Working Papers

www.cesifo.org/wp

Grade Retention and Unobserved Heterogeneity

Robert J. Gary-Bobo
Marion Gousse
Jean-Marc Robin

CESIFO WORKING PAPER NO. 4846
CATEGORY 5: ECONOMICS OF EDUCATION
JUNE 2014

CESifo *Working Paper No. 4846*

Grade Retention and Unobserved Heterogeneity

Abstract

We study the treatment effect of grade retention, using a panel of French junior highschool students, taking unobserved heterogeneity and the endogeneity of grade repetitions into account. We specify a multi-stage model of human-capital accumulation with a finite number of types representing unobserved individual characteristics. Class-size and latent student-performance indices are assumed to follow finite mixtures of normal distributions. Grade retention may increase or decrease the student's knowledge capital in a type-dependent way. Our estimation results show that the Average Treatment effect on the Treated (ATT) of grade retention on test scores is small but positive at the end of grade 9. The ATT of grade retention is higher for the weakest students. We also show that class size is endogenous and tends to increase with unobserved student ability. The Average Treatment Effect (ATE) of grade retention is negative, again with the exception of the weakest group of students. Grade repetitions reduce the probability of access to grade 9 of all student types.

JEL-Code: M500.

Keywords: secondary education, grade retention, unobserved heterogeneity, finite mixtures of normal distributions, treatment effects, class-size effects.

Robert J. Gary-Bobo
CREST, ENSAE
15 boulevard Gabriel Péri
France – 92245 Malakoff cedex
robert.gary-bobo@ensae.fr

Marion Gousse
Sciences Po Paris
28 rue des Saints Pères
France – 75007 Paris
marion.gousse@gmail.com

Jean-Marc Robin
UCL London &
Sciences Po Paris
Paris / France
jeanmarc.robin@sciencespo.fr

16 April 2014
We thank Costas Meghir for useful comments. Robin gratefully acknowledges financial support from the Economic and Social Research Council through the ESRC Centre for Microdata Methods and Practice, grant RES-589-28-0001, and from the European Research Council (ERC), grant ERC-2010-AdG-269693-WASP. Robin and Gary-Bobo gratefully acknowledge support from the Agence Nationale de la Recherche (ANR), projet blanc "Econométrie des redoublements".

1 Introduction

Grade retention practices are common in the schools of some countries but absent from others. Some educational systems have been designed to play the role of public certification agencies. If this is the case, a student is promoted to the next grade only if her (his) test scores are sufficiently high, and the students who can't pass are tracked or retained. France and Germany are good instances of such systems, in which grade retention is familiar. In contrast, social promotion, that is, the practice of passing students to the next grade, regardless of school performance, seems to prevail in more egalitarian societies, or in countries promoting mass education. Scandinavian countries and the UK are good instances of the latter system. At the same time, grade retention is a form of second-best remedial education; in some countries it is the main if not the only form of remedial education, but it entails substantial costs. Grade repetitions consume resources, since they permanently increase the stock of enrolled students. There are opportunity costs, since grade repeaters could become productive sooner or have a longer productive life. There also exists substantial costs in the long run, since grade repeaters tend to obtain lower wages on the labor market, conditional on their highest credential[1]. Grade retention may also entail some benefits. The mere presence of grade repetitions acts as an incentive device and may increase study effort.[2] Finally, the distribution of skills in a given cohort of outgoing students may be improved if grade repeaters benefit from a longer period of schooling. Yet, many important aspects of a cost-benefit analysis are imperfectly known. As a consequence, in spite of its widespread use, it is hard to tell if grade retention dominates social promotion, or which of the two systems has the highest value as a social policy. As is well known, the question is hotly debated and international comparisons show trends in both directions. For instance, in the recent years, France has relied less often on grade repetitions, while in the US, grade retention has made a certain comeback, as an ingredient of school accountability policies.

The consequences of grade retention are not easy to estimate. This is essentially due to the endogenous character of the decision to hold a student back and to unobservable heterogeneity. Many studies in the past may have found a negative impact of grade retention on various outcomes because grade repeaters are a selected population with abilities below the average. In the sequel, we propose a way of evaluating the treatment effects of grade repetition in French junior high schools (grades 6 till 9), using a rich set of micro-data, and taking the endogeneity of retention decisions and class size into account. We do not observe the students' wages and focus on educational outcomes.[3]

In a preliminary study of the data, we find that the local average treatment effect (*i.e.*, the LATE[4]) of grade retention on value-added, defined here as the difference between grade-9 and grade-6 scores, is significant and positive, using the quarter of birth as an instrument for retention. But the result doesn't seem to be very robust. We know that when treatment

[1] On this question, see Brodaty *et al.* (2012)

[2] On study effort, see De Fraja *et al.* (2010).

[3] For a study of the impact of grade retention on wages, using French data, see Brodaty *et al.* (2012).

[4] On this concept, see Imbens and Angrist (1994)

effects are heterogeneous, the linear Instrumental Variable (IV) estimator is a weighted average of marginal treatment effects (see the work of Heckman and Vytlacil (2005); see also Heckman (2010)). It follows that the IV estimates obtained with a particular instrument may not correctly identify the relevant effects. Indeed, in the following, we show that the treatment effect of grade repetition varies with unobserved characteristics of students, being positive for some individuals and negative for others.

Taking our inspiration from the work of Heckman and his co-authors, we propose a tractable model in which treatment effects are heterogeneous (see, *e.g.*, Carneiro, Hansen, Heckman (2003)). We assume the existence of a finite number of *latent student types* and that the effects of retention may vary from one type of individual to the next. Our approach is parametric: the observed outcomes and the latent variables, such as unobserved test scores, are modeled as finite mixtures of normal distributions. The model can then be used to compute counterfactuals and treatment effects.

We take dynamics into account, exploiting the data's panel structure. Our approach is similar in spirit to that of Cunha and Heckman (2007, 2008) and Cunha, Heckman and Schennach (2010), but different (and somewhat simpler) in a number of technical details. The educational outcomes of the same individuals are observed recursively through time, either completely (quantitative test scores) or partially (qualitative promotion decisions). The successive observations are used to identify the model parameters and the latent student types. In particular, the coefficients of student types, that is, their impact on the different outcomes, are identified under a limited set of reasonable assumptions.

To be more precise, we specify a structural model of knowledge-capital accumulation in junior high school. The model explains grade retention, class size, promotion decisions and test scores. It is estimated using panel data, on scores in grades 6 and 9, information on class sizes and on student transitions (promotion to next grade, retention and redirection towards vocational education). The panel provides a rich set of control variables describing family background and the environment of students. Repeated grades contribute to the accumulation (or destruction) of human capital (or skills) in a specific and type-dependent way. We present estimation results for a variant of our model with four unobserved student types or *groups*. Groups are clearly distinct and a clear hierarchy appears in terms of student ability. Groups are ranked in the same way if we use test scores in Math, in French, at the beginning of grade 6 or at the end of grade 9. The ranking of groups explains a similar ranking in the students' probabilities of grade retention (or promotion to the next grade). In a parallel fashion, the weaker the group, the smaller the class-size, in every grade. This result shows the endogeneity of class-size, which is used as a remediation instrument. Finally, to assess the impact of grade repetition on test scores at the end of grade 9, we compute the ATT and the ATE of the grade-repetition treatment. To this end, with the help of the model, we compute the counterfactual class-size and test scores of grade repeaters (resp. non-repeaters) that would be observed if they had not repeated a grade (resp. if they had repeated a grade), averaging over students and all possible types of each student, using their posterior probabilities of belonging to a group. We find that the ATE is negative, while the

ATT is positive, but small and barely significant. The ATE and ATT are also computed within each of the four groups separately. This confirms that treatment effects are heterogeneous: grade retention is detrimental to able students but has some positive effects on the weakest students' final test scores. It is also shown that grade repetition has a negative impact on the student's probabilities of access to grade 9. We conclude that grade retention should be replaced by some other form of remediation.

There is a substantial literature on grade retention, but many early contributions did not address endogeneity or selection problems in a convincing way (see, *e.g.*, Holmes and Matthews (1984), Holmes (1990)). Few contributions have managed to propose a causal econometric evaluation of grade retention. An early attempt, providing IV estimates on US High-School data is due to Eide and Showalter (2001). Also in the US, Jacob and Lefgren (2004, 2009) use regression discontinuity methods to evaluate grade repetitions in the Chicago Public-Sector Schools. Jacob and Lefgren (2004) find some positive short-term effects of grade retention on test scores for primary school children. Neal and Whitmore-Shanzenbach (2010) also propose an evaluation of the 1996 reforms that ended social promotion in Chicago Public Schools. Ying Ying Dong (2010) studies grade retention in Kindergarten and finds positive effects. The same data is used by Fruewirth-Cooley, Navarro and Takahashi (2010) to estimate a multi-period structural model in which the treatment effect of retention depends on the year of application. They also find positive effects. Recently, Baert, Cockx and Picchio (2013) used a structural dynamic choice model, estimated with Belgian data, and found that grade retention has a positive impact on the next evaluation, and persistent effects. On Latin American countries see, *e.g.*, Gomes-Neto and Hanushek (1984). Manacorda (2009) applies a regression discontinuity approach to Uruguayan junior high-school data and finds negative effects on the dropout rate. In France, contributions on this topic (with a causal approach) are due to Mahjoub (2007), Alet (2010), d'Haultfoeuille (2010), Brodaty *et al.* (2012) and Alet, Bonnal and Favard (2013). Among these authors, d'Haultfoeuille (2010) applies a new non-parametric method for the estimation of treatment effects to French primary education data and also finds positive effects. Finally, Brodaty *et al.* (2012) find negative signaling effects of grade retention on wages. None of the quoted papers use the methods and the data employed in the present article.

In the following, Section 2 describes the data. Section 3 presents a preliminary analysis of grade retention using linear IV methods. Section 4 presents our multi-stage skill accumulation model. The estimation strategy is exposed in Section 5. Section 6 and 7 present the estimation results and the average treatment effects. Concluding remarks are in Section 8.

