Thema

UMR 8184

THéorie Économique, Modélisation et Applications

Thema Working Paper n°2012-47
Université de Cergy Pontoise, France

Is grade repetition one of the causes of early school dropout? Evidence from Senegalese primary schools.

Pierre André

UNIVERSITÉ de Cergy-Pontoise

CNRS CENTRE NATIONAL DE LA RECHERCHE SCIENTIFIQUE

November, 2012

Is grade repetition one of the causes of early school dropout? Evidence from Senegalese primary schools.*

Pierre André†

November 19, 2012

JEL Classification: D12, I28, O12
Key words: Grade repetition, School demand, School dropouts, Senegal

Abstract

This paper investigates the connection between grade repetition and school outcomes. It uses the fact that pupils need to meet class-specific standards to pass to the next grade. It measures the differences in the link between learning achievement and grade repetition between classes with different requirements to pass to the next grade. This double difference identifies the effect of grade repetition. The results show a negative effect of the grade repetition decision on the probability to be enrolled at school the next year, and on the probability to start secondary school.

Despite this mechanism, pupils from schools with tough grade repetition policies are on average more likely to be enrolled during the follow-up survey and to start secondary school. These schools do not seem to be located in particularly favorable places for this. This emphasizes that grade repetition policies might have other consequences than affecting repeating pupils.

*I am grateful to Sylvie Lambert for detailed comments on many versions of this paper. Luc Behaghel, Christelle Dumas, Paul Glewwe, Marc Gurgand, Marco Manacorda and Sandrine Mesplé-Somps should also be praised for very useful comments and help on earlier versions of this paper.

†Université de Cergy-Pontoise, THEMA, F-95000 Cergy-Pontoise, pierre.andre@u-cergy.fr

1 Introduction

Primary education in many African countries is characterized by particularly high repetition rates. Some 7.5% of the pupils enrolled in Senegalese primary schools in 2011 were repeating their grades in 2005,[1] whereas the African average is 13% (in 2002)[2]. Besides, dropouts before completion of primary school are frequent in these countries: some 40% of the Senegales children enrolled in the first grade of primary school (and 32% of African pupils) do not achieve the last one.[1,2] Using a cross-country regression, Manacorda (2012) remarks that grade repetition is more widespread in countries where gross enrolment rates in secondary school are low, raising the question of the causality behind this correlation.

High repetition rates in primary school can decrease national school attainment rates because grade repetition is very expensive for the state and households alike. Indeed, for a given final grade, a grade repetition increases the time spent at school. Hence both private and public costs of schooling increase with a grade repetition, and grade repetition decreases the returns to human capital investments.

Besides, grade repetition can decrease school attainment because it is discouraging for children. Some psychologists as Jimerson, Carlson, Rotert, Egeland, and Sourie (1997) consider that early grade repetition has a negative effect on socio-emotional adjustment. In economic terms, grade repetition may be a negative signal about a child's ability. If the parents observe their children's ability noisily, then grade repetition diminishes parents' belief in their children's ability (and/or the children's beliefs on their own abilities).

The discouraging effect of grade repetition can be mitigated by its pedagogic effect. The pedagogic benefits of grade repetition are nevertheless uncertain. When children repeat grades, they may consolidate the skills taught at those grades. However, it is unclear whether this offsets their failure to acquire the skills taught at the next grade. The net effect of grade repetition on the acquisition of knowledge is therefore ambiguous.

Jacob and Lefgren (2004) control for this potential bias using a discontinuity in school policy in Chicago. Pupils there take standardized tests at the end of grades 3, 6 and 8. They were promoted if their test score was higher than a minimum score. Regression-discontinuity analysis revealed a small and positive effect of grade repetition on academic achievement after one year. Doing the same with a similar retention policy in Florida, Greene and Winters (2007) find a positive effect of grade repetition in third grade on reading ability after two years.

Grade repetition policies may also have pedagogic effects on non-repeating pupils. Jacob (2005) studies the test-based grade repetition policy in Chicago. An accountability policy has simultaneously been implemented, and made teacher and schools accountable for student achievement. He uses a diff in diff strategy, and shows that the policy increased learning achievement in classes where a lot of pupils where likely to repeat ex-ante, and increases learning achievement for at-risk students. This is consistent with the fact that the grade repetition policy changes the incentives in the class. Unfortunately, it is impossible in this case to disentangle the effect of incentives for pupils caused by grade repetition from the effect of the incentives for teachers caused by the accountability policy.

This paper inquires whether frequent early school dropout in Senegal is in part a consequence of high repetition rates. The question of dropouts induced by grade-repetition is important for at least two reasons. First, education provides the individuals with basic capabilities: dropping out from primary school is *per se* an element of poverty. This is the reason why the Millenium Development Goals include universal completion of primary school. In addition, under imperfect information, school dropout decisions are probably inefficient. Two recent controlled experiments in developing countries

[1] Ministry of Education, Senegal (2005)
[2] www.poledakar.org

(Nguyen, 2008 and Jensen, 2007) have shown that further information on the returns to schooling affect the school enrollment decisions. In the end, both political commitment and economic efficiency make it necessary to fight against endogenous primary school dropout.

The effect of grade repetition on school attainment in developing countries has been extensively studied with control-based identification strategies.[3] However, conditional on a test-score prior to the grade repetition decision, teacher's grade repetition decisions are probably not taken as random.[4] Manacorda (2012) uses the Uruguayan grade retention policy in junior high schools to estimate the effect of grade repetition on dropout. Grade repetition was automatic when a pupil had failed more than 4 subjects. Using a regression discontinuity design based on the number of failed subjects, he finds that grade repetition decreases school achievement by 1 grade on average.[5]

This paper estimates the effect of grade repetition decision on immediate school dropout, and mid-term academic achievement. It controls for the potential correlation between the children's unobservable characteristics and grade repetition with an original instrumental variables strategy. My instrumental strategy is based on the fact that grade repetition probability is strongly non-linear between pupils whose learning achievement are above and below the learning achievement to pass to the next grade. I use a double difference strategy, between learning achievement and between classes with tough or lenient grade repetition policies, to identify the effect of grade repetition on school dropout.

The results reveal a negative effect of grade repetition on the probability of enrollment at school the next year. The estimated effect is fairly high: the estimations show that grade repetition increases the probability of school dropout by approximately 13 percentage points on average. In the mid-term, the results show that grade repetition decreases the probability to reach secondary school by the follow-up survey.

This paper's second result is that schools with tough grade repetition policies are relatively successful: pupils from these schools are on average more likely to be enrolled during the follow-up survey and to start secondary school. These schools are not located in places with particularly favorable observable characteristics. Therefore, this mitigates the conclusion that grade repetition is harmful for school outcomes.

Section 2 presents the dataset used to identify the causal effect of grade repetition on school dropout. Section 3 presents the strategies used here for identifying this effect while brief remarks are made by way of conclusion.

2 The data

This paper uses two datasets, PASEC[6] and EBMS.[7] Both contain detailed information about schooling and are combined here to estimate the effect of grade repetition.

---

[3]See, for example, PASEC (2004) or Glick and Sahn (2009) with the same data than this paper or King, Orazem, and Paterno (2008).

[4]Glick and Sahn (2009) claim that for a given test score, differences in grade repetition decisions depend on variations across schools in test score thresholds for promotion. Grade repetition may nevertheless also depend on pupils' motivation at school for a given test score. Motivation at school can be correlated with parental preferences for education.

[5]The number of failed subjects being an integer, one can still discuss the validity of regression discontinuity designs in this context.

[6]*Programme d'analyse des systèmes éducatifs* set up by CONFEMEN *Conférence des ministres de l'éducation des pays ayant le français en partage.*

[7]*Education et Bien-être des Ménages au Sénégal.* This survey was designed in collaboration by a team from from LEA-INRA, France and from Cornell University, USA. It was implemented in association with the *Centre de Recherche en Economie Appliquée* (Dakar, Senegal). I thank Sylvie Lambert and Christelle Dumas for having made the data available.

