,561	14,622	10,790
R^2	0.297	0.294	0.353
Adjusted R^2	0.260	0.258	0.310

Notes: ***Significant at the 1 percent level.
**Significant at the 5 percent level.
*Significant at the 10 percent level.

Subgroup estimation results

Table 15: Instrumental variable and reduced form estimates by sex - school year 2009 - 2010

	Dependent variable : Total test score			
	IV		RF	
	females	males	females	males
	(1)	(2)	(3)	(4)
Age at test	0.365	0.237		
	(0.048)***	(0.048)***		
Assigned relative age			0.308	0.203
			(0.037)***	(0.039)***
N	6,666	6,895	6,666	6,895
R^2	0.007	0.019	0.081	0.087
Adjusted R^2	−0.104	−0.086	−0.022	−0.011
Notes:		***Significant at the 1 percent level.		
		**Significant at the 5 percent level.		
		*Significant at the 10 percent level.		

Table 16: Instrumental variable and reduced form estimates by sex - school year 2010 - 2011

	Dependent variable : Total test score			
	IV		RF	
	females	males	females	males
	(1)	(2)	(3)	(4)
Age at test	0.355	0.228		
	(0.045)***	(0.046)***		
Assigned relative age			0.282	0.188
			(0.033)***	(0.036)***
N	7,325	7,297	7,325	7,297
R^2	0.019	0.057	0.101	0.126
Adjusted R^2	−0.085	−0.043	0.006	0.034
Notes:		***Significant at the 1 percent level.		
		**Significant at the 5 percent level.		
		*Significant at the 10 percent level.		

Table 17: Instrumental variable and reduced form estimates by sex - school year 2011 - 2012

	Dependent variable : Total test score			
	IV		RF	
	females	males	females	males
	(1)	(2)	(3)	(4)
Age at test	0.277	0.273		
	(0.050)***	(0.054)***		
Assigned relative age			0.235	0.232
			(0.041)***	(0.044)***
N	5,493	5,297	5,493	5,297
R^2	0.029	0.039	0.083	0.106
Adjusted R^2	−0.102	−0.098	−0.041	−0.021

Notes: ***Significant at the 1 percent level.
**Significant at the 5 percent level.
*Significant at the 10 percent level.

Table 18: Instrumental variable estimates by socio-professional category - school year 2009-2010

	Dependent variable : Total test score							
	Farmers	Entrepreneurs	Executives	Intermediates	Employees	Workers	Retired	Unemployed
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Age at test	0.948	0.110	0.283	0.090	0.286	0.402	2.869	0.356
	(0.383)**	(0.160)	(0.131)**	(0.159)	(0.084)***	(0.116)***	(4.318)	(0.060)***
Sex - male	−0.198	−0.340	−0.211	−0.101	−0.311	−0.239	−0.674	−0.291
	(0.317)	(0.093)***	(0.065)***	(0.074)	(0.047)***	(0.061)***	(0.613)	(0.027)***
N	198	710	1,040	842	2,053	1,523	114	4,776
R^2	0.0001	0.019	0.0003	0.003	0.00001	0.018	0.018	0.013
Adjusted R^2	−3.690	−1.095	−0.560	−1.001	−0.399	−0.559	−11.336	−0.142

Notes:
***Significant at the 1 percent level.
**Significant at the 5 percent level.
*Significant at the 10 percent level.

Table 19: Reduced form estimates by socio-professional category - school year 2009-2010

	Dependent variable : Total test score							
	Farmers	Entrepreneurs	Executives	Intermediates	Employees	Workers	Retired	Unemployed
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Age at test	0.952	0.100	0.227	0.077	0.269	0.311	0.835	0.287
	(0.338)***	(0.145)	(0.101)**	(0.135)	(0.075)***	(0.084)***	(0.542)	(0.044)***
Sex - male	−0.274	−0.330	−0.209	−0.098	−0.286	−0.212	−0.552	−0.258
	(0.283)	(0.090)***	(0.062)***	(0.074)	(0.044)***	(0.056)***	(0.242)**	(0.025)***
N	198	710	1,040	842	2,053	1,523	114	4,776
R^2	0.104	0.040	0.024	0.004	0.040	0.027	0.211	0.036
Adjusted R^2	−3.203	−1.051	−0.523	−0.999	−0.343	−0.544	−8.912	−0.115

Notes:
***Significant at the 1 percent level.
**Significant at the 5 percent level.
*Significant at the 10 percent level.