2 Data

The data set used in this study is the 1995 secondary education panel of the French Ministry of Education (DEPP[5] Panel 1995), which follows 17,830 students in junior high-school (*i.e., collège*) from grade 6 to grade 9 (grade 6 is the equivalent of the French *classe de sixième*) during the

[5] Département de l'Evaluation, de la Prospective et de la Performance

Table 1: Individual Grade Histories

Grade History	Count	
1234	9403	71,58%
12334	732	
12234	910	
11234	684	
Subtotal	2326	17,71%
1233V	33	
1223V	114	
1123V	154	
123V	147	
122V	146	
112V	246	
11V	7	
12V	560	
Subtotal	1407	10,71%
Total	13136	

years 1995-2001. The principals of a sample of junior high-schools were asked to collect data on all pupils born on the 17th day of each month, with the exception of March, July, and October, and entering grade 6 in September 1995 — about 1/40th of the whole cohort. A recruitment survey was conducted at the beginning of the first school year (1995-96). Then, a number of follow-up questionnaires were filled by the principals in every subsequent year until 2001, and a questionnaire was filled by the families in 1998 (with a response rate of 80%). Each student's junior high-school history was recorded without interruption, even when the student moved to another school. For each pupil and each year, we know the attended grade (6 to 9), the size of the class, and the promotion decision made by the teachers at the end of the year. In fact there are three possible decisions: promotion to next grade, grade retention or redirection to vocational education (*i.e.* "steering"). These transition decisions are made during the last staff meeting (*i.e.*, the *conseil de classe*), at the end of every school year, on the basis of test scores and other more or less objective assessments of the pupil's ability and potential in the next grade. Test scores in Mathematics and French are available at the beginning of grade 6 and at the end of grade 9. Grade 9 test scores are missing for the individuals who dropped out of general education for apprenticeship or vocational training, and therefore never reached grade 9 in the general (non vocational) middle schools. In addition, matching these data with another source from the Ministry of Education, the *Base Scolarité*, we obtain further information on school characteristics. In particular, total school enrollment and total grade enrollment (in each grade) for each year during the 1995-2001 period. These data will allow us to compute instruments for class-size, based on local variations of enrollment. There are some missing data, but the quality

Table 2: Students Promoted, Retained or Redirected in Each Grade and Year

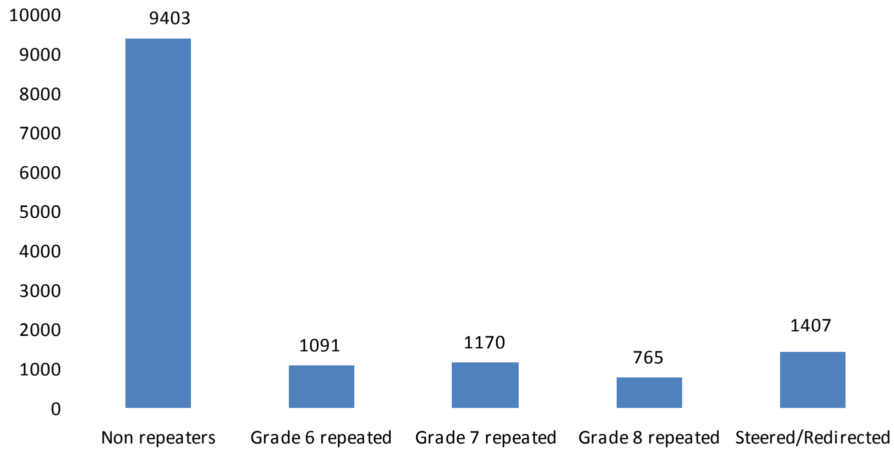
Year t	Grade	Initial Stock	Promoted P	Retained R	Redirected V
$t = 1$	Grade 6	13136	12045	1091	0
$t = 2$	Grade 6	1091	1084	0	7
	Grade 7	12045	10315	1170	560
$t = 3$	Grade 7	2254	1862	0	392
	Grade 8	10315	9403	765	147
$t = 4$	Grade 8	2627	2326	0	301
	Grade 9	9403			

of the panel is very good. For example, initial test scores are known for 95% of the sampled individuals. Discarding observations with obvious coding errors and missing data, and slightly more than 450 histories of pupils registered in special education programs (for mentally retarded children), we finally ended up with a sample of more than 13,000 individuals: 9,403 of them are in grade 9 in 1999, 2,594 are in grade 8 and 250 in grade 7. The last subset contains the few individuals who repeated a grade twice. We chose to discard these observations to reduce the number of cases. The final sample has 13,136 Students, which amounts to almost 75% of the individuals in the initial survey.

In the following, grades are denoted by g, and $g \in \{1, 2, 3, 4\}$, where $g = 1$ corresponds to grade 6, and so on. The year is denoted t with $t \in \{1, 2, 3, 4, 5\}$, where $t = 1$ corresponds to year 1995, etc. Individuals are indexed by i. Let g_{it} denote the grade of individual i in year t. With this notation system, a student i who doesn't repeat any grade is such that $g_{it} = t$. A grade repeater is such that $g_{it} = t - 1$. Table 1 gives the observed distribution of grade histories (in junior high school). Each row corresponds to a different type of trajectory. Letter V stands for vocational education. For example, the sequence 11234 means that grade 6 was repeated and therefore, that the student is observed in grade $g = 4$ in year $t = 5$. The sequence $123V$ indicates that the student was steered towards vocational education after grade 8. In total, about 30 % of the pupils do not complete junior high-school in four years: 18% are retained in one grade, 11 % are redirected.

Individual histories are described by Table 2 and on Fig. 1. Table 2 presents two rows per year, except in year $t = 1$. During the first year, all students are in grade 6. Out of the 13136 students initially enrolled in grade $g = 1$, 12045 are promoted, and 1091 are retained. In year $t = 2$, we see that 1084 repeaters in grade $g = 1$ are promoted and only 7 students have been redirected. In year $t = 3$ there are $2254 = 1170 + 1084$ students in grade 7; 1170 students repeating grade 7 and 1084 students that were in grade 6 the year before, etc. Figure 1 shows that the 9403 non-repeaters constitute a majority of more than 70% of the students. Repeaters amount to less than 9% of the latter cohort each year.

Figure 1: Number of Repeaters in each Grade



3 Preliminary Analysis : IV Estimates

We start our study of the causal effect of grade retention on educational achievement, using the student's quarter of birth as an instrument for grade retention, in a linear model. The quarter or the month of birth has been used by various authors as an instrument (see, *e.g.*, Angrist and Krueger (1991)). Recent work has shown that the month of birth can have long-lasting effects (see, *e.g.*, Bedard and Dhuey (2006), Grenet (2010)). In his dissertation, and a recent paper, Mahjoub (2007, 2009), used the quarter of birth as an instrument for grade retention. This approach yields a positive impact of grade retention on value-added scores, defined as the difference between standardized grade-6 and grade-9 scores, in Mathematics and in French. We follow the same approach here, as a preliminary step.

Value added is higher for repeaters than for non-repeaters. This is true both in French and Mathematics. There exists a strong link between the age of a child, as measured by the month of birth, or quarter of birth, and the probability of grade repetition (for details, see the Appendix). The probability of grade retention is clearly higher for children born later in the year. In principle, children must be 6 years old on September 1rst of year t to be admitted in primary school, grade 1, year t. First-quarter students tend to be relatively older in their class, with an age difference that can reach 11 months, and relatively older children tend to perform better. At the same time, teachers are reluctant to retain older children in a grade, as retention may change a difference —being older— into a stigma —being too old.

It follows that the month, quarter or season of birth is a candidate instrument for the grade-retention treatment, because it has good chances of being independent of the error term in an outcome equation with many controls. Note, in addition, as emphasized by Mahjoub (2007), that the value-added outcome being the difference of two test scores, possible specific and persistent effects of the birth quarter are "differenced out".

We now estimate the effect of grade retention on value added by 2SLS, using the quarter of birth as an instrument for grade retention. Some descriptive statistics on value-added, as well as further details on this IV approach are relegated in the Appendix. Scores are standardized to

Table 3: Grade Retention Probability

Variables	Grade retention
First quarter	-0.0513***
	(0.0110)
Second quarter	-0.0459***
	(0.00991)
Third quarter	-0.0133
	(0.0109)
R^2	0.054
F statistic for instruments	31.74

Estimated by OLS. The dependent variable is the grade-retention dummy here. The following list of control variables were included in the regressions: gender; number of siblings; birth order (rank among siblings); parental occupation; parental education; indicator of grade repetition in primary school; total school enrollment. Standard errors are in parentheses; ***, ** and * indicate significance at the levels of 1, 5, and 10%, resp.

have a mean of 50 and a standard deviation of 10 in grade 6 and in the whole sample (including all redirected pupils). Scores in grade 9 are standardized in the same way, using the sub-sample of individuals who reached grade 9. The first-stage is a linear regression of the grade-retention dummy on birth quarter dummies and controls (the linear probability model). Results are displayed in Table 3. The fourth quarter being the reference in the regressions, we see that relatively older students have a significantly lower probability of being held back.

Table 4 presents OLS and 2SLS estimates of the effect of grade retention on value-added scores using the same set of controls. Instrumenting grade retention by the quarter of birth has dramatic effects on the sign and the size of the effect. Grade retention increases the score by about twice the standard deviation of value-added. These results confirm that the retention decision is endogenous.

Now, trying to estimate the impact of grade repetition in variants of this model, we found that the 2SLS results of Table 4 were not very robust, being very sensitive to the set of controls introduced in the equation of interest. But it is well known that IV estimates can be difficult to interpret when treatment effects vary with unobservable characteristics of individuals. To see this, let R denote the grade retention indicator. The outcome variable Y is value added. Let Y_1, Y_0 denote counterfactual outcomes for grade repeaters and non-repeaters. Let Z denote a dummy variable indicating whether the student was born in the first half of the year or in the second half. $Z = 1$ thus points at relatively older children. Let R_1, R_0 denote the counterfactual grade retention dummies, conditional on the instrument Z being 1 or 0. Under monotonicity, *i.e.*, $\mathbb{P}(R_1 \leq R_0) = 1$, the IV estimator converges to the Local Average Treatment Effect :

$$LATE = \mathbb{E}(Y_1 - Y_0 | R_1 - R_0 = -1)$$

Table 4: OLS and IV Estimates of Grade-Retention Effects

	OLS		2SLS	
Dependent Variable	Math VA	French VA	Math VA	French VA
Grade repetition	1.757***	1.899***	21.94***	14.79***
	(0.200)	(0.196)	(5.391)	(4.510)
R^2	0.035	0.043		

The table reports the estimated coefficient of the retention dummy in different regressions. Gender is included as a control in all regressions in addition to number of siblings; birth order (rank among siblings); parental occupation; parental education; indicator of grade repetition in primary school; total school enrollment. Standard errors are in parentheses; ***, ** and * indicate significance at the levels of 1, 5, and 10%, resp.