2.1 The PASEC panel

The PASEC conducted a panel survey among primary school pupils of 98 Senegalese schools between 1995 and 2000. Twenty second grade students were chosen at random in randomly chosen second grade classes in each school at the beginning of the 1995-1996 school year. They passed learning achievement tests at the end of each school year, and were monitored throughout their school careers (including grade repetitions) until the first of them finished primary school (sixth grade) in 2000. The tests were marked by the PASEC team. Consequently, test scores could not be influenced by teachers.

Whenever a child took a PASEC test in a given school year, the information includes his current grade. Grade repetitions are inferred from this longitudinal information on the school careers. The pupil questionnaire also included some information about living conditions. In particular, the household wealth index used in this paper is based from the PASEC information.

2.2 EBMS Survey

The EBMS survey provides additional information about a sample of PASEC pupils in 2003. It includes some of the pupils from 59 of the schools surveyed between 1995 and 2000. The objective was to resurvey households in each community (village or urban districts) with children who had been in the PASEC panel. Of the 1177 pupils attending the 59 schools surveyed by PASEC, 921 are in EBMS data after deletion of questionable matches. Information was collected about the living conditions and educational levels of the household members. Retrospective data about the school careers of the children surveyed by PASEC meant dropout could be differentiated from other causes of attrition. Consequently, school-leaving dates are known for almost every child re-surveyed (if they had left in 2003). In addition, the EBMS data include the parent's education of the PASEC pupils and retrospective information about living conditions includes self-reported shocks on harvests.

2.3 Aggregate dataset

Both datasets provide reliable retrospective information about enrollment. Together they give enough information to reconstruct most instances of grade repetition. This information is necessary for evaluating the impact of repetition on drop out. Another advantage of the aggregate dataset is that it evaluates the individual learning achievement (test scores), which is a crucial determinant of grade repetition. Definition of all the variables used in this paper can be found in appendix A.

3 Empirical strategy and results

This paper seeks to identify the effect of grade repetition, denoted R_{ik}, on school dropout (enrolment during the next school year is denoted $E_{ik,t+1}$ for child i of group k[8], at date $t+1$), which is the coefficient γ in the equation (1) below. The other determinants of dropout are the PASEC test score S_{ik}, and a vector of covariates X_{ik}.[9]

$$E_{ik,t+1} = \mathbb{1}\left[\beta_{e1}S_{ik} + X_{ik}\beta_{e2} + \gamma R_{ik} + u_{ik} > 0\right] \tag{1}$$

The main difficulty in identifying γ is to control for the potential endogeneity of grade repetition. This paper uses an original instrumental variables strategy to control for the potential correlation between the children's unobservable characteristics (and the measurement error) and grade repetition.

[8] A group is composed of all the observations from the same school, the same year and the same grade. This is an approximation of a class, since in some schools, there are several classes per grade.

[9] This vector includes grade-year dummies, household wealth parents' education, and group mean test score when not included in the model.

3.1 Identification strategy

The identification strategy is based on the widespread idea that a certain learning achievement is required to pass to the next grade. This "target achievement" is denoted t_k. Grade repetition is a non-linear function of the difference between own achievement (measured by the test score S_{ik}) and target achievement. The grade repetition equation writes:

$$R_{ik} = \mathbb{1}\left[\beta_{r1} S_{ik} - \beta_{r2} t_k + f_r(S_{ik} - t_k) + X_{ik}\beta_{r4} + \epsilon_{ik} < 0\right] \tag{2}$$

f_r is a non-parametric function. In practice, it is approximated with dummy variables for intervals of $S_{ik} - t_k$. Unfortunately, the "target achievement" is not observed, and probably varies between classes. The "test score of the last passer" is used in this paper as a proxy for t_k. "Passers" are those peers of a given pupil in a given year who are admitted to the next grade. Among the passers, the pupil with the lowest test score is called the last passer. Her test score is denoted LP_{-ik}:

$$LP_{-ik} = min_{\{j \neq i, R_{jk}=0\}}(S_{jk}) \tag{3}$$

This paper uses the position to the "target achievement" as an instrument for grade repetition, to measure the effect of grade repetition on dropout in model (4):

$$\begin{cases} E_{ik,t+1} & = \mathbb{1}[\quad \beta_{e1} S_{ik} \quad +\beta_{e2} LP_{-ik} \quad \quad \quad +\gamma R_{ik} \quad +X_{ik}\beta_{e4} + u_{ik} > 0] \\ R_{ik} & = \mathbb{1}[\quad \beta_{r1} S_{ik} \quad -\beta_{r2} LP_{-ik} \quad +f_r(S_{ik} - LP_{-ik}) \quad \quad +X_{ik}\beta_{r4} + \epsilon_{ik} < 0] \end{cases} \tag{4}$$

Identification of the parameters in the $E_{ik,t+1}$ equation is assured by the simple non-linearity of the two equations (1)-(2). However, this is likely better handled via an exclusion restriction, i.e. a variable that ideally influences the R_{ik} equation but not the $E_{ik,t+1}$ equation. This is the role of function f_r in model (4). This model controls for the test score S_{ik} and the last passer's score LP_{-ik} and uses the relative position of individual test score to the last passer's score $f_r(S_{ik} - LP_{-ik})$ as an instrument for grade repetition. With a control for S_{ik}, the specification compares pupils with similar learning achievement, and potentially similar background in terms of unobservable characteristics. With a control for LP_{-ik}, the specification includes an homogeneous[10] effect of "target achievement" on grade repetition and dropout. Hence the function $f_r(S_{ik} - LP_{-ik})$ is measured with the differences in the effect of LP_{-ik} on grade repetition between different levels of S_{ik}.

In sum, we measure the double difference between the high-achievement pupils and the low-achievement pupils and between classes with tough and soft grade retention policies. For high-achievement pupils, an increase of LP_{-ik} does not change $f_r(S_{ik} - LP_{-ik})$. Hence it increases the grade repetition probability only through $\beta_{r2} LP_{-ik}$. For low-achievement pupils, an increase of LP_{-ik} can change $f_r(S_{ik} - LP_{-ik})$, and increase much more grade repetition probability. In other words, the paper uses the fact that low-achievement pupils are much more vulnerable to "tough" teachers (teachers with high target achievement).

In the equation of interest, the effect of grade repetition on dropout in model (4) is therefore identified with the same variations of grade repetition probability. The paper is based on the comparison of the correlation between last passer's score and enrollment for low-achievement and high-achievement pupils. This double difference identifies the effect of grade repetition on dropout in model (4).

3.2 Identification questions

School failure or grade repetition? This paper uses a double difference strategy to measure the effect of grade repetition on dropout. The specifications compare the correlation between test

[10] I mean linear in the latent variable.

scores and dropout between classes with different requirements to pass the grade. Our results show that students whose learning achievement lag behind teacher's standards tend to repeat more often, and drop out more often. Model (4) assumes this is entirely due to the effect of grade repetition on dropout.

In this paper, the approximation for target achievement allows to disentangle:

- The link between test score and dropout
- The link between the relative position of individual test score to class mean test score and dropout
- The link between the relative position of individual test score to the last passer's score and dropout

It is possible to make clear only the latter matters in terms of dropout. (This is discussed in the last paragraph of section C.2) This can probably rule out most endogeneity issues.

However, families might directly observe the relative position of their children's test score to the last passer's score. In this case, this position probably affects dropout. The specifications would therefore identify the effect of the inability to meet teacher's requirements on dropout, rather than the effect of grade repetition on dropout. This paper neglects the subtle distinction between the inability to meet teacher's requirement and grade repetition.