Table 20: Instrumental variable estimates by socio-professional category - school year 2010-2011

	Dependent variable : Total test score							
	Farmers	Entrepreneurs	Executives	Intermediates	Employees	Workers	Retired	Unemployed
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Age at test	−0.147	0.027	0.390	0.351	0.290	0.246	5.289	0.263
	(0.550)	(0.199)	(0.129)***	(0.133)***	(0.089)***	(0.161)	(7.512)	(0.054)***
Sex - male	−0.129	−0.320	−0.102	−0.228	−0.203	−0.320	−0.237	−0.264
	(0.201)	(0.122)***	(0.057)*	(0.073)***	(0.040)***	(0.068)***	(1.728)	(0.024)***
N	154	510	1,068	986	2,189	957	97	5,248
R^2	0.049	0.038	0.006	0.00000	0.006	0.001	0.061	0.005
Adjusted R^2	−3.156	−1.511	−0.588	−0.866	−0.392	−0.938	−5.439	−0.143

Notes: ***Significant at the 1 percent level.
**Significant at the 5 percent level.
*Significant at the 10 percent level.

Table 21: Reduced form estimates by socio-professional category - school year 2010-2011

	Dependent variable : Total test score							
	Farmers	Entrepreneurs	Executives	Intermediates	Employees	Workers	Retired	Unemployed
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Age at test	−0.097	0.021	0.293	0.305	0.241	0.192	1.302	0.206
	(0.369)	(0.152)	(0.095)***	(0.110)***	(0.071)***	(0.120)	(0.737)*	(0.041)***
Sex - male	−0.116	−0.317	−0.093	−0.207	−0.188	−0.303	−0.383	−0.246
	(0.185)	(0.118)***	(0.054)*	(0.070)***	(0.039)***	(0.066)***	(0.641)	(0.022)***
N	154	510	1,068	986	2,189	957	97	5,248
R^2	0.007	0.045	0.019	0.032	0.022	0.048	0.339	0.030
Adjusted R^2	−3.341	−1.492	−0.568	−0.806	−0.368	−0.847	−3.534	−0.113

Notes: ***Significant at the 1 percent level.
**Significant at the 5 percent level.
*Significant at the 10 percent level.

Table 22: Instrumental variable estimates by socio-professional category - school year 2011-2012

	Dependent variable : Total test score							
	Farmers	Entrepreneurs	Executives	Intermediates	Employees	Workers	Retired	Unemployed
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Age at test	0.940	0.424	0.181	0.306	0.288	0.235	2.446	0.217
	(0.566)	(0.178)**	(0.132)	(0.158)*	(0.099)***	(0.112)**	(5.977)	(0.052)***
Sex - male	−0.272	−0.291	−0.104	−0.260	−0.260	−0.351	−1.475	−0.277
	(0.228)	(0.098)***	(0.053)**	(0.073)***	(0.046)***	(0.060)***	(2.567)	(0.027)***
N	160	645	978	938	1,952	1,412	87	4,568
R^2	0.004	0.003	0.006	0.00002	0.001	0.003	0.038	0.0003
Adjusted R^2	−2.683	−1.334	−0.593	−0.881	−0.417	−0.604	−8.193	−0.163

Notes:
***Significant at the 1 percent level.
**Significant at the 5 percent level.
*Significant at the 10 percent level.