This measures the average value-added score for the individuals whose retention in a grade would have been avoided, had they been born at the beginning of the year instead of at the end. The question is whether these marginal individuals are representative of the whole sample or not. To help answering this question, suppose that counterfactual scores follow a generalized Roy model[6]:

$$\begin{aligned} Y = Y_1 = m_1 + U_1 \quad & \textit{if} \quad Y_1 - Y_0 > c(Z) + V \\ = Y_0 = m_0 + U_0 \quad & \textit{if} \quad Y_1 - Y_0 \leq c(Z) + V, \end{aligned}$$

where $c(Z)$ is an increasing function of Z, interpreted as a cost. Assume that U_1, U_0 and V are independent given Z and that they are normally distributed. It is easy to show that

$$LATE = m_1 - m_0 + \frac{Var(U_1 - U_0)}{Var(U_1 - U_0) + Var(V)} \left(\frac{\phi(d_0) - \phi(d_1)}{\Phi(d_1) - \Phi(d_0)} \right)$$

with

$$d_z = c(z) - (m_1 - m_0) \ , \ z = 1, 0,$$

where ϕ is the normal density and Φ is the normal c.d.f. The cost of grade retention is higher for older individuals, so $c(1) > c(0)$, hence, $d_1 > d_0$. Let us assume, for the sake of the argument, that $c(1) > m_1 - m_0 > c(0)$, so $d_1 > 0 > d_0$. It is clear in this case that the $LATE$ may be positive or negative, without this telling us anything certain about the sign of $m_1 - m_0$. The $LATE$ being a marginal effect, it may predominantly reflect cost parameters and may not be informative about treatment effects. This is why, in the next section, we design a structural model to uncover the mechanisms of grade repetition and their impact on educational attainment.

[6] see Heckman and Vytlacil (2005).

4 A Model of Knowledge-Capital Accumulation

We construct a model of knowledge capital accumulation with unobserved heterogeneity. We found a source of inspiration in a series of influential papers by James Heckman and his coauthors, in which heterogeneity is captured by means of dynamic factor models. See, *e.g.*, Cunha and Heckman (2008) and Cunha, Heckman and Schennach (2010). Although close in spirit, the present approach relies on a somewhat simpler model. We use a multi-period setting. We rely on the idea that, in the educational process, inputs are imperfectly observed and outputs are imperfectly measured, by means of test scores and teacher's decisions. Unobserved heterogeneity is modeled by means of a discrete set of unobserved individual types, generating finite mixtures of normal distributions.

The model is designed to match the following data features. We observe test scores, in French and Mathematics, but only at the beginning of grade 6 and at the end of grade 9. Promotion decisions (promotion to the next grade, grade retention or redirection to vocational training) are observed in all years. In addition to these test scores and transitions, we also observe class size and total school enrollment. The students who do not drop off into vocational education at some point reach the terminal grade after four or five years, depending on retention, during the period 1995-2000. For children who never repeat a grade, we have observations in years $t = 1, 2, 3, 4$. For those who repeat a grade once and are not redirected to a vocational track, t can take all five values 1, 2, 3, 4, 5. Redirected children are the cause of attrition. Pupils are indexed by $i = 1, \dots, N$. Let $g_{it} \in \{1, 2, 3, 4\}$ denote the grade of student i in year t, and let $S_{it} \in \{P, R, V\}$ denote the promotion decision (*i.e.*, promotion, retention and redirection) at the last staff meeting of year t. $g_{i,t+1}$ is missing if $S_{it} = V$. All students start in grade 6 in year 1 ($g_{i1} = 1$), so we set $S_{i0} = P$ for all i. There is no redirection of children towards vocational education in grade 6, so $S_{i1} \in \{P, R\}$.

4.1 Initial conditions

Initial scores in Mathematics and French measure initial knowledge-capital in Mathematics and in French, denoted h_{m0} and h_{f0} respectively. We assume that individuals have four possible unobservable types, or equivalently, belong to one of four possible *groups*. Let G_{ik} denote the dummy which is equal to 1 if i belongs to group k and equal to 0 otherwise. Let p_k denote the unconditional probability of belonging to group k and, of course, $p_1 + p_2 + p_3 + p_4 = 1$. Knowledge-capital levels, at the beginning of grade 6, *i.e.*, h_{m0} and h_{f0}, have the following form:

$$h_{mi0} = c_{m01} + c_{m02}G_{i2} + c_{m03}G_{i3} + c_{m04}G_{i4}, \tag{1}$$

$$h_{fi0} = c_{f01} + c_{f02}G_{i2} + c_{f03}G_{i3} + c_{f04}G_{i4}. \tag{2}$$

In this formulation, Group 1 is the reference group. It follows that c_{m01} and c_{f01} are the average initial levels of knowledge-capital in Mathematics, and French, respectively, for Group 1 individuals. Subscript m, resp. f, indicates a coefficient related to the initial Mathematics capital, resp., the French language-capital equation. The average initial Mathematics-capital of Group k is thus $c_{m01} + c_{m0k}$, for $k = 2, 3, 4$, etc.

Human capital is therefore discrete, but this should not be taken literally. We could add a random term with a continuous distribution, representing other unobserved inputs to the expressions of h_{mi0} and h_{fi0}, but the distribution of this term would not be identifiable, because it could not be distinguished from the teachers' "grading error", defined below. At this stage, we could also have added a list of controls, including indicators of family-background characteristics, but we omitted them, mainly to limit the number of parameters to be estimated. It follows that the groups may capture some of the effects of family background. Family-background variables and other controls will later be used to explain the probability of belonging to a given group, in separate regressions. We suppose that the test scores in French, denoted y_f, and in Math, denoted y_m, at the beginning of grade 6, are two different measures of the same knowledge-capital, that is,

$$y_{mi} = h_{mi0} + \varepsilon_{mi0} \tag{3}$$

$$y_{fi} = h_{fi0} + \varepsilon_{fi0}. \tag{4}$$

where ε_{m0}, ε_{f0} are random variables with a normal distribution and a zero mean, representing "grading" errors. The latter regression functions will identify the variance of ε_{m0} and ε_{f0}.

During the schooling of each student, we observe different variables that we regroup in different categories. There are time-invariant characteristics of the individual, such as family background observations, denoted X_0; time-varying characteristics of the individual denoted X_t, $t = 1, \ldots, 5$ and time-varying characteristics of the school, used as instruments for class size, denoted Z_t. The variables used in regressions are listed in Table 5.

The instrument for class size exploits discontinuities induced by the application of a class-opening threshold, as in Angrist and Lavy (1999) and Hoxby (2000). Let N_{it} denote total grade enrollment in i's school in year t. The *theoretical class size* in year t, denoted Z_{it}, is the class size that would obtain if the headmaster's rule was to open a new class, as soon as total grade enrollment in grade g_{it} became greater than τq and to minimize class-size differences, where τ is the class-opening threshold and q is an integer. Given these definitions, the theoretical *number* of classes in grade g_{it}, denoted κ_{it}, is by definition,

$$\kappa_{it} = int\left[\frac{N_{it} - 1}{\tau}\right] + 1,$$

where $int[x]$ is the largest integer q such that $q \leq x$. The theoretical number of students per class in grade g_{it} is simply

$$Z_{it} = \frac{N_{it}}{\kappa_{it}}.$$

Piketty and Valdenaire (2006), Gary-Bobo and Mahjoub (2013) show how this function of total grade enrollment fits the observed data in the French Educational system. We set the threshold value $\tau = 25$ because it seems to provide the best fit with Panel 1995. We will see below that Z_{it} has a strong effect in class-size regressions.

Table 5: Sets of Variables

Time-invariant characteristics X_0	Time-varying characteristics X_1,X_2,X_3,X_4,X_5	Time-varying instruments Z_1,Z_2,Z_3,Z_4,Z_5
Gender.	Foreign language studied.	Theoretical class size (*i.e.*, Maimonides' Rule)
Father's occupation.	Special education zone.	
Mother's education.	Number of foreigners in school.	
Number of Siblings.	Class size.	
Grade retention in primary school.	Total school enrollment.	
Private sector in primary school.	Size of the urban area.	
	Private sector.	

4.2 Knowledge-capital accumulation

Knowledge, or human capital, accumulates according to the following equation:

$$h_{i1} = a_1 n_{i1} + b_1 X_{i1} + c_{11} + c_{12} G_{i2} + c_{13} G_{i3} + c_{14} G_{i4}, \tag{5}$$

where n_{i1} denotes class size in individual i's class, grade $g_{i1} = 1$. Again, in equation (5), Group 1 is the reference, so that c_{11} is the impact of Group 1 on h_{i1}, and the impact of group k is $c_{11} + c_{1k}$ for all $k > 1$.

Many studies have established that class size is an endogenous variable. In particular, available evidence for France shows that class size is positively correlated with student performance because smaller classes are typically used to redistribute resources in favor of weaker students, or in favor of schools located in areas targeted for special help in education (see Piketty and Valdenaire (2006), Gary-Bobo and Mahjoub (2013)). We therefore model class size n_{i1} separately, as follows. Using Group 1 as the reference, we have,

$$n_{i1} = \alpha_{11} X_{i1} + \alpha_{12} Z_{i1} + \beta_{11} + \beta_{12} G_{i2} + \beta_{13} G_{i3} + \beta_{14} G_{i4} + \zeta_{i1} \tag{6}$$

The random term ζ_{i1} is an independent, normally distributed error.

Since we do not have any quantitative measure of performance at the end of grades $g \in \{1, 2, 3\}$, repeated or not, we define a single, latent education score for those years. In grade 6, *i.e.*, if $g_{it} = 1$, we define the latent variable,

$$y_{i1} = h_{i1} + \varepsilon_{i1} \tag{7}$$

where ε_1 is an independent normal error with a zero mean.

An individual is promoted to grade 7, *i.e.*, $g_{i,2} = 2$, if his(her) human capital is high enough, and repeats a grade otherwise. The promotion decision is modeled as a simple Probit. Let C_{11} be a human-capital threshold above which students are promoted. We have,

$$S_{i1} = \begin{cases} P & \textit{if} \quad y_{1i} \geq C_{11} \\ R & \textit{if} \quad y_{1i} < C_{11}. \end{cases} \tag{8}$$

The distribution of ε_1 is assumed to be standard normal, as usual in such a case, to identify the coefficients of the latent index. Given our specification of h_1 given by (5) above, we see that the model will only identify the constant

$$\delta_{11} = C_{11} - c_{11}.$$

This is of course technically equivalent to normalizing C_{11}, but, in principle, C_{11} is the human-capital level above which students pass, while c_{11} is the specific mean level reached by Group 1 students in the hypothetical situation $n_1 = X_1 = 0$. In essence, our model identifies differences between groups, not the absolute mean level of a group.