Peer effects This paper uses a characteristic of the peers, the last passer's score LP_{-ik}, to measure the target achievement t_k. The extensive literature on peer effects in education economics emphasizes the potential correlation between peers' unobservable characteristics. In addition there may be a "mirror effect", where the outcomes of child i's and i's peer's interact in both directions. In this paper, interactions between peers can probably affect our estimations mostly through LP_{-ik}. Child i's unobservable characteristics may be correlated with LP_{-ik} due the potential correlation between peers' unobservable characteristics or the "mirror effect". In that case, LP_{-ik} can be correlated with dropout because of peer effects.

Child i's unobservable characteristics can theoretically affect LP_{-ik} through 2 mechanisms: through the identity of the last passer, and through the test score of the peers. The second mechanism is probably controlled for. Indeed, all the specifications presented in this paper control for own test score and group mean test score. Hence there is a control for the fact that child i's characteristics help or prevent the last passer to learn and improve her test score. Concerning the first mechanism, grade repetition decisions are probably partly based on a relative evaluation of pupils in Senegal (See the discussion of Table C.1 in the end of this section). When i has favorable unobservable characteristics, i's peers may therefore be more likely to repeat, and LP_{-ik} tends to increase.

In this paper, the main coefficient of interest is not the coefficient on LP_{-ik}, but a double difference. The specifications compare between classes with different LP_{-ik} the correlation between test scores and dropout. When i has favorable unobservable characteristics, i will rarely dropout even in adverse situations (the dropout probability is $\approx 2\%$ in the sample). Hence the link between test scores and dropout is probably tiny, as dropout is unlikely in any situation. When i has favorable unobservable characteristics, LP_{-ik} is probably higher (see previous paragraph). Hence peer effects can decrease the correlation between test scores and dropout when LP_{-ik} is high.

In the data, when LP_{-ik} is high, many pupils are at risk of grade repetition. In these classes, more pupils repeat, and our estimations predict that repeaters are likely to dropout. Hence the link between test scores and dropout is greater. This may be slightly attenuated by peer effects.

Can the coefficients on LP_{-ik} be interpreted as causal? This paper measures the effect of children's unability to meet teacher's requirements in classes. The most challenging way to fight against this is to help children to meet those requirements. It is obviously desirable, but hard to reach. The policy recommendation that could be easily addressed is to change teacher's requirements, or in other words to decrease grade repetition rates for a given learning achievement.

Grade repetition policy might nevertheless have direct effects. Firstly, the threat of grade failure may be an incentive to learn for low-achievement pupils. In addition, a grade repetition may be less discouraging when grade repetition rates are low; and passing the grade may be a signal for high learning ability only when grade repetition rates are high. Besides, a consistent grade repetition policy may decrease variability of learning achievements in the class.

It is therefore desirable to identify the direct effect of grade repetition practices, measured by the direct effect of LP_{-ik} in model (4). In this paper, it is not possible to claim the variations in grade repetition practices (measured by the coefficient on LP_{-ik}) are due to a well identified and exogenous source. However, conditionally on test scores, grade repetition practices seems uncorrelated with individual and location characteristics. In appendix, Table C.1 shows that the proxy for grade repetition practices strongly depends on group mean test score. The coefficient in the linear regression is 1, as if the grade repetition rates and average learning achievement were independent. The proxy for grade repetition practices is not correlated with either observable community characteristics or observable household characteristics. Hence, conditionally on group mean test score, grade repetition practices does not seem to be correlated with the context.

However, all the traits of teacher's pedagogy are likely to be correlated with each other, and to cause dropout. Conditional on test scores, grade repetition practices may be one of these traits, and hence correlated with other determinants of dropouts. To illustrate this, assume dropout rates are lower with tough teachers. It is hard to say whether this is due to their grade repetition practices or to some other trait of tough teachers. In the general case, it is hard to give the sign of the potential bias. So it is not possible to prove the coefficients on the proxy grade repetition practices in the estimations are causal. However, the proxy for grade repetition practices seems uncorrelated with the context: the estimations fail to reject the exogeneity of grade repetition policy.

3.3 Selection issues

Selection on test participation Although children were randomly selected among the second grade pupils of the schools in 1995, attrition and grade repetition meant that the children in the same grade-year were increasingly selected as time elapsed.

There were two causes for attrition in this panel. First, dropouts did not take the PASEC tests. Second, the PASEC team organized the tests and collected the data in each of the schools on a given day in each school year. Children missing school that day or no longer attending the surveyed school were not tested.

Table 1 shows the number of children attending each test in the sample and reveals children often missed a test even though still enrolled. All 921 children were enrolled in school year 1995-1996 although only 817 attended the test. Our regressions are based on test scores, so when a child does not take a test, this year of observation is excluded from the panel. This can pose a selection bias. We do not correct this potential bias in the paper, but its sign is probably predictable.

The paper measures the effect of the difference between pupil's achievement and grade repetition standards on dropout. Dropouts are potentially less likely to take the test: pupils expecting to drop out at the end of the school year may have irregular school attendance. Hence some dropouts due to grade repetition may be selected from the sample because they have anticipated their dropout. The magnitude of the effect of grade repetition on dropout may therefore be underestimated.

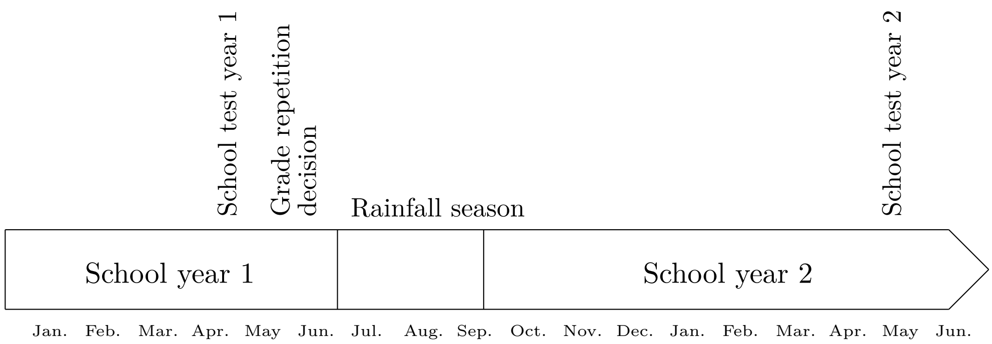
Selection on grade repetition observation Not all grade repetition decisions can be observed in the EBMS-PASEC data. The information for grade repetition is mostly inferred from this longitudinal information on the school careers. The Figure 1 summarizes the timing of the PASEC panel survey.

Table 1: Number of children attending the tests during the panel, by grade and school year

Initial tests (1995)	school year 1995 - 1996	school year 1996 - 1997	school year 1997 - 1998	school year 1998 - 1999	school year 1999 - 2000	
					214	Sixth grade (CM2)
				357	236	Fifth grade (CM1)
			412	204	86	Fourth grade (CE2)
		594	154	53	15	Third grade (CE1)
789	817	102	no test			Second grade (CP)
789	817	696	566	614	551	Total attendance

Note: This table reports the attendance among the 921 children of PASEC sample resurveyed by EBMS

Figure 1: Sequence of the main events during the PASEC panel



It provides us with an exclusion restriction for selection: rainfall shocks. Indeed, the rainfall season happens to be after the end of the school year. Hence rainfalls cannot affect grade repetition decisions.

Information on grade repetition decision at the end of school year t is known if a child took the tests in school year t and school year $t+1$.[11] Grade repetition decisions are not known for the children who dropped out immediately after this decision: if a child dropped out before the tests of school year $t+1$, there is no way of knowing what the repetition decision was at the end of school year t, as grade repetition is inferred from the school career. The structure of the data is therefore summarized in Table 2.