Table 23: Reduced form estimates by socio-professional category - school year 2011-2012

	Dependent variable : Total test score							
	Farmers	Entrepreneurs	Executives	Intermediates	Employees	Workers	Retired	Unemployed
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Age at test	0.807	0.387	0.148	0.273	0.240	0.203	0.778	0.181
	(0.453)*	(0.153)**	(0.106)	(0.134)**	(0.077)***	(0.093)**	(1.206)	(0.042)***
Sex - male	−0.306	−0.279	−0.103	−0.246	−0.245	−0.333	−0.361	−0.262
	(0.230)	(0.090)***	(0.052)**	(0.070)***	(0.044)***	(0.057)***	(0.608)	(0.026)***
N	160	645	978	938	1,952	1,412	87	4,568
R^2	0.063	0.049	0.008	0.033	0.031	0.044	0.092	0.030
Adjusted R^2	−2.463	−1.226	−0.588	−0.820	−0.374	−0.537	−7.679	−0.128

Notes:
***Significant at the 1 percent level.
**Significant at the 5 percent level.
*Significant at the 10 percent level.

References

Alet, Elodie, Liliane Bonnal, and Pascal Favard. 2013. "Repetition: Medicine for a Short-Run Remission." *Annals of Economics and Statistics/ANNALES D'ÉCONOMIE ET DE STATISTIQUE*. JSTOR, 227–50.

Aliprantis, Dionissi. 2014. "When Should Children Start School?" *Journal of Human Capital* 8 (4). University of Chicago Press Chicago, IL: 481–536.

Angrist, Joshua D, and Guido W Imbens. 1995. "Two-Stage Least Squares Estimation of Average Causal Effects in Models with Variable Treatment Intensity." *Journal of the American Statistical Association* 90 (430). Taylor & Francis Group: 431–42.

Angrist, Joshua D, Guido W Imbens, and Donald B Rubin. 1996. "Identification of Causal Effects Using Instrumental Variables." *Journal of the American Statistical Association* 91 (434). Taylor & Francis: 444–55.

Arellano, Manuel. 1987. "PRACTITIONERS'CORNER: Computing Robust Standard Errors for Within-Groups Estimators." *Oxford Bulletin of Economics and Statistics* 49 (4). Wiley Online Library: 431–34.

Bedard, Kelly, and Elizabeth Dhuey. 2006. "The Persistence of Early Childhood Maturity: International Evidence of Long-Run Age Effects." *The Quarterly Journal of Economics* 121 (4). MIT Press: 1437–72.

Bell, John F, and Sandra Daniels. 1990. "Are Summer-Born Children Disadvantaged? The Birthdate Effect in Education." *Oxford Review of Education* 16 (1). Taylor & Francis: 67–80.

Black, Sandra E, Paul J Devereux, and Kjell G Salvanes. 2011. "Too Young to Leave the Nest? The Effects of School Starting Age." *The Review of Economics and Statistics* 93 (2). MIT Press: 455–67.

Buckles, Kasey S, and Daniel M Hungerman. 2013. "Season of Birth and Later Outcomes: Old Questions, New Answers." *Review of Economics and Statistics* 95 (3). MIT Press: 711–24.

Cahan, Sorel, and Nora Cohen. 1989. "Age Versus Schooling Effects on Intelligence Development." *Child Development*. JSTOR, 1239–49.

Cahan, Sorel, and Daniel Davis. 1987. "A Between-Grade-Levels Approach to the Investigation of the Absolute Effects of Schooling on Achievement." *American Educational Research Journal* 24 (1). Sage Publications: 1–12.

Cascio, Elizabeth U, and Ethan G Lewis. 2006. "Schooling and the Armed Forces Qualifying Test Evidence from School-Entry Laws." *Journal of Human Resources* 41 (2). University of Wisconsin Press: 294–318.

Cascio, Elizabeth, and others. 2008. "How and Why Does Age at Kindergarten Entry Matter?" *FRBSF*

Economic Letter. Federal Reserve Bank of San Francisco.