4.3 From second to fifth year

Similarly, still using Group 1 as the reference, in the second and third years, the human capital has the following representation.

If $g_{it} = t$ (non-repeaters),

$$h_{it} = a_t n_{it} + b_t X_{it} + c_{t1} + c_{t2} G_{i2} + c_{t3} G_{i3} + c_{t4} G_{i4}. \tag{9}$$

If $g_{it} < t$ (repeaters), we have

$$h_{it} = a_{tr} n_{it} + b_{tr} X_{it} + c_{t1r} + c_{t2r} G_{i2} + c_{t3r} G_{i3} + c_{t4r} G_{i4}. \tag{10}$$

The class-size equations are specified as follows.

If $g_{it} = t$ (non-repeaters), we have,

$$n_{it} = \alpha_{t1} X_{it} + \alpha_{t2} Z_{it} + \beta_{t1} + \beta_{t2} G_{i2} + \beta_{t3} G_{i3} + \beta_{t4} G_{i4} + \zeta_{it}, \tag{11}$$

where ζ_{it} is an independent normal random variable.

If $g_{it} < t$ (repeaters), we have,

$$n_{it} = \alpha_{t1r} X_{ti} + \alpha_{t2r} Z_{it} + \beta_{t1r} + \beta_{t1r} G_{i2} + \beta_{t3r} G_{i3} + \beta_{t4r} G_{i4} + \zeta_{itr}. \tag{12}$$

where ζ_{itr} is an independent normal random variable.

At the end of the second and third years, if the student has not repeated a grade before, he or she can either pass to the next grade (P), repeat the year (R) or be redirected towards a vocational track (V). We model these three different transitions with an Ordered Probit. Promotion or retention decisions are made by the teachers' staff meetings (*i.e.*, the *conseils de classe*), at the end of every school year. In essence, these staff meetings base decisions on the student's grade-point average (hereafter GPA) at the end of the year, and decide wether to promote, to hold back, or to "steer" the student towards vocational education. Students with a GPA above a certain threshold are promoted; students with a low record are "steered"; students with a mediocre, below-the-average record repeat the grade, if the teachers' committee thinks that they can benefit from the repetition. It seems reasonable to assume that the promotion decision is based on some average of the teachers' assessments of the student's cognitive capital, plus an unobserved individual effect, reflecting other unobservable factors that the members of the teaching staff take into consideration. We have in mind that the student's unobservable GPA in year t is highly correlated with the latent capital h_{it}, or to fix ideas, that h_{it} is the GPA in year t plus some random factor. We then model the unobservable capital h_{it} as an educational output, which is the result of some educational inputs: class-size, time-varying variables, and individual ability, as captured by the group indicator G_{ik}. Given this, and given the clear hierarchy of the three possible decisions, it seems reasonable to use an Ordered Probit structure.

Define first the latent variable

$$y_{it} = h_{it} + \varepsilon_{it},$$

where ε_t is an independent normal error. The decision S_{it} is then specified as follows,

$$S_{it} = \begin{cases} V & \text{if} \quad y_{it} < C_t \\ R & \text{if} \quad C_t \leq y_{it} < D_t \\ P & \text{if} \quad y_{it} \geq D_t, \end{cases} \tag{13}$$

where C_t and D_t are the Probit cuts. We assume that ε_t has a standard normal distribution. As above, the model in fact identifies only the differences,

$$\delta_{t1} = C_t - c_{t1}, \quad \text{and} \quad \delta_{t2} = D_t - c_{t1}.$$

In the sample, a student never repeats a grade twice. Thus, the model embodies the fact that, if the student has already repeated a grade, he or she cannot repeat a second time. For repeaters, the possible decisions are: promotion to the next grade or redirection. We model the two different transitions with a simple Probit. We first define the latent variable,

$$y_{itr} = h_{it} + \varepsilon_{itr},$$

where ε_{tr} is an independent normal error. The decision S_{itr} is then specified as follows,

$$S_{itr} = \begin{cases} P & \text{if} \quad y_{itr} \geq C_{tr} \\ V & \text{if} \quad y_{itr} < C_{tr}, \end{cases} \tag{14}$$

where C_{tr} is a threshold, and we assume that ε_{tr} has a standard normal distribution. The model identifies only the difference, $\delta_{tr} = C_{tr} - c_{t1r}$.

It follows from these assumptions that the latent human capital h_{it} is affected by the promotion and retention decisions, because all the coefficients are free to vary in expressions (9) and (10), as well as in the auxiliary class-size equations (11)-12), to describe a different productivity of inputs for students who repeated a grade.

The test scores in French, denoted y_{f4}, and in Math, denoted y_{m4} are two different measures of the final human capital. For non-repeaters, with obvious notations for the random error terms, we have,

$$y_{mi4} = h_{mi4} + \varepsilon_{mi4}, \tag{15}$$

$$y_{fi4} = h_{fi4} + \varepsilon_{fi4}, \tag{16}$$

where ε_{m4} and ε_{f4} are independent normal random variables. For repeaters, at the end of grade 9, test scores in French are observed in year $t = 5$ and denoted y_{f5}. Similarly, test scores in Mathematics are denoted y_{m5}. We have two different measures of the repeaters' final human capital, with obvious notations for the independent random error terms,

$$y_{mi5} = h_{mi5} + \varepsilon_{mi5}, \tag{17}$$

$$y_{fi5} = h_{fi5} + \varepsilon_{fi5}. \tag{18}$$

The functions h_{mit} and h_{fit}, with $t = 4, 5$ have the same specification as h_{it} (as given by (9) above), with coefficients a_{mt}, b_{mt}, c_{mt} and a_{ft}, b_{ft}, c_{ft}, etc., that may be different for Mathematics and French. Our model is now fully specified.

5 Estimation Method

The estimation method is a variation on the EM algorithm. Let Y_i be the set of outcomes observed for individual i : $Y_i = (y_{mi0}, y_{fi0}, S_{i1}, ..., S_{i4}, y_{mi4}, y_{fi4})$. Let $X = (X_1, X_2, X_3, X_4, X_5)$ and $Z = (Z_1, Z_2, Z_3, Z_4, Z_5)$. Then, we denote θ the vector of all model parameters, namely, $\theta = (p_1, p_2, p_3, p_4, a_i, b_i, c_{ij}, \alpha_i, \dots)$. We replicate each individual i in the sample to create 4 different artificial observations of i. Student i's replicas differ by the unobserved type, or group k only, but the values of X_i, Y_i and Z_i are the same for each replica. We arbitrarily choose initial values for the unconditional prior probabilities of the groups p_k, $k = 1, \dots, 4$, and for the posterior probabilities of belonging to a certain group knowing the observed characteristics of i, that is, $p_{ik} = \mathbb{P}(G_{ik} = 1|Y, X, Z)$. They will be updated after each iteration.

The estimation algorithm can be described as follows.

1. We first run 20 *weighted* regressions and Ordered Probits.
   (a) Two regressions for the initial test scores in Math and French.
   (b) Two regressions of class size by grade: one for the repeaters and one for the non-repeaters (except for the first year, because there are only non-repeaters in year $t = 1$ and for year $t = 5$, because there are only repeaters). This amounts to 8 regressions.
   (c) One simple Probit to model the transition at the end of grade 6 in year $t = 1$. Two Ordered Probits to model the decision at the end of grades 7 and 8 for non-repeaters. Three simple Probits to model steering decisions relative to repeaters in grades 6, 7 and 8. There are 4 Probits and 2 Ordered Probits in total.
   (d) Two final test-score regressions in Math and French, for repeaters and non-repeaters (4 regressions).
2. We obtain an estimation of θ by means of our system of weighted regressions and weighted Probits.

3. The residuals of regressions and the probabilities of passing to the next grade are collected to compute the individual contributions to likelihood, that is, by definition,

$$l_i(X,Z,Y,\theta) = \sum_{k=1}^{K} p_k l_i(Y,X,Z,\theta \mid G_{ik}=1). \tag{19}$$

4. Individual posterior probabilities p_{ik} of belonging to a group are then updated, using Bayes' rule and the likelihood as follows,

$$p_{ik} = \mathbb{P}(G_{ik}=1|Y,X,Z,\theta) = \frac{p_k l_i(Y,X,Z,\theta \mid G_{ik}=1)}{\sum_{j=1}^{K} p_j l_i(Y,X,Z,\theta|G_{jk}=1)}. \tag{20}$$

   These individual probabilities are then averaged to update the prior probabilities p_k, as follows,

$$p_k = \mathbb{P}(G_k=1) = \frac{1}{N}\sum_{i=1}^{N} p_{ik}. \tag{21}$$

5. A new iteration begins until convergence of the estimated unconditional probabilities.

All standard deviations have been bootstrapped, using 50 drawings with replacement in the sample.

The estimation method used here has been advocated and justified by various authors (see, *e.g.*, Arcidiacono and Jones (2003), Bonhomme and Robin (2009)).

6 Estimation Results

6.1 Distribution of groups

The results of the algorithm, using $K=4$ groups, are given by Table 6. We chose to use only 4 groups because of weak identifiability and computational problems when $K>4$. In Table 7, we compare the most likely groups of individuals, estimated with the full model, called Classification 1, with the results of a limited sub-model, based on grade 6 entry scores only, called Classification 2. Both models have 4 unobserved types or groups. This has been done to try to assess the impact of initial test scores on the individual's posterior probabilities of belonging to a group. In other words, are students fully predetermined by their initial stock of knowledge? We observe that, according to Classification 2, 75% of Group 1 individuals are also most likely to become members of Group 1, according to Classification 1 (the full model). Observing the grade 6 scores

Table 6: Estimated Group Probabilities

	Group 1	Group 2	Group 3	Group 4
Probabilities	15.54 %	31.16 %	33.56 %	19.74 %
	(0.69)	(0.64)	(0.58)	(0.82)

Table 7: Comparison of Two Classifications

	Classification 2				
Classification 1	Group 1	Group 2	Group 3	Group 4	Total
Group 1	74 %	1 %	0 %	3 %	2021
Group 2	24 %	59 %	2 %	61 %	4076
Group 3	0 %	38 %	48 %	34 %	4383
Group 4	2 %	2 %	50 %	2 %	2656
	100 %	100 %	100 %	100 %	
Total	2547	2883	4967	2739	13136

in Math and French only allows us to assign the student to the first group, to a large extent. But Group 4 students are not predetermined by their entry test scores, since less that 2% of the students assigned to Group 4 on the basis of the latter scores end up being members of Group 4 in the full model. The corresponding percentages are 59% and 48% for Groups 2 and 3, respectively. We conclude that, with the exception of Group 1, unobserved types are far from being perfectly predicted in year $t = 1$ (*i.e.*, in grade 6). It seems that the weakest students are easily detected from the beginning, but the brightest students are not. We will come back to this point in the general discussion of estimation results below.