This selection problem makes questionable the identification of the effect of grade repetition deci-

[11] The details and other cases are explained in appendix A.

Table 2: Observation of grade repetition decision

date t	date $t+1$	
Enrolled	Enrolled	} Grade repetition decision is observed
	Enrolled	} Grade repetition decision is not observed: did not take the tests of year t+1
	Drops out	

sions on school dropout: if grade repetition causes dropout, then it causes its own selection. However, this selection bias can probably be corrected. This paper claims it is possible to control for the selection and hence to identify the determinants of grade repetition and the effect of grade repetition on school dropout in model (5):

$$\begin{cases} E_{ik,t+1} & = \mathbb{1}[\quad \beta_{e1}S_{ik} \quad +\beta_{e2}LP_{-ik} \quad +\beta_{e3}Z_s \quad +\gamma R_{ik} \quad +X_{ik}\beta_{e4} + u_{ik} > 0] \\ R_{ik} & = \mathbb{1}[\quad \beta_{r1}S_{ik} \quad -\beta_{r2}LP_{-ik} \quad +f_r(S_{ik} - LP_{-ik}) \quad +X_{ik}\beta_{r4} + \epsilon_{ik} < 0] \\ \\ selection & = \mathbb{1}[\quad \beta_{s1}S_{ik} \quad +\beta_{s2}LP_{-ik} \quad +\beta_{s3}Z_s \quad +\gamma_s R_{ik} \quad +X_{ik}\beta_{s4} + v_{ik} > 0] \end{cases} \tag{5}$$

Z_s is a dummy taking value 1 when the household head reported negative shocks on harvests during the calendar year. It is an exclusion restriction to ensure the identification of the selected equation predicting grade repetition. These shocks are not expected to be a determinant of grade repetition because the rainfall season in Senegal is from July to September, during the school vacations (see Figure 1). Accordingly, grade repetition is known when the rainfall season begins. Theoretically, then, it can be ruled out that teachers might use this information for grade repetitions.

Appendix B.1 proves that in model (5):

- If $(\epsilon_{ik}, u_{ik}, v_{ik})$ is independent of $(S_{ik}, Z_R, Z_s, X_{ik})$
- If $\lambda_2 \neq 0$ and $\beta_{s3} \neq 0$
- Under certain technical assumptions[12]

all the coefficients of model (5) could be identified without any parametric assumption about the distribution of $(\epsilon_{ik}, u_{ik}, v_{ik})$. This is based on a simple intuition: there is an instrument for grade repetition and an instrument for selection. In this case the system of all the probability function derivatives has a single solution. γ and γ_s are not identified by this system, since R_{ik} is binary. However, a simple adaptation of Vytlacil and Yildiz (2007) show the coefficient for the endogenous variable is identified.

Appendix B.2 even shows that under much simplier hypotheses and without Z_s, the sign of the effect of grade repetition on dropout is still identified. The intuition for that is rather simple. Indeed, the derivative of the probability of grade repetition towards $f_r(S_{ik} - LP_{-ik})$ gives the sign of α regardless of selection. Therefore the effect of grade repetition on enrollment is positive if the derivatives of the probability of grade repetition and of the probability of enrollment towards $f_r(S_{ik} - LP_{-ik})$ have the same sign, and negative if they have opposite signs.

This paper does not intend to identify model (5) semiparametrically. All the models in this paper are estimated using a standard maximum likelihood method. However, this result shows that there is enough information to identify the effect grade repetition on dropout in the EBMS-PASEC data without parametric assumption. Hence appendix B.1 suggests the results rely essentially on the information from the data, and not so much on the distributional assumptions of the models.

3.4 Main results

Table 3 gives the estimation of model (5). The model is estimated with a maximum likelihood method, as a "trivariate probit" specification. The distribution of the error terms follow a trivariate normal distribution, simulated with a GHK simulator. The three columns of Table 3 correspond to the model's three equations. The data are pooled for the various grades and years. Each specification includes grade-year dummies in each equation.

[12] Hypotheses about the support of the distribution of $(\epsilon_{ik}, u_{ik}, v_{ik})$ and of the distribution of the observables.

Table 3: Joint estimation of the determinants of grade repetition, selection and school dropouts corresponding to the model (5)

	repetition	$enrolled_{t+1}$	*selection*
	(1)	(2)	(3)
Test score	-.223 (.223)	-.023 (.149)	-.238 (.123)*
LP_{-ik}	.074 (.207)	.379 (.137)***	.305 (.095)***
$S_{ik} - LP_{-ik} < -1$	.695 (.407)*		
$-1 < S_{ik} - LP_{-ik} < -0.75$	1.131 (.347)***		
$-0.75 < S_{ik} - LP_{-ik} < -0.5$	.631 (.262)**		
$-0.5 < S_{ik} - LP_{-ik} < -0.25$	.485 (.193)**		
$-0.25 < S_{ik} - LP_{-ik} < 0$	.269 (.155)*		
$0 < S_{ik} - LP_{-ik} < 0.25$	Ref.		
$0.25 < S_{ik} - LP_{-ik} < 0.5$	-.379 (.159)**		
$0.5 < S_{ik} - LP_{-ik} < 0.75$	-.451 (.192)**		
$0.75 < S_{ik} - LP_{-ik} < 1$	-.401 (.226)*		
$1 < S_{ik} - LP_{-ik} < 1.5$	-.492 (.274)*		
$1.5 < S_{ik} - LP_{-ik}$	-.815 (.451)*		
Negative shock on harvests		.150 (.254)	.569 (.207)***
Grade repetition		-2.259 (.337)***	-2.740 (.463)***
Average marginal effect of grade repetition		*-.138 (.051)****	
Household wealth and Parents' education, Previous year's test score, Group[a] mean test score	Yes	Yes	Yes
Grade*year dummies	Yes	Yes	Yes
Obs.	1818	1818	1818
log likelihood	-1258.516	-1258.516	-1258.516
χ^2 exclusion restrictions	30.290		7.575
corresponding p value	.0008		.006

The model is estimated with a maximum likelihood method, as a "trivariate probit" specification. The distribution of the error terms follow a trivariate normal distribution, simulated with a GHK simulator (25 iterations). Note: ***, ** and * mean respectively that the coefficient is significantly different from 0 at the 1%, 5% and 10% level. The standard deviations of the estimators are corrected for the correlation of the residuals between different observations of the same child.

a: A group is composed of all the observations from the same school, the same year and the same grade. This is an approximation of a class, since in some schools, there are several classes per grade.

Determinants of grade repetition In the grade repetition equation, the effect of test score on enrollment is not significant. This does *not* mean that good pupils have the same probability to repeat grades than the others. This means that the effect of learning achievement on grade repetition is entirely captured by the other coefficients in the regression. In other words, grade repetition probability is not a function of learning achievement *per se*, but a function of the difference between learning achievement and teacher's standards. Similarly, last passer's test score does not seem to affect grade repetition likelihood, which means that its effect is captured by the difference between learning achievement and last passer's score.

The difference between own learning achievement and last passer's score is strongly correlated with grade repetition. The corresponding variables are strongly significant, the χ^2 test for the significance is about 30. The reference is pupils with a test score higher than the last passer's score by 0 to 0.25 points. Pupils with a test score lagging behind the last passer's score by more than 0.5 point have a higher probability of grade repetition, by 0.5 to 1 probit point. Pupils with a test score higher than the last passer's score by more than 0.25 point have a lower probability to repeat the grade by approximately 0.5 probit point. The magnitude of the effect is substantial: on average, changing $S_{ik} - LP_{-ik}$ can change the grade repetition risk from 70% to 10%.