Crawford, Claire, Lorraine Dearden, and Costas Meghir. 2007. "When You Are Born Matters: The Impact of Date of Birth on Child Cognitive Outcomes in England." Institute for Fiscal Studies.

Datar, Ashlesha. 2006. "Does Delaying Kindergarten Entrance Give Children a Head Start?" *Economics of Education Review* 25 (1). Elsevier: 43–62.

DeCos, Patricia L. 1997. *Readiness for Kindergarten: What Does It Mean?* California State Library, California Research Bureau.

Dhuey, Elizabeth, and Stephen Lipscomb. 2008. "What Makes a Leader? Relative Age and High School Leadership." *Economics of Education Review* 27 (2). Elsevier: 173–83.

Dhuey, Elizabeth, David Figlio, Krzysztof Karbownik, and Jeffrey Roth. 2017. "School Starting Age and Cognitive Development." National Bureau of Economic Research.

Dobkin, Carlos, and Fernando Ferreira. 2010. "Do School Entry Laws Affect Educational Attainment and Labor Market Outcomes?" *Economics of Education Review* 29 (1). Elsevier: 40–54.

Elder, Todd E, and Darren H Lubotsky. 2009. "Kindergarten Entrance Age and Children's Achievement Impacts of State Policies, Family Background, and Peers." *Journal of Human Resources* 44 (3). University of Wisconsin Press: 641–83.

Fertig, Michael, and Jochen Kluve. 2005. "The Effect of Age at School Entry on Educational Attainment in Germany."

Fletcher, Jason, and Taehoon Kim. 2016. "The Effects of Changes in Kindergarten Entry Age Policies on Educational Achievement." *Economics of Education Review* 50. Elsevier: 45–62.

Fredriksson, Peter, and Björn Öckert. 2006. "Is Early Learning Really More Productive." *The Effect of School Starting Age on School and Labor Market Performance, IZA WP* 1659.

———. 2013. "Life-Cycle Effects of Age at School Start." *The Economic Journal* 124 (579). Oxford University Press Oxford, UK: 977–1004.

Graue, M Elizabeth, and James DiPerna. 2000. "Redshirting and Early Retention: Who Gets the' Gift of Time' and What Are Its Outcomes?" *American Educational Research Journal* 37 (2). Sage Publications: 509–34.

Grenet, Julien. 2009. "Academic Performance, Educational Trajectories and the Persistence of Date of Birth

Effects. Evidence from France.” *Unpublished Manuscript*.

Hart, Robert A, and Mirko Moro. 2017. “Date of Birth and Selective Schooling.”

Hámori, Szilvia. 2007. “The Effect of School Starting Age on Academic Performance in Hungary.” Budapest Working Papers on the Labour Market.

Hámori, Szilvia, and János Köllő. 2012. “Whose Children Gain from Starting School Later?–evidence from Hungary.” *Educational Research and Evaluation* 18 (5). Taylor & Francis: 459–88.

Heckman, James J, and Richard Robb. 1986. “Alternative Methods for Solving the Problem of Selection Bias in Evaluating the Impact of Treatments on Outcomes.” In *Drawing Inferences from Self-Selected Samples*, 63–107. Springer.

Imbens, Guido W., and Joshua D. Angrist. 1994. “Identification and Estimation of Local Average Treatment Effects.” *Econometrica* 62 (2). [Wiley, Econometric Society]: 467–75. http://www.jstor.org/stable/2951620.

Jürges, Hendrik, and Kerstin Schneider. 2011. “Why Young Boys Stumble: Early Tracking, Age and Gender Bias in the German School System.” *German Economic Review* 12 (4). Wiley Online Library: 371–94.

Kaila, Martti. 2017. “The Effects of Relative School Starting Age on Educational Outcomes in Finland.”