Table 8 presents the parameters obtained when we regress the individual posterior probabilities of belonging to a certain group k, defined as p_{ik} above, on the socio-demographic and family-background variables X_0. We find that the probabilities of belonging to the two extreme groups, Group 1 and Group 4, are quite well predicted by the social background, with an R^2 superior to 14%. These results also show, among other things, that when the mother is educated and the father is an executive, the probability of belonging to Group 4 is significantly increased. Group 2 and Group 3 are not so easy to distinguish on the basis of observed student characteristics.

6.2 Group effects on test scores

We present here the estimated parameters of group effects and class size. Table 9 shows the estimated coefficients for the initial test scores (at the beginning of grade 6) and the final test scores (at the end of grade 9). Group 1 is the reference. We see how well the four groups are defined. Scores in French and Math increase with group k and the estimated coefficients yield the same ranking of ability groups in all columns, except the rightmost column of Table 9. More precisely, Group 4 has everywhere the highest scores, with the exception of Group 4 repeaters, in French, but the latter coefficient is estimated with less precision than the others. Intuitively, this is because Group 4 students have a low probability of repeating a grade. Apart from this exception, Group 4 is above Group 3, which in turn dominates Group 2, and Group 1 is unambiguously the lowest ability group.

Table 8: Individual Group Probabilities and Family Background

	Group 1	Group 2	Group 3	Group 4
Female	0.0493***	0.0388***	-0.00773	-0.0804***
	(0.00565)	(0.00766)	(0.00779)	(0.00638)
Mother education:	-0.0126	-0.0167	0.0234**	0.00594
Junior High School	(0.00873)	(0.0118)	(0.0120)	(0.00985)
Mother education:	-0.0521***	-0.0175	0.0531***	0.0165
Vocational Certificate	(0.00937)	(0.0127)	(0.0129)	(0.0106)
Mother education:	-0.0901***	-0.103***	0.0550***	0.138***
High-School Graduate	(0.0109)	(0.0147)	(0.0150)	(0.0123)
Mother education:	-0.0864***	-0.154***	0.0832***	0.157***
2 years of college	(0.0118)	(0.0160)	(0.0162)	(0.0133)
Mother education:	-0.103***	-0.174***	0.0240	0.253***
4 years of college and more	(0.0142)	(0.0192)	(0.0195)	(0.0160)
Father occupation:	-0.0514***	-0.0373	0.0100	0.0786***
Executives and educated professionals	(0.0181)	(0.0245)	(0.0250)	(0.0204)
Father occupation:	-0.00141	0.0674***	-0.0314	-0.0346*
White collars	(0.0184)	(0.0249)	(0.0253)	(0.0207)
Father occupation:	0.0557***	0.0777***	-0.0696***	-0.0638***
Blue collars	(0.0169)	(0.0229)	(0.0233)	(0.0191)
More than three children in family	0.0845***	0.0004	-0.0432***	-0.0418***
	(0.00840)	(0.0114)	(0.0116)	(0.00949)
Retention in primary school	0.206***	0.0412***	-0.169***	-0.0781***
	(0.00731)	(0.00990)	(0.0101)	(0.00825)
Quarter of birth				
Q2	0.0000	0.0144	-0.0193*	0.0049
	(0.0077)	(0.0105)	(0.0107)	(0.0087)
Q3	0.0198**	0.0256**	-0.0331***	-0.0123
	(0.0085)	(0.0115)	(0.0117)	(0.0096)
Q4	0.0145*	0.0463***	-0.0260**	-0.0349***
	(0.0087)	(0.0117)	(0.0119)	(0.0098)
R^2	0.187	0.059	0.060	0.143

Linear regressions of probabilities p_{ik} on controls X_0. Standard errors are in parentheses; ***, ** and * indicate significance at the levels of 1, 5, and 10%, resp. There are 12,937 observations.

Table 9: Estimated Impact of Groups and Class Size on Test Scores

	Score in Math			Score in French		
	Initial	Final		Initial	Final	
		Non-repeaters	Repeaters		Non-repeaters	Repeaters
Class size $t = 4$		-0.25***			-0.25***	
		(0.03)			(0.04)	
Class size $t = 5$			-0.19***			-0.25***
			(0.07)			(0.05)
Group 2	10.44***	8.14***	5.32***	10.82***	9.10***	5.80***
	(0.27)	(0.57)	(0.67)	(0.23)	(0.65)	(0.69)
Group 3	19.17***	15.80***	9.07***	19.16***	16.65***	10.13***
	(0.22)	(0.62)	(0.91)	(0.30)	(0.61)	(0.73)
Group 4	25.42***	26.18***	16.05***	25.60***	27.50***	9.22**
	(0.25)	(0.62)	(5.24)	(0.28)	(0.68)	(5.18)
Constant	35.34***	41.88***	43.31***	35.20***	40.87***	44.29***
	(0.24)	(0.92)	(1.71)	(0.26)	(0.94)	(1.20)
R^2	0.68	0.60	0.18	0.68	0.63	0.21

Standard errors are in parentheses; ***, ** and * indicate significance at the levels of 1, 5, and 10%, resp.

If we now focus on final scores, it is easy to see that Group 1 gets higher scores on average when a grade was repeated (*i.e.*, this is because the constant is higher). In contrast with Group 1, individuals in Groups 3 and 4 who didn't repeat a grade obtain higher scores than the repeaters of these two groups. Take Group 3 for instance. To obtain the final score in Math of the average Group 3 student who repeated a grade, we add the constant in the column, *i.e.*, 43.31 to the differential impact of Group 3, *i.e.*, 9.07. The total is 52.38. But if we compute the corresponding term for Group 3 non-repeaters, in Math, we obtain, $15.80 + 41.88 = 57.68$. Grade repetition seems detrimental to Group 3. The same is true with Group 4. For the latter group, the corresponding additions yield 68.06 in the non-repeaters' column and 59.36 in the repeaters' column. However, individuals in Group 2 get approximately the same increase in their score, whether they repeat or not.

6.3 Promotion decision model and effects of class size

If we now look at the top rows in Table 9, we find that increasing class-size has a negative impact in grade 9 for all students. The standard deviation of class size is around 3. [7] It follows that the estimated impact of a standard deviation of class size is around three quarters of a normalized test-score point for non-repeaters, or 7.5% of the standard deviation of test scores. The significant negative coefficient on class-size appears because we control for unobserved heterogeneity, and therefore, for the endogeneity of this variable. Otherwise, the coefficient on class size would be positive (we return to this question below, when we discuss the class-size regressions). This being said, we do not find a very strong class-size effect on final scores (a quarter of a point, or 1/40th of the standard deviation of test scores, for a one-student reduction in class-size). Table 10 shows the main parameters of the promotion decision model. Dependent variables determine rows, while the coefficients of a given explanatory variable in equations are displayed in the same column. A higher group label means a higher average knowledge-capital. As a consequence, the greater the group label, the greater the probability of passing to the next grade, for non-repeaters as well as for repeaters, in each grade. The estimated coefficients reflect this ranking of groups very clearly, again, with the exception of the impact of Group 4 in the Probit concerning grade 8 repeaters (*i.e.*, S_4 repeaters). The latter coefficient is not estimated with precision because Group 4 students have a small probability of repeating a grade. Apart from this exception, all other coefficients are estimated with good precision. The first column of Table 10 shows that increasing class size decreases the probability of promotion to grade 7, but has a non-significant (or even a positive impact) on pass rates in later grades.

6.4 Endogeneity of class size

Table 11 finally gives the coefficients of group dummies and of instruments in class-size equations. Each row in the table corresponds to a dependent variable. One of the class-size instruments is

[7] To be precise, the standard deviation of class size in year t, denoted σ_{nt} has the following values $\sigma_{n1} = 3.02$, $\sigma_{n2} = 2.90$, $\sigma_{n3} = 3.32$, $\sigma_{n4} = 3.38$.

Table 10: Estimated Impact of Groups and Class Size on Promotion Decisions

Dependent variable ↓	Class size	Group 2	Group 3	Group 4	Cut 1 δ_{t1}	Cut 2 δ_{t2}	Cut R δ_{tr}
S_1	-0.025^{***}	0.67^{***}	2.24^{***}	2.45^{***}	-1.13^{***}		
	(0.007)	(0.04)	(0.12)	(0.72)	(0.16)		
S_2 repeaters	0.010^{**}	4.29^{***}	4.22^{***}	3.17^{***}			-1.80^{*}
	(0.04)	(0.42)	(0.61)	(1.3)			(1.12)
S_2	-0.004	0.63^{***}	1.62^{***}	2.72^{***}	-0.85^{***}	-0.08	
	(0.006)	(0.044)	(0.057)	(0.64)	(0.14)	(0.14)	
S_3 repeaters	-0.016^{*}	0.38^{***}	0.92^{***}	4.43^{***}			-0.93^{***}
	(0.012)	(0.017)	(0.24)	(1.37)			(0.29)
S_3	0.045^{***}	0.33^{***}	0.92^{***}	1.67^{***}	-0.64^{***}	0.34^{**}	
	(0.006)	(0.05)	(0.06)	(0.17)	(0.18)	(0.16)	
S_4 repeaters	0.035^{***}	0.33^{***}	0.65^{***}	0.55			-0.002
	(0.01)	(0.07)	(0.12)	(1.91)			(0.24)

The promotion decisions S_t are modeled with the help of an Ordered Probit. They take the value 0 for redirection, 1 for retention and 2 for pass. Standard errors are in parentheses; ***, ** and * indicate significance at the levels of 1, 5, and 10%, resp.