The effect of grade repetition on school dropout Table 3 considers that the only reason why lagging behind teacher's standard affects dropout is the effect of grade repetition on dropout. Using this identification restriction, Table 3 finds a negative effect of grade repetition on dropout. The average marginal effect is -14%, which is impressive given that the average dropout rate is 2% in our sample. The specification simply predicts that all the dropouts have failed to pass their grade. Indeed, in the fitted model, the sample average of the probability of dropout without grade repetition is 0.15% and the sample average of the probability of dropout with grade repetition is 14%.

This result means that all dropouts in the sample dropped out because of grade repetition. This is not completely unlikely. First, all the dropouts observed here took place during primary school, and not between primary school and secondary school. School dropouts during school cycles are the most likely to be due to grade repetition. In addition, during the follow-up EMBM survey in 2003, household members told the reason of their dropout. Most of the dropouts in our sample[13] told why they dropped out. According to them, 60% dropped out because of school failure, and 27% for "other reasons". Only 13% of them gave a precise answer uncompatible with grade repetition. It is therefore credible that dropout within a school cycle is massively associated with grade repetition.

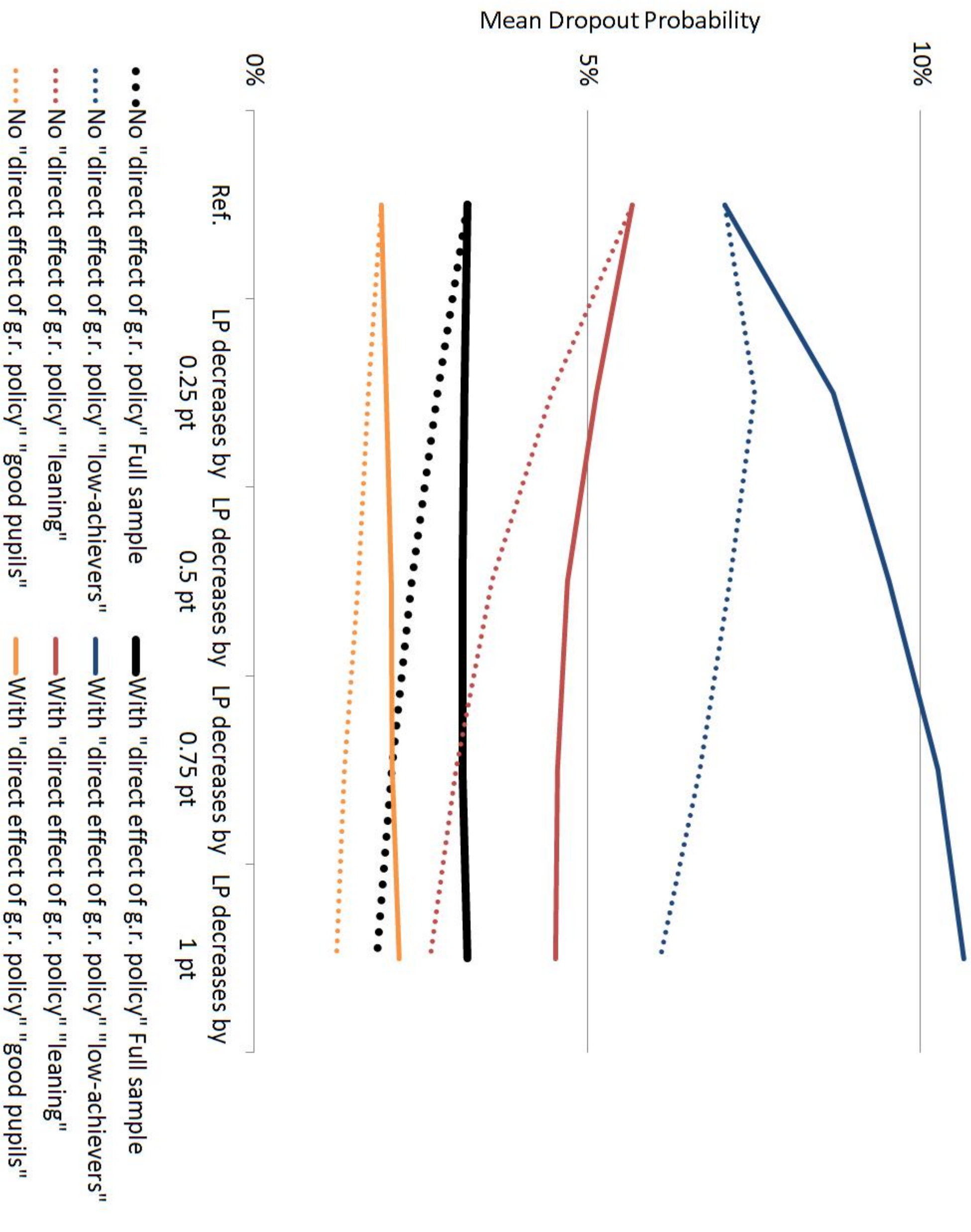
Appendix B.1 suggests the results presented in Table 3 rely on the information from the data, and not so much on the distributional assumptions of the models. So as to make it even clearer, it is possible to identify the reduced-form results corresponding to Table 3. This is done in appendix C.2, Table C.2. This appendix shows there is a link between school dropout and the position relative to the last passer in a reduced-form estimation. This links mirrors the link between the position relative to the last passer and grade repetition (see Figure C.1).

Potential direct effect of grade repetition practices In the enrollment equation, LP_{-ik}, the proxy for grade repetition practices, is positively associated with the probability of being enrolled at school the next year. It is necessary to be cautious with a causal interpretation of this coefficient. However, should this interpretation be valid, this would dramatically change the consequences of a limitation of grade repetition.

Simulations illustrated in Figure 2 illustrate this (The full results are in Table D.1). The thick line shows the dropout probability for the full sample. The thin lines split the sample in 3. The "good pupils" ($0.25 < S_{ik} - LP_{-ik}$) probably pass. The "leaning" pupils might pass if the teacher is more lenient ($-1 < S_{ik} - LP_{-ik} < 0.25$). The "low-achievers" probably repeat anyway ($S_{ik} - LP_{-ik} < -1$).

[13]Those who lived in 2003 in the same household than during the PASEC panel starting in 1995.

Figure 2: Simulations: effect of a decrease of last passer's score on dropout probability



Notes: Simulations based on the estimates of Table 3. Figures in Table D.1.
LP stands for last passer's score. The unit for test scores is the standard deviation of distribution of the test for the year-grade.
"low-achievers": $S_{ik} - LP_{-ik} < -1$
"at risk": $-1 < S_{ik} - LP_{-ik} < 0.25$
"good pupils": $0.25 < S_{ik} - LP_{-ik}$
No "direct effect of grade repetition policy": The simulation assesses the consequences of decreasing LP_{-ik} on grade repetition, and measures the indirect effect on enrollment due to "the effect of grade repetition on enrollment"
With "direct effect of grade repetition policy": The simulations assess the consequences of decreasing LP_{-ik} on grade repetition. It measures the sum of the direct effect of LP_{-ik} measured in the enrollment equation and of the indirect effect on enrollment due to "the effect of grade repetition on enrollment".

With the estimates of Table 3, Figure 2 simulates the consequences of a decrease in LP_{-ik}, the proxy for grade repetition practices, by 0.25, 0.5, 0.75 and 1 standard deviation of the test scores.[14] In the simulations, this represents a fairly substantial decrease of the number of failing pupils, from 26% to 15%.

The dotted lines in Figure 2 show the consequences of lenient grade repetition practices due to the effect of grade repetition on dropout. It means we model the consequences of a decrease of LP_{-ik} (the proxy for grade repetition practices) on grade repetition, and assume the only consequences on dropout are due to the subsequent decrease of grade repetition rates.[14] The dropout probability decreases sharply, from 3.2% to 1.9%. This is especially strong for the "leaning" pupils, whose dropout probability decreases from 5.7% to 2.6%.