Kawaguchi, Daiji. 2011. “Actual Age at School Entry, Educational Outcomes, and Earnings.” *Journal of the Japanese and International Economies* 25 (2). Elsevier: 64–80.

Langer, Philip, John Michael Kalk, and Donald T Searls. 1984. “Age of Admission and Trends in Achievement: A Comparison of Blacks and Caucasians.” *American Educational Research Journal* 21 (1). Sage Publications: 61–78.

Leuven, Edwin, Mikael Lindahl, Hessel Oosterbeek, and Dinand Webbink. 2004. “New evidence on the effect of time in school on early achievement.” HEW 0410001. University Library of Munich, Germany. https://ideas.repec.org/p/wpa/wuwphe/0410001.html.

Mayer, Susan E, David Knutson, and others. 1997. “Does Age at Enrollment in First Grade Affect Children’s Cognitive Test Scores.” Northwestern University/University of Chicago Joint Center for Poverty Research.

Mühlenweg, Andrea M, and Patrick A Puhani. 2010. “The Evolution of the School-Entry Age Effect in a School Tracking System.” *Journal of Human Resources* 45 (2). University of Wisconsin Press: 407–38.

Pellizzari, Michele, and Francesco C Billari. 2012. “The Younger, the Better? Age-Related Differences in

Academic Performance at University." *Journal of Population Economics* 25 (2). Springer: 697–739.

Peña, Pablo A. 2017. "Creating Winners and Losers: Date of Birth, Relative Age in School, and Outcomes in Childhood and Adulthood." *Economics of Education Review* 56. Elsevier: 152–76.

Piketty, Thomas, Mathieu Valdenaire, and others. 2006. *L'impact de La Taille Des Classes Sur La Réussite Scolaire Dans Les écoles, Collèges et Lycées Français: Estimations à Partir Du Panel Primaire 1997 et Du Panel Secondaire 1995.* Direction de l'évaluation et de la prospective.

Puhani, PA, and AM Weber. 2005. "Does the Early Bird Catch the Worm? Instrumental Variable Coefficients of Educational Effects of Age of School Enty in Germany." IZA Discussion Paper.

Robertson, Erin. 2011. "The Effects of Quarter of Birth on Academic Outcomes at the Elementary School Level." *Economics of Education Review* 30 (2). Elsevier: 300–311.

Rubin, Donald B. 1974. "Estimating Causal Effects of Treatments in Randomized and Nonrandomized Studies." *Journal of Educational Psychology* 66 (5). American Psychological Association: 688.

Schneeweis, Nicole, and Martina Zweimüller. 2014. "Early Tracking and the Misfortune of Being Young." *The Scandinavian Journal of Economics* 116 (2). Wiley Online Library: 394–428.

Staiger, Douglas, and H James. 1997. "Stock. 1997.'Instrumental Variables with Weak Instruments."' *Econometrica* 65 (3): 557–86.

Stipek, Deborah. 2002. "At What Age Should Children Enter Kindergarten? A Question for Policy Makers and Parents." *Social Policy Report* 16 (2): 1–20.

Strøm, Bjarne. 2004. "Student Achievement and Birthday Effects." Citeseer.

Suziedelyte, Agne, and Anna Zhu. 2015. "Does Early Schooling Narrow Outcome Gaps for Advantaged and Disadvantaged Children?" *Economics of Education Review* 45. Elsevier: 76–88.

White, Halbert, and others. 1980. "A Heteroskedasticity-Consistent Covariance Matrix Estimator and a Direct Test for Heteroskedasticity." *Econometrica* 48 (4). Princeton: 817–38.

Woessmann, Ludger. 2016. "The Importance of School Systems: Evidence from International Differences in Student Achievement." *Journal of Economic Perspectives* 30 (3): 3–32.

Wooldridge, Jeffrey M. 2013. *Introductory Econometrics : A Modern Approach.* Book; Book/Illustrated. 5th ed., international ed. Australia : South Western, Cengage Learning.