Table 11: Estimates of Class-Size Equation Parameters

Dependent variable ↓	Maimonides' Rule	Constant	Group 2	Group 3	Group 4	R^2
Class size $t = 1$	0.32***	16.09***	1.12***	1.75***	1.78***	0.20
	(0.02)	(0.36)	(0.18)	(0.15)	(0.16)	
Class size $t = 2$	0.49***	14.81***	0.68***	-5.75***	3.15*	0.25
(repeaters)	(0.05)	(1.09)	(0.27)	(1.51)	(2.35)	
Class size $t = 2$	0.37***	15.17***	1.07***	1.85***	1.96***	0.21
	(0.02)	(0.17)	(0.30)	(0.14)	(0.16)	
Class size $t = 3$	0.36***	16.06***	0.53***	1.13***	-1.96*	0.18
(repeaters)	(0.05)	(0.91)	(0.17)	(0.32)	(1.27)	
Class size $t = 3$	0.35***	13.66***	1.87***	2.90***	3.10***	0.24
	(0.05)	(0.40)	(0.22)	(0.20)	(0.26)	
Class size $t = 4$	0.33***	15.61***	0.95***	2.00***	1.85	0.19
(repeaters)	(0.05)	(0.90)	(0.22)	(0.29)	(2.16)	
Class size $t = 4$	0.35***	14.05***	1.62***	2.62***	2.94***	0.26
	(0.02)	(0.46)	(0.34)	(0.26)	(0.29)	
Class size $t = 5$	0.26***	16.34***	0.91***	2.67***	0.32	0.22
(repeaters)	(0.04)	(0.73)	(0.32)	(0.31)	(2.92)	

Standard errors are in parentheses; ***, ** and * indicate significance at the levels of 1, 5, and 10%, resp.

theoretical class size (*i.e.*, Maimonides' rule), that is, the class size that would be experienced by the student if a class-opening threshold of 25 was applied, given total grade enrollment. The coefficient of this variable is significant and positive, as expected. We also find that class size increases with the ability (*i.e.*, the group) of students. The only exceptions are the coefficients on Group 4 dummies, that cannot be estimated with precision among grade repeaters. These results prove that class-size is strongly endogenous, and that it is used as a remediation instrument by school principals.

Our estimates are robust if the group dummies are exogenous variables in each year. To check this, we regressed the posterior probabilities of belonging to a group over a set of permanent individual characteristics X_0 and the time-varying characteristics X_1, X_2, X_3, X_4, X_5. The results of these latter regressions are not presented here, but they show that, if the coefficients on X_0 are strongly significant, in contrast, time-varying characteristics are not significant. Thus, our model seems well specified (and we found a confirmation of well-known results). A better social background (that is, richer, more educated and more qualified parents) significantly increases the initial capital and therefore, the probability of belonging to high-ability groups.

7 The Treatment Effects of Grade Retention

We now turn to the key question of the present paper: the treatment effects of grade repetition. The model will be used to compute counterfactuals.

7.1 Effect of grade retention on grade 9 scores

Each individual i has a posterior conditional probability p_{ik} of belonging to each of the four groups $k = 1, \dots, 4$. For each individual and each of his (her) possible types, we compute a counterfactual class size and a counterfactual final test score. Each individual has four counterfactual final scores and four counterfactual final class sizes. Using the posterior probabilities, we can then compute expected counterfactual grades.

For each group, and for each student who hasn't repeated a grade,

1. we compute the class size he or she would have experienced in grade 9, if he or she had repeated a grade.

   To do this, we assume that the student doesn't move to a different school and that his class environment has the same characteristics (same number of foreigners, same foreign language chosen, same size of the urban area, same sector (private or public), same classification as priority education zone). However, we use the information that we have on total school enrollment and total grade enrollment in the same school one year later.

2. we compute the grade predicted in grade 9 if the student had repeated a grade (this counterfactual is denoted Y_r^c).

For each grade repeater and each group,

1. we compute the class size predicted in grade 9 if the student had not repeated a grade;

2. we compute the student's predicted grade in grade 9 if he or she had not repeated a grade (this counterfactual is denoted Y^c).

Let N_r denote the number of individuals who repeated a grade and let N_p denote the number of individuals who didn't repeat a grade. Of course, we have, $N = N_p + N_r$. Let y_{ri} be the observed final grade of i, if i is a repeater. Let y_i be the observed final grade of i, if i never repeated a grade. We can now compute the following treatment effects.

The average treatment effect (*i.e.*, ATE) is defined as follows.

$$ATE = \frac{1}{N}\left(\sum_{i\in\mathbf{N}_p}\sum_{k=1}^{4}(\mathbb{E}(Y_{ri}^c \mid G_{ik}=1) - y_i)p_{ik} + \sum_{i\in\mathbf{N}_r}\sum_{k=1}^{4}(y_{ri} - \mathbb{E}(Y_i^c \mid G_{ik}=1))p_{ik}\right), \quad (22)$$

where $p_{ik} = \mathbb{P}(G_{ki} = 1|X, Z, Y)$ is i's posterior probability of belonging to Group k. In the above expression, $\mathbb{E}(Y_{ri}^c|G_{ik} = 1)$ and $\mathbb{E}(Y_i^c|G_{ik} = 1))$ are the predictions of i's final grades, in the counterfactual situations of grade repetition and not repeating, respectively, using the estimated regression functions, and conditional on belonging to Group k.

The average treatment effect on the treated (*i.e.*, ATT) is then,

$$ATT = \frac{1}{N_r}\sum_{k=1}^{4}\sum_{i\in\mathbf{N}_r}(y_{ri} - \mathbb{E}(Y_i^c \mid G_{ik}=1))p_{ik}. \quad (23)$$

We also compute an ATE by group. For Group k, the average treatment effect ATE_k is defined as,

$$ATE_k = \frac{1}{Np_k}\left(\sum_{i\in\mathbf{N}_p}(\mathbb{E}(Y_{ri}^c \mid G_{ik}=1) - y_i)p_{ik} + \sum_{i\in\mathbf{N}_r}(y_{ri} - \mathbb{E}(Y_i^c \mid G_{ik}=1))p_{ik}\right), \quad (24)$$

where $p_k = (1/N)\sum_i p_{ik}$. The, ATT within group k, denoted ATT_k, can be defined in a similar way,

$$ATT_k = \frac{1}{\sum_{i\in\mathbf{N}_r} p_{ik}}\sum_{i\in\mathbf{N}_r}(y_{ri} - \mathbb{E}(Y_i^c \mid G_{ik}=1))p_{ik}. \quad (25)$$

7.2 Effect of grade retention on the probability of access to grade 9

Individual i's estimated probability of access to grade 9, knowing Group k, is denoted P_{9ik} and can be decomposed in the following way:

$$\begin{aligned}
P_{9ik} &= \Pr(S_{i1}=P \mid k)\Pr(S_{i2}=P \mid k)\Pr(S_{i3}=P \mid k) \ \textit{(does not repeat)}\\
&+ \Pr(S_{i1}=P \mid k)\Pr(S_{i2}=P \mid k)\Pr(S_{i3}=R \mid k)\Pr(S_{i4r}=P \mid k) \ \textit{(repeats grade 8)}\\
&+ \Pr(S_{i1}=P \mid k)\Pr(S_{i2}=R \mid k)\Pr(S_{i3r}=P \mid k)\Pr(S_{i4r}=P \mid k) \ \textit{(repeats grade 7)}\\
&+ \Pr(S_{i1}=R \mid k)\Pr(S_{i2r}=P \mid k)\Pr(S_{i3r}=P \mid k)\Pr(S_{i4r}=P \mid k), \ \textit{(repeats grade 6)}
\end{aligned}$$

Table 12: Counterfactuals required to compute the probabilities of accessing grade 9

	Grade 7		Grade 6R		Grade 8		Grade 7R		Grade 8R	
History	$\Pr(S_2)$	n_2	$\Pr(S_{2r})$	n_{2r}	$\Pr(S_3)$	n_3	$\Pr(S_{3r})$	n_{3r}	$\Pr(S_{4r})$	n_{4r}
1234			C	C			C	C	C	C
12334			C	C			C	C		
12234			C	C	C	C				
11234	C	C			C	C				
1233V			C	C			C	C		
1223V			C	C	C	C				
1123V	C	C			C	C				
123V			C	C			C	C	C	C
122V			C	C	C	C			C	C
112V	C	C			C	C			C	C
12V			C	C	C	C	C	C	C	C
11V	C	C			C	C	C	C	C	C

Letter C indicates that a counterfactual value has been computed. Letter R indicates that a grade-repeater model is used. $\Pr(S_t)$ means the probability distribution of decision $S_t \in \{P, V, R\}$. n_t denotes class-size in year t. Subscript r indicates the specific model for grade repeaters, $S_{tr} \in \{P, V\}$.

where, to simplify notation, we denote $\Pr(S_{it} = X \mid k) = \Pr(S_{it} = X \mid G_{ik} = 1)$, for all $X = P, R, V$. If the government decides to abolish grade retention (but keeps the possibility of steering students towards the vocational track) then, the only way of reaching grade 9 is to pass the three grades directly. Let P^c_{9ik} be the counterfactual probability of accessing grade 9 when grade retention is abolished. Given that no student is redirected to the vocational track at the end of grade 6, this probability can be expressed as follows,

$$P^c_{9ik} = \Pr(S_{i2} = P \mid k) \Pr(S_{i3} = P \mid k).$$

To find the average treatment effect of grade retention, we need to compute the individual probabilities P_{9ik} and P^c_{9ik} for all the students in the sample, including those who have actually been redirected. This requires the computation of many counterfactuals. For those who repeated grade 6 and then passed or were redirected, we need counterfactual class sizes and counterfactual school-environment characteristics for year 2 and 3, that they would have experienced, had they not repeated a grade. For those who repeated grade 5 or have been redirected at the end of grade 5, we need their counterfactual class size and counterfactual characteristics for year 3, as if they hadn't repeated this grade. Finally, for those who were never held back, we need the counterfactual class size and characteristics that they would have experienced, had they repeated a grade. Table 12 summarizes the counterfactual probabilities and the counterfactual class size we computed for each different grade history. Then we can compute the following treatment

Table 13: Average Treatment Effects of Grade Retention

	Mathematics		French		Probability of access to grade 9	
	ATE	ATT	ATE	ATT	ATE	ATT
Group 1	2.43	2.45	3.09	3.20	−0.11	−0.11
	(0.76)	(0.76)	(0.81)	(0.80)	(0.014)	(0.014)
Group 2	0.12	0.36	0.18	0.47	−0.12	−0.12
	(0.42)	(0.42)	(0.41)	(0.42)	(0.012)	(0.012)
Group 3	−3.79	−2.92	−3.66	−2.77	−0.09	−0.10
	(0.76)	(0.75)	(0.60)	(0.59)	(0.022)	(0.023)
Group 4	−6.68	−14.08	−6.86	−14.22	−0.06	−0.06
	(4.61)	(4.53)	(4.54)	(4.52)	(0.07)	(0.07)
All	−2.56	0.27	−3.73	0.71	−0.09	−0.11
	(0.85)	(0.31)	(0.94)	(0.33)	(0.017)	(0.008)