The solid lines in Figure 2 show the consequences of a decrease of LP_{-ik} (the proxy for grade repetition practices) if we assume the direct effect of grade repetition practices are measured in Table 3.[14] It shows a totally different picture: dropout rates do *not* decrease with lenient grade repetition practices. Indeed, the direct effect of grade repetition practices measured in Table 3 completely offset by the gains due to fewer grade repetitions. For example, among the "low-achievers", decreasing lenient grade repetition practices does not decrease much grade repetition risk. On the contrary, it increases dropout risk conditional on grade repetition. The dropout risk for the "low-achievers" increases from 7.1% to 10.7%.

Determinants in the selection equation The estimation of selection in model (5) is intended to control for selection bias in the observation of R_{ik}. The determinants of selection may be the determinants of moving or missing school the day of the tests in addition to the determinants of dropout. Accordingly there is no particular interpretation of these coefficients.

Nevertheless, it is necessary to focus on the effect of the negative shocks on harvests, since this variable is the exclusion restriction in the equation for R_{ik}. These shocks positively affect selection: when there is a negative shock, the child is more likely to take the test the next year. Negative shocks on harvests may decrease opportunity costs, so children may be more likely to take the tests when there is a shock. The F-test for the significance of this instrument is 7.2.

Non-linearities in the coefficients of other variables In Table 3, the effect of the difference (between own test score and last passer's score) is non-linear, and we assume that the effect of all the other explanatory is linear (in the latent variable). Hence, our measure of the effect of the difference may catch some other non-linearities of the model. It is nevertheless possible to check this is not the case. In Table C.3 in appendix, the effect of some of the other explanatory variables is treated as non-linear in a reduced-form estimation. The results are very similar to the main reduced-form estimation in Table C.2 (see Figure C.2). In addition, the non-linearities added in this specification are not statistically significant.

4 Mid-term consequences of grade repetition

Section 3 identifies the short-term effect of grade repetition on dropout. This effect is of limited interest if not persistent. For example, grade repetition could push out of school only pupils who intended to stay in school for a single additional school year. The retrospective information on education in the EBMS data can give an insight on this. Precisely, grade repetitions studied in this paper take place between 1996 and 2000, and the EBMS can give information on the retrospective school trajectory as of 2003. This section assesses the mid-term consequences of grade repetition. Table 4 uses the same specifications as Table 3, changing the dependent variable.

[14]Details are given in appendix D.1.

$$\begin{cases} S_{ik,t+\delta} & = \mathbb{1}[\quad \beta_{e1}S_{ik} \quad +\beta_{e2}LP_{-ik} \quad +\beta_{e3}Z_s \quad +\gamma R_{ik} \quad +X_{ik}\beta_{e4} + u_{ik} > 0] \\ R_{ik} & = \mathbb{1}[\quad \beta_{r1}S_{ik} \quad -\beta_{r2}LP_{-ik} \quad +f_r(S_{ik} - LP_{-ik}) \quad +X_{ik}\beta_{r4} + \epsilon_{ik} < 0] \\ selection & = \mathbb{1}[\quad \beta_{s1}S_{ik} \quad +\beta_{s2}LP_{-ik} \quad +\beta_{s3}Z_s \quad +\gamma_s R_{ik} \quad +X_{ik}\beta_{s4} + v_{ik} > 0] \end{cases} \tag{6}$$

In equation (6), $S_{ik,t+\delta}$ is an outcome related to school enrollment of child ik at date $t+\delta$, δ years after the grade repetition decision. Table 4 estimates this model with 8 variables. $enrolled_{t+1}$ is the same variable as in Table 3, and Table 4, column 1 recalls the main quantitative results of Table 3.

Table 4 measures the effect of grade repetition on enrollment 2 years, 3 years and 4 years after the grade repetition decision and finds no effect. It finds no effect of grade repetition on the probability to be enrolled at school during the follow-up survey in 2003.

A grade repetition increases the age for grade. Hence, if grade repetition does not affect dropout date in the mid-term, it may still affect the last grade attended. Table 4 finds that grade repetition decreases the likelihood to reach grade 7 (the first grade of secondary school) and grade 8 until 2003 by 19 and 12 percentage points. However, these estimates may be polluted by completions of grade 7 and 8 posterior to 2003.

The specifications in Table 4 also measure whether our proxy for grade repetition practices (LP_{-ik}) is correlated with long-term achievement (conditionally on grade repetition). Indeed, it is positively correlated with the likelihood to be enrolled in 2003, and with the likelihood to reach grade 5, 6 and 7. Again, there is no strict evidence that this is a causal effect. Should it be a causal effect, the estimated negative effect of lenient grade repetition practices would override the estimated positive effects.

The simulations in Table 4 estimate the consequences of lenient grade repetition policies, with a decrease of LP_{-ik} by 0.25 pt. This represents a decrease in the share of failing students from 26% to 22% (see Table D.1).

The first row of simulations shows the mid-term benefits of lenient grade repetition practices due to the effect of grade repetition. They find that it would increase the probability to be enrolled the next year by 0.5 percentage points, and the probability to reach the first and second grade of secondary school (grade 7 and 8) by 0.7 and 0.4 percentage points.

The second row of simulations in Table 4 shows the consequences of lenient grade repetition practices with a direct effect of grade repetition practices. It assumes this direct effect is measured by the coefficients of LP_{-ik} in the first row of Table 4. The results are in the opposite direction: lenient grade repetition practices would decrease the share of pupils enrolled in 2003 by 1.5 percentage points, and decrease the share of pupils reaching grades 7 and 8 by 1.4 and 1.1 percentage points.

Table C.4 in appendix checks the reduced-form counterpart of these simulations. In other words, it checks the correlation between mid-term outcomes and grade repetition practices (with probit specifications). It finds the same results: conditionally on test scores, tough grade repetition practices (high LP_{-ik}) are correlated with better long-term outcomes.

5 Conclusion

This paper assesses whether grade repetition can deteriorate school outcomes. Its instrumental strategy measures the differences in the link between learning achievement and grade repetition between classes with different requirements to pass to the next grade. This double difference identifies the effect of grade repetition, and shows grade repetition negatively affects school outcomes. Indeed, the probability to be enrolled at school the next year and the probability to start secondary school by the follow-up survey are negatively affected by grade repetition.

However, grade repetition policies can have other consequences than affecting repeating pupils, and it is hard (and probably not achieved in this paper) to find a convincing statistical identification of these

Table 4: Effect of grade repetition on short-term and mid-term outcomes - Full Sample

	$enrolled_{t+1}$	$enrolled_{t+2}$	$enrolled_{t+3}$	$enrolled_{t+4}$	*Still enrolled* (2003)	*Last Grade* > 5	*Last Grade* > 6	*Last Grade* > 7
	(1)[a]	(2)	(3)	(4)	(5)	(6)	(7)	(8)
LP_{-ik}	.379	.128	.087	.135	.249	.176	.296	.246
(enrollment equation)	(.137)***	(.187)	(.148)	(.124)	(.075)***	(.092)*	(.076)***	(.080)***
Grade repetition	-2.259	-.609	.193	.072	-.442	-.423	-.668	-.525
(enrollment equation)	(.337)***	(.880)	(.516)	(.442)	(.270)	(.360)	(.296)**	(.270)*
Average marginal effect of	*-.138*	*-.055*	*.027*	*.013*	*-.156*	*-.118*	*-.189*	*-.119*
repetition (enrollment eq.)	*(.051)****	*(.098)*	*(.073)*	*(.082)*	*(.098)*	*(.107)*	*(.083)***	*(.058)***
Obs.	1818	1449	1449	1449	1789	1777	1777	1777
log-likelihood	-1258.516	-1117.741	-1267.688	-1382.891	-2197.15	-1942.983	-2023.004	-1889.06
χ^2 instruments	30.290	21.903	25.626	20.998	28.988	28.196	27.124	28.886
corresponding p value	.0008	.016	.004	.021	.001	.002	.002	.001
Simulation: LP_{ik}decreases by 0.25 pts, no "direct effect of grade repetition policy"[b]	.005	.002	-.001	-.0006	.006	.005	.007	.004
	(.002)***	(.004)	(.004)	(.004)	(.004)	(.004)	(.003)**	(.002)**
Simulation: LP_{ik}decreases by 0.25 pts, with "direct effect of grade repetition policy"[b]	.0004	-6.06e-06	-.005	-.007	-.015	-.006	-.014	-.011
	(.001)	(.002)	(.003)	(.004)*	(.005)***	(.005)	(.004)***	(.004)***

Notes: The Table reports the results of trivariate specifications as in Table 3 (model (5)), with different dependent variables in the enrollment equation.
***, ** and * mean respectively that the coefficient is significantly different from 0 at the 1%, 5% and 10% level. The standard deviations of the estimators are corrected for the correlation of the residuals between different observations of the same child.
a: Recalls the estimates of Table 3.
b: The simulations assess the consequences of decreasing LP_{-ik} on grade repetition, and measures the indirect effect on enrollment due to "the effect of grade repetition on enrollment"
c: The simulations assess the consequences of decreasing LP_{-ik} on grade repetition. It measures the sum of the direct effect of LP_{-ik} measured in the enrollment equation and of the indirect effect on enrollment due to "the effect of grade repetition on enrollment"

effects. Schools with tough grade repetition policies emphasize similar or better mid-term outcomes than other schools. I did not find any evidence that their environment is particularly favorable to these outcomes. Hence claims that lenient grade repetition policies improve school outcomes are based on incomplete evidences.

References

Filmer, D., Pritchett, L., 2001. Estimating wealth effects without expenditure data - or tears: en application of educational enrollment in states of india. Demography 38, 115–132.

Glick, P., Sahn, D., 2009. Ability, grade repetition and school attainment in Senegal: A panel data analysis, SAGA Working Paper, Cornell University.

Greene, J. P., Winters, M. A., 2007. Revisiting grade retention: an evaluation of florida's test-based retention policy. Journal of School Psychology 2, 319–340.

Jacob, B., Lefgren, L., 2004. Remedial education and student achievement: A regression - discontinuity analysis. Review of economics and statistics 86, 226–244.

Jacob, B. A., 2005. Accountability, incentives and behavior: the impact of high-stakes testing in chicago public schools. Journal of Public Economics 89, 761–796.

Jensen, R., 2008. The perceived returns to education and the demand for schooling, mimeo, Watson Institute for International Studies, Brown University.

Jimerson, S., Carlson, E., Rotert, M., Egeland, B., Sourie, L. A., 1997. A prospective, longitudinal study of the correlates and consequences of early grade retention. Journal of School Psyqhology 35, 3–25.

King, E. M., Orazem, P. F., Paterno, E. M., 2008. Promotion with and without learning: Effects on student dropout. Working paper, World Bank.

Manacorda, M., 2012. The cost of grade retention. Review of Economics and Statistics 94, 596–606.

Manski, C., 1988. Identification of binary response models. Journal of the American Statistical Association 83, 729–738.

Ministry of Education, Senegal, 2005. Situation des indicateurs de l'education 2000 - 2005. Tech. rep., Ministry of education, Senegal.

Nguyen, T., 2008. Information, role models and perceived returns to education: Experimental evidence from Madagascar, mimeo, Massachussets Institute of Technology.

PASEC, 2004. Le redoublement : pratiques et conséquences dans l'enseignement primaire au sénégal. Tech. rep., Conferency of the Ministers of Education of French-speaking countries.

Vytlacil, E., Yildiz, N., 2007. Dummy endogenous variables in weakly separable models. Econometrica 75, 759–779.

A Variables

A.1 Dependant variables

Enrolled is the fact that the child is still enrolled at school in a given year. The information is inferred from the EBMS dataset so as to distinguish attrition in the panel from school dropout.

Last grade is the last grade attended. Grade 6 is the last grade of primary school. The information is inferred from the EBMS dataset.

Repetition is a dummy taking value 1 if the child repeated the grade, and 0 otherwise. Information is from the PASEC panel. In each case, I tried to infer each year whether the child passed at the end of the school year. Table A.2 sums up the various possible cases in the PASEC data and specifies whether anything can be learned about the child's progression. Case 1 is the basic case: the child took all the tests. He repeated after school year 1995 - 1996, and has passed all the subsequent grades. In case 2, the child did not take the tests in 1996 - 1997. The reason why he did not take the test is not reported. Consequently, whether he repeated the second or the third grade is unknown. In case 3, the child dropped out in 1996. Consequently whether he was admitted to third grade after school year 1995 - 1996 is unknown. In case 4, the child is not in the sample after 1997 - 1998, so whether he repeated during the subsequent grades remains unknown. In cases 5 and 6, grade repetitions are not ambiguous: we know the child repeated twice (case 6) or passed twice (case 5) when he was not observed.

A.2 Test scores

Test scores are a proxy for learning achievement at the end of the current school year. In fact the PASEC panel contains school tests at the end of each academic year until the end of the survey.[15] The tests were marked by the PASEC team. Consequently, test scores could not be influenced by teachers. Table 1 reports the number of children taking each test.

The tests were designed to ensure easy comparisons within grade-years. They nevertheless differed between different grades and years of the panel. The test scores have a mean of 0 and a standard deviation of 1 within each grade-year.

Group mean test score A group is composed of all the observations from the same school, the same year and the same grade. This is an approximation of a class, since in some schools, there are several classes per grade.

Last passer's test score is a proxy for the "target achievement", i.e. the learning achievement is required to pass to the next grade. "Passers" are those peers of a given pupil in a given year who are admitted to the next grade. Among the passers, the pupil with the lowest test score is called the last passer. Her test score is denoted LP_{-ik}:

$$LP_{-ik} = min_{\{j \neq i, R_{jk}=0\}}(S_{jk})$$

[15] The second grade classes were not surveyed from 1997 - 1998, so pupils still in this grade at that time were not surveyed until they passed the third grade.

Table A.1: Descriptive statistics

	N	mean	standard deviation	min.	max.
Enrolled next year	1823	.979		0	1
$enrolled_{t+2}$	1454	.959		0	1
$enrolled_{t+3}$	1454	.920		0	1
$enrolled_{t+4}$	1454	.882		0	1
Still enrolled (2003)	1794	.673		0	1
Last Grade > 5	1782	.749		0	1
Last Grade > 6	1782	.437		0	1
Last Grade > 7	1782	.291		0	1
Grade repetition	1823	.148		0	1
Selection (on grade repetition observation)	1823	.867		0	1
Negative shocks on harvests	1823	.087	.308	0	2
Test score	1823	-.055	.953	-3.20	3.34
LP_{-ik}	1818	-.770	.849	-3.20	2.63
Previous year's test score	1823	.00481	1.01	-2.34	3.81
Group[a] mean test score	1823	-.0709	.536	-1.58	1.91
Household wealth	1823	-.591	2.07	-3.12	4.38
Parent's education	1823	2.05	1.47	1	8
Head is not Muslim (Christian or Animist)	1820	.0335		0	1
Ethnic group of the head: Wolof	1813	.398		0	1
Ethnic group of the head: Pulaar-Halpulaar	1813	.187		0	1
Ethnic group of the head: Serere	1813	.250		0	1
Ethnic group of the head: Dioola	1813	.0281		0	1
Ethnic group of the head: Mandingue-Sose	1813	.102		0	1
Ethnic group of the head: Others	1813	.0342		0	1
Community: mean asset index	1823	.0884	1.68	-2.29	3.33
Community: mean education index	1823	2.20	.730	1.19	4
log(Village or city population)	1714	10.0	2.81	5.60	14.6
Community main activity: trade (ref:agri.)	1823	.404		0	1
Community has electricity	1823	.813		0	1
Rural	1823	.535		0	1
Distance to health center	1823	.0730	.260	0	1
Distance to hospital	1823	1.63	1.26	0	3

Notes: Standard deviations are not reported for binary variables.

a: A group is composed of all the observations from the same school, the same year and the same grade. This is an approximation of a class, since in some schools, there are several classes per grade.