Standard deviations are in parentheses.

effects. The average treatment effect is,

$$ATE = \frac{1}{N}\sum_{k=1}^{4}\left(\sum_{i\in\mathbf{N}_r}(P_{9ik}-P^c_{9ik})p_{ik}+\sum_{i\in\mathbf{N}_p}(P_{9ik}-P^c_{9ik})p_{ik}\right).$$

The average treatment effect on the treated is then,

$$ATT = \frac{1}{N_r}\sum_{k=1}^{4}\sum_{i\in\mathbf{N}_r}(P_{9ik}-P^c_{9ik})p_{ik}.$$

7.3 Results and discussion

Table 13 displays the results of the various computations. The last row in this table shows the overall results. If we consider the final tests scores in Math and French (at the end of grade 9), the ATT is positive, but small. Given that the mean value of the scores is 50 with a standard deviation of 10, the effects are smaller than a tenth of a standard deviation and barely significant. The ATE is clearly negative in Math and in French. As we will see, this is mainly due to the fact that the most able students would suffer from grade repetitions. If we now look at the values of ATE_k and ATT_k, the treatment effects within group k, it is easy to see that only Group 1 students benefit for grade repetitions. The effect of grade repetitions is not significantly different from zero for Group 2 students. In contrast, in the case of Group 3, and Group 4, both the ATE and the ATT are negative, in Math and in French. This shows that grade repetition hurts the students belonging to top groups.[8] We conclude that grade repetitions have some usefulness for the weakest students, with an effect of the order of a quarter of a standard deviation on the final grades.

We now discuss the effect on the probability of access to grade 9. The treatment effects of grade repetition on final scores rely essentially on the regression equations determining the final test scores, and the latter equations are estimated with the subset of individuals who reached grade 9. The fact that this population is selected is taken into account by the posterior individual probabilities p_{ik}. But it is reassuring to derive results for an outcome that depends on the entire structure of the model. This is the case of access to grade 9, because the probabilities P_{9ik}, defined above, depend on all the decision and class-size equations.

It is striking to see that in Table 13, the ATTs and ATEs of grade retention are all negative, even if we consider within-group treatment effects. This means that introducing grade retention, if grade retention does not already exists, will be detrimental to students, on average, and detrimental to students of each group, taken separately. The effects are particularly strong for Groups 1 and 2. To see this, we computed the distribution of the individual probabilities P_{9ik} and individual counterfactual probabilities P^c_{9ik} in the student population. The histograms of these distributions are displayed on Figure 2.

On Fig. 2 it is easy to see that the counterfactual probabilities have a mass near 1, meaning that the abolition of grade repetitions would help many students to reach grade 9. Yet, there are clearly subgroups of individuals that keep a low probability of access: these individuals bear a high risk of being tracked in vocational programs. We will understand the effect of grade repetition on access to grade 9 more fully if we compute the histograms of P_{9ik} and P^c_{9ik} separately for each group. This is done in the following figures. Figure 3 gives the distributions of P_{9ik}, while Figure 4 displays the distributions of the counterfactual P^c_{9ik}.

Comparing the histograms, it immediately comes to mind that when grade repetitions are abolished, access to grade 9 becomes certain for Group 3 and Group 4 students. The effect of abolition is less obvious for the weakest groups, 1 and 2, but in fact, these probabilities increase and become more favorable. To sum up, these effects explain why the treatment effects of grade repetitions on access to grade 9 are unambiguously negative. We see also that these effects are very strong, since a drop of 11 or 12 points of probability, very roughly, amounts to 50% of the best chances of access to grade 9 among Group 1 and Group 2 students.

The treatment effects are positive only for the weakest students, and these effects are weak when they are positive. Given these results, and the results of Table 13 in general, it seems that we can only recommend the abolition of grade retention. The results of Table 7 suggests a path for reform. Coming back to this table, we see that the weakest (*i.e.*, the Group 1) students are more easily detected in grade 6 than other types. In cases of grade retention, forcing weaker

[8]Note that ATT_k and ATE_k should be equal for each k, if Group k was the only variable used to predict counterfactual scores. But other control variables are used to predict these scores, such as class size, family background characteristics, etc. This determines differences between ATT_k and ATE_k in Table 13. However, the differences are neither large nor significant.

Figure 2: Histogram of Individual Probabilities of Access to Grade 9

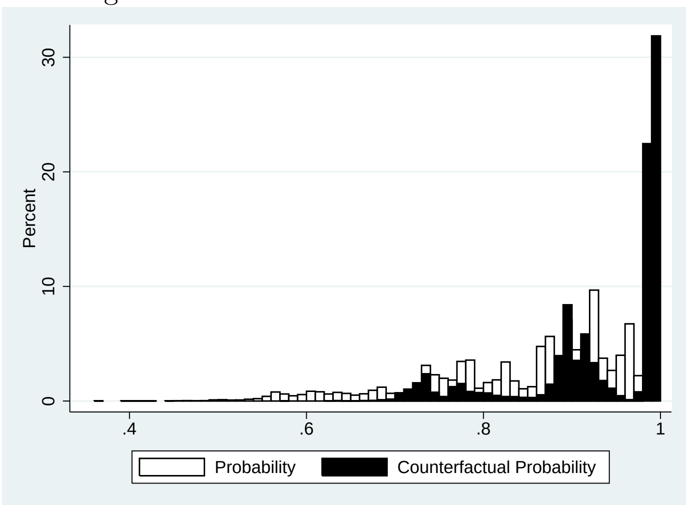


Figure 3: Histograms of Probabilities of Access to Grade 9, by Group

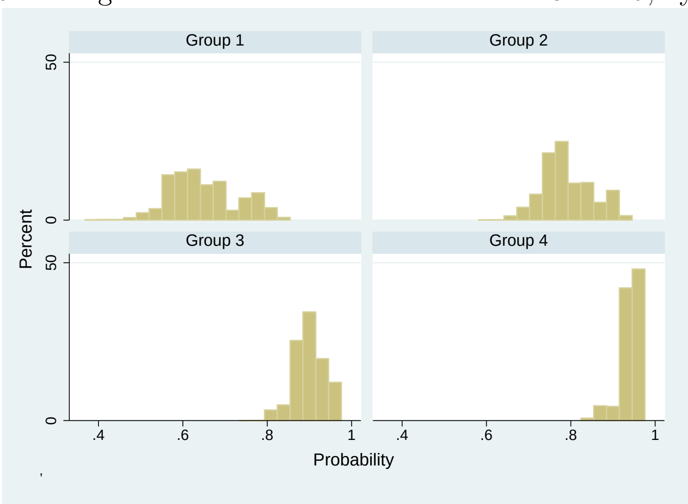
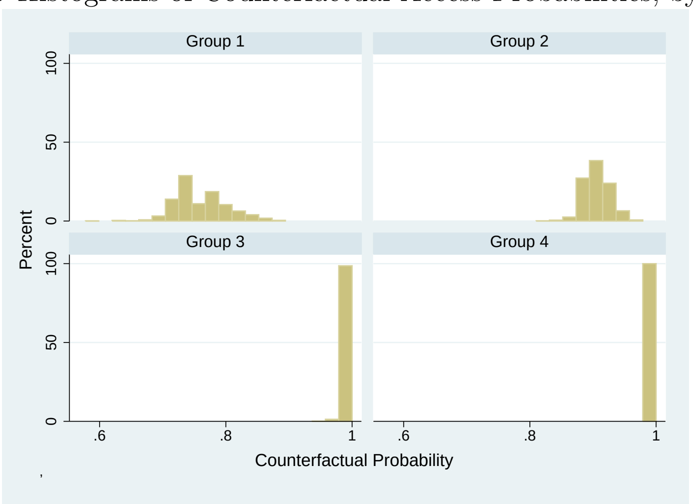


Figure 4: Histograms of Counterfactual Access Probabilities, by Group



students to follow the same teaching twice is only a rough second best. It would be more efficient to track these students from the start of junior high school, with additional remediation resources. One could imagine a slow track and a fast track, with, say, a year of difference in duration to reach the certification exams at the end of grade 9, and with flexible possibilities of track changes in both directions. To avoid the stigma of tracking, the slow track should probably be the norm, and students that seem promising would be steered towards the fast track. A system of that sort would lead to a more efficient use of resources than grade repetitions. It would clearly give weak students better chances of reaching the end of grade 9 with the required stock of knowledge and skills.

8 Conclusion

Grade retention is difficult to evaluate because grade repeaters have been selected on the basis of many characteristics that the econometrician doesn't observe. The difficult problem is to find a reasonable model to compute what would be the counterfactual performance of a student who has repeated a grade, if instead of being held back, he or she had been promoted to the next grade. To this end, we have assumed that the distribution of student test scores can be represented by a finite mixture of normal distributions, conditional on observed covariates, during each year of the observation period. The class size experienced by a student is also assumed to be distributed as a mixture of normals. All such mixtures are relying on the same finite number of latent student classes, called *groups*. In a flexible formulation, we show that class-size, probabilities of grade retention and test scores all depend on the unobserved group in a non-trivial and consistent way. We estimated a model with four groups and found that the four groups are unambiguously ranked. The higher the group index, the larger the student's ability, and the larger his class size. This proves that class size is endogenous, smaller classes being used by school principals to redistribute resources towards weaker students. With the help of our model, we computed counterfactual test scores to evaluate the average treatment effect and the average treatment effect *on the treated* of grade retention. We found that the ATE is negative, while the ATT is generally positive, but small. We computed treatment effects in each student group separately, and found that the ATE is positive for less able students and negative for more able students. Finally we computed the ATT and ATE of grade retention on the probability of access to grade 9, and found that this effect is significant and negative. Grade retention is a form of remedial education and seems to help the weakest students, insofar as it tends to increase their test scores at the end of grade 9. But these effects are weak. It follows that grade retention could probably be replaced by a form of tracking, or by different forms of remediation. Other studies have shown that grade retention is a stigma, that repeated years are interpreted as a negative signal by employers (on this point, see Brodaty *et al.* (2012)). The long-run effects of grade retention seem to be detrimental. We can only conclude that grade retention is unlikely to be an efficient public policy, because its impact on student performance — when positive — is weak.

9 References

ALET, Elodie (2010). "Is Grade Repetition a Second Chance?" *Manuscript. Toulouse School of Economics*, Toulouse, France.

ALET, Elodie, BONNAL, Liliane, and Pascal FAVARD (2013). "Repetition : Medicine for a Short-run Remission", *Annals of Economics and Statistics*, 111, 227-250.

ANGRIST, Joshua D., and Alan B. KRUEGER (1991). "Does Compulsory School Attendance Affect Schooling and Earnings?", *Quarterly Journal of Economics*, 106(4), 979-1014.

ANGRIST, Joshua D., and Victor LAVY (1999); "Using Maimonides' Rule to Estimate the Effect of Class Size on Scholastic Achievement", *Quarterly Journal of Economics*, 114, 533-575.

ARCIDIACONO, Peter and John B. JONES (2003), "Finite Mixture Distributions, Sequential Likelihood, and the EM Algorithm", Econometrica, 71(3), 933-946

BAERT, Stijn, COCKX, Bart, and Matteo PICCHIO (2013), "On Track Mobility, Grade Retention and Secondary School Completion", *manuscript, Ghent University*, Belgium.

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), 1437-1472.

BONHOMME, Stéphane and Jean-Marc ROBIN (2009), "Assessing the Equalizing Force of Mobility Using Short Panels: France, 1990-2000", *Review of Economic Studies*, 76(1), 63-92.

BRODATY, Thomas, GARY-BOBO, Robert J. and Ana PRIETO (2012). "Does Speed Signal Ability? The Impact of Grade Retention on Wages", manuscript,*CREST-ENSAE*, France.

CARNEIRO, Pedro, HANSEN, Karsten T., and James J. HECKMAN (2003),"Estimating Distributions of Treatment Effects with an Application to the Returns to Schooling and Measurement of the Effects of Uncertainty on College Choice", *International Economic Review*, 44(2), 361-422.

COOLEY, Jane, NAVARRO, Salvador and Yuha TAKAHASHI (2011). "How the Timing of Grade Retention Affects Outcomes: Identification and Estimation of Time-Varying Treatments Effects", manuscript.

CUNHA, Flavio, and James J. HECKMAN (2007). "The Technology of Skill Formation". Technical report, *National Bureau of Economic Research*, Cambridge, Massachusetts.

CUNHA, Flavio, and James J. HECKMAN (2008). "Formulating, Identifying and Estimating the Technology of Cognitive and Noncognitive Skill Formation", *Journal of Human Resources*, 43(4), 738-782.

CUNHA, Flavio, HECKMAN, James J. and Susanne M. SCHENNACH (2010). “Estimating the Technology of Cognitive and Noncognitive Skill Formation”, *Econometrica*, 78(3), 883-931.

DE FRAJA, Gianni, OLIVEIRA, Tania and Luisa ZANCHI (2010). “Must Try Harder: Evaluating the Role of Effort in Educational Attainment”, *Review of Economics and Statistics*, 92(2), 577-597.

D’HAULTFOEUILLE, Xavier (2010). “A New Instrumental Method for Dealing with Endogenous Selection”. *Journal of Econometrics*, 154(1), 1-15.

DONG, Yingying (2010). “Kept Back to Get Ahead? Kindergarten Retention and Academic Performance", *European Economic Review*, 54(2), 219-236.

EIDE, Eric R., and Mark H. SHOWALTER (2001). “The Effect of Grade Retention on Educational and Labor Market Outcomes”. *Economics of Education Review*, 20(6), 563-576.

GARY-BOBO, Robert J., and Mohamed-Badrane MAHJOUB (2013). "Estimation of Class-Size Effects, Using Maimonides’ Rule and Other Instruments: The Case of French Junior High Schools", *Annals of Economics and Statistics*, 111, 193-225.

GOMES-NETO, Joao Batista, and Eric A. HANUSHEK (1994). “Causes and Consequences of Grade Repetition: Evidence from Brazil. Economic Development and Cultural Change”, 43(1), 117-148.

GRENET, Julien (2010),“Academic Performance, Educational Trajectories, and the Persistence od Date-of-birth Effects: Evidence from France”, *manuscript, Center for Economic Performance*, London School of Economics, London, UK.

HECKMAN, James J. and Edward VYTLACIL (2005), “Structural Equations, Treatment Effects, and Econometric Policy Evaluation”, *Econometrica*, 73(3), 669-738.

HECKMAN, James J. (2010), “Building Bridges between Structural and Program Evaluation Approaches to Evaluating Policy”, *Journal of Economic Literature*, 48(2), 356-398.

HOLMES, Thomas C. (1989). “Grade Level Retention Effects: a Meta-Analysis of Research Studies”. in: *Flunking Grades : Research and Policies on Retention*, Falmer Press, Bristol.

HOLMES, Thomas C., and Kenneth M. MATTHEWS (1984). “The Effect on Nonpromotion on Elementary and Junior High-School Pupils: A Meta-Analysis”. *Review of Educational Research*, 54(2), 225-236.

HOXBY, Caroline M. (2000); “The Effects of Class Size on Student Achievement: New Evidence from Population Variation”, *Quarterly Journal of Economics*, 115, 1239-1285.

Imbens, Guido, W. and Joshua D. Angrist (1994), “Identification and Estimation of Local Average Treatment Effects”, *Econometrica*, 62(2), 467-475.

Jacob, Brian A., and Lars Lefgren (2004). “Remedial Education and Student Achievement: A Regression-Discontinuity Analysis”. *Review of Economics and Statistics*, 86(1), 226-244.

Jacob, Brian A., and Lars Lefgren (2009). “The Effect of Grade Retention on High School Completion”. *American Economic Journal: Applied Economics*, 1(3), 33-58.

Mahjoub, Mohamed-Badrane (2007). "The Treatment Effect of Grade Repetitions", manuscript, *Paris School of Economics*, Paris, France.

Mahjoub, Mohamed-Badrane (2009). "Essais en micro-économie de l’éducation", *University of Paris I Panthéon-Sorbonne*, doctoral dissertation, R. Gary-Bobo, supervisor.

Manacorda, Marco (2012). “The Cost of Grade Retention.” *Review of Economics and Statistics*, 94(2), 596-606.

Neal, Derek, and Diane Whitmore-Schanzenbach (2010). “Left Behind by Design: Proficiency Counts and Test-Based Accountability”. *Review of Economics and Statistics*, 92(2), 263-283.

Piketty, Thomas, and Mathieu Valdenaire (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”, *Les Dossiers, Ministère de l’Education Nationale*, no 173.

Table 14: Descriptive Statistics for Value Added

	Math		French	
Standardized score	Balanced sample[a]	Repeaters	Balanced sample[a]	Repeaters
Grade 6	51.10	43.25	51.21	43.38
	(9.55)	(8.48)	(9.47)	(8.44)
Grade 9	50	43.37	50	43.46
	(10)	(8.23)	(10)	(7.87)
VA = Grade 9 − Grade 6	−1.10	0.11	−1.21	0.08
	(8.55)	(9.63)	(8.39)	(9.18)

Note [a]. Sample of all pupils for whom a test score is available both in grade 6 and in grade 9.

10 Appendix: Details on Quarter of Birth as an Instrument for Grade Retention

Table 14 displays descriptive statistics on value-added. Scores in grade 6, ranging between 0 and 20, as is usual in French schools, are standardized to have a mean of 50 and a standard deviation of 10 in the whole sample in grade 6 (including all redirected pupils). Scores in grade 9 are standardized in the same way in the sample of individuals who reached grade 9. Table 14 shows that value added, the sign of which is irrelevant because scores are measures of performance relative to each grade, is nevertheless higher for repeaters than for non-repeaters. This is true both in French and Mathematics. There exists a strong link between the age of a child, as measured by the month of birth, or quarter of birth, and the probability of grade repetition. A look at Figure 5 shows the frequency of grade retention by quarter of birth[9]. The probability of grade retention is clearly higher for children born later in the year. In principle, children must be 6 years old on September 1rst of year t to be admitted in primary school, grade 1, year t. In practice, many 5-year-old children born between October and December are admitted, but the 5-year-old children born in the first quarter typically have to wait until the next year. It follows that first-quarter students tend to be relatively older in their class, with an age difference that can reach 11 months. Older children being more mature, they tend to perform better. At the same time, teachers are reluctant to retain older children in one grade as retention may change a difference —being older— into a stigma —being too old.

Figure 6 shows that initial (grade 6 entry) scores decrease with quarter of birth. The decreasing trend also exists for final scores but is less pronounced. Figure 7 shows that value-added scores tend to be higher for relatively younger students, who seem to be catching up during their junior high-school years. In a first attempt to check if this is attributable to grade retention, we plot value-added by quarter of birth separately for repeaters and non-repeaters. Figure 8 clearly shows that value-added age profiles are steeper for repeaters than for non-repeaters.

[9]Due to the survey protocol, there are no observations for students born in March, July and October.

Figure 5: Probability of Grade Retention by Quarter of Birth

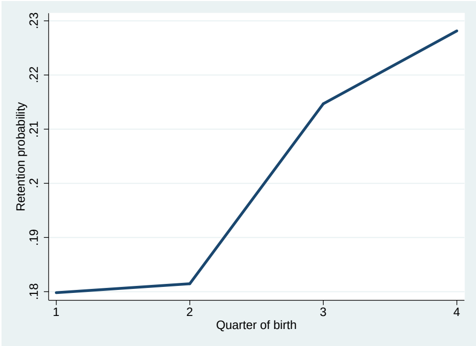


Figure 6: Scores by Quarter of Birth

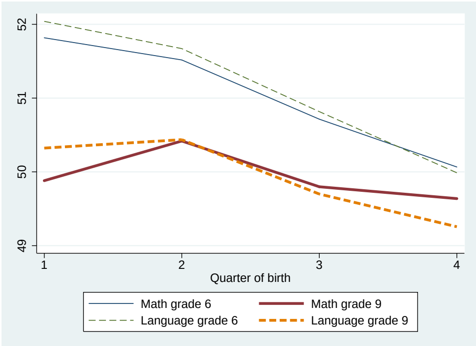


Figure 7: Value Added by Quarter of Birth

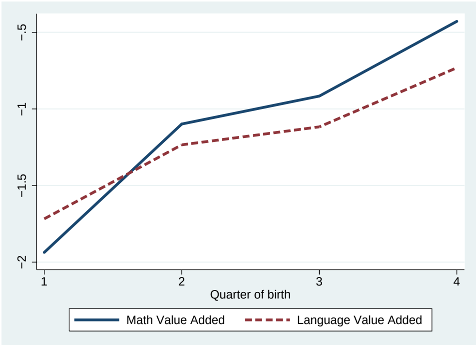


Figure 8: Value Added by Quarter of Birth for Repeaters and Non-Repeaters