Table A.2: Grade attended during the PASEC panel for six imaginary cases

case 1	case 2	case 3	case 4	case 5	case 6	
2	**2**	**2**	**2**	**2**	**2**	school year 1995 - 1996
2	**2,3**	**drop.**	**3**	3	**3**	school year 1996 - 1997
3	**3**		**3,4**	**4**	3	school year 1997 - 1998
4	**4**		**3,4,5**	**5**	**3**	school year 1998 - 1999
5	**5**		**3,4,5,6**	**6**	**4**	school year 1999 - 2000

(When the child did not take the tests, the possible grades are in grey)

Previous year's test scores are a proxy for learning achievement prior to the current school year. During the panel, the children took tests at the end of each school year. In each grade-year of the panel, most of the children had been in the preceding grade the year before. The others had been in the same grade the year before, and were currently repeating their grade. The tests for currently repeating children and others had been different. Yet, some items had been common to both, and those items are used to compare the knowledge of the pupils prior to the current school year. Again, this variable has a mean of 0 and a standard deviation of 1 within each grade-year. This comparison relies exclusively on skills acquired in the preceding grade, since the tests never included items about the skills supposed to be acquired in the following grades.

A.3 Other explanatory variables in main regression

Household wealth is a composite indicator for possession of durable goods, obtained by a principal component analysis (see Filmer and Pritchett, 2001). It is based on children's declarations in 1995, and so avoids reverse causality due to the children's education.

Negative shocks on harvests is a dummy taking value 1 if the head of the household reports a negative shock on harvests during the current calendar year or the next. These shocks are taken into account if the child or his parents were still in the household visited by EBMS in 2003. Otherwise this dummy equals 0, because the child was not really affected by these shocks. (140 cases out of 1823)

Parents' education is the mean of both parents' education. The education of an individual is 1 if the individual never went to school, 2 if the person began but did not finish primary school, 3 if he finished primary school but did not begin secondary school, etc. It takes the highest value, 8, if the individual attended to higher education. If information about the father's education or the mother's education was missing, it is replaced by the mean education of the other adults (aged more than 25 in 1995) in the household.

A.4 Other variables

Community has electricity is from the community questionnaire of the EBMS survey

Community main activity is from the community questionnaire of the EBMS survey

Distance to health center is from the community questionnaire of the EBMS survey

Distance to hospital is from the community questionnaire of the EBMS survey

Religion, ethnic group of the head are taken from the EBMS survey

Village or city population is taken from the EBMS survey for rural area, and from the national census for cities.

B Proofs for the semiparametric identification of model (5)

B.1 model (5)

This section proves that model (5) can be semiparametrically identified.
The model (5) is :

$$\begin{cases} r &= \mathbb{1}(X\beta_r + \gamma_r Z_1 & & +\varepsilon_r > 0) \\ s &= \mathbb{1}(X\beta_s + \gamma_s Z_2 & +\alpha_s r & +\varepsilon_s > 0) \\ e &= \mathbb{1}(X\beta_e + \gamma_e Z_2 & +\alpha_e r & +\varepsilon_e > 0) \end{cases} \tag{B.1}$$

(For simplicity r is *repetition*, s is *selection*, and e is $enrolled_{t+1}$. For the same reason, the equations have been written in a simple form $X\beta + \gamma Z + \varepsilon$.)

Let us recall r is observed if and only if $s = 1$. $f(\varepsilon_r, \varepsilon_s, \varepsilon_e)$ is the distribution function of $(\varepsilon_r, \varepsilon_s, \varepsilon_e)$. Manski (1988) shows that in the one-dimensional binary model case, the parameters are identified by the derivatives of the distribution function. This idea is used to show that all the parameters of model (5) are identified without any parametric assumption on $f(\varepsilon_r, \varepsilon_s, \varepsilon_e)$.

Θ is the support of (X, Z_1, Z_2). Let us make the following assumptions:

1. The distribution of $(\varepsilon_r, \varepsilon_s, \varepsilon_e)$ is independent of (X, Z_1, Z_2).

2. $\gamma_r \neq 0$ and $\gamma_s \neq 0$

3. $\forall j \in \{r, s, e\}$, $\beta_{j1} = 1$

4. $\exists (X_0, Z_{10}, Z_{20}) \in \Theta$ verifying :

   (a) In the neighborhood of (X_0, Z_{10}, Z_{20}), $(X, Z_1, Z_2) \in \Theta$

   (b) $\begin{pmatrix} \frac{d\mathbb{P}(r=1,s=1)}{dZ_1}(X_0, Z_{10}, Z_{20}) & \frac{d\mathbb{P}(r=1,s=1)}{dZ_2}(X_0, Z_{10}, Z_{20}) \\ \frac{d\mathbb{P}(r=0,s=1)}{dZ_1}(X_0, Z_{10}, Z_{20}) & \frac{d\mathbb{P}(r=0,s=1)}{dZ_2}(X_0, Z_{10}, Z_{20}) \end{pmatrix}$ has full rank

   (c) $\forall (X, Z_1, Z_2)$ in the neighborhood of (X_0, Z_{10}, Z_{20}), $0 < f(-X\beta_r - \gamma_r Z_1, -X\beta_s - \gamma_s Z_2, -X\beta_e - \gamma_e Z_2) < \infty$

5. $\exists (a = (X_a, Z_{1a}, Z_{2a}), b = (X_b, Z_{1b}, Z_{2b})) \in \Theta^2$

   (a) $\begin{cases} X_a\beta_r + \gamma_r Z_{1a} = X_b\beta_r + \gamma_r Z_{1b} \\ X_a\beta_s + \gamma_s Z_{2a} + \alpha_s = X_b\beta_s + \gamma_s Z_{2b} \\ X_a\beta_e + \gamma_e Z_{2a} + \alpha_e = X_b\beta_e + \gamma_e Z_{2b} \end{cases}$

   (b) In the neighborhood of a and b, $(X, Z_1, Z_2) \in \Theta$ and $0 < f(-X\beta_r - \gamma_r Z_1, -X\beta_s - \gamma_s Z_2, -X\beta_e - \gamma_e Z_2) < \infty$

Assumption 1 is necessary in Manski (1988) and is still necessary here. It ensures that the derivatives of the probability functions with respect to X, Z_1 or Z_2 are not caused by variations of $f(\varepsilon_r, \varepsilon_s, \varepsilon_e)$.

Assumption 2 ensures the instruments have a real causal effect on the endogenous variables.

In model (5), only the signs of the latent variables ($X\beta_r + \gamma_r Z_1 + \varepsilon_r$, $X\beta_s + \gamma_s Z_2 + \alpha_s r + \varepsilon_s$ and $X\beta_e + \gamma_e Z_2 + \alpha_e r + \varepsilon_e$) are observed. Accordingly, the parameters are identified up to the scale of the parameter vector. Assumption 3 easily fixes that scale.

Assumption 4a ensures it is possible to compute the derivatives of the probability functions with the data since the points in the neighborhood of (X_0, Z_0) are in the support of (X, Z). It is certainly possible to extend the identification result when X contains some binary variables.

Assumption 4b ensures some of the derivatives of the probability functions are not all zero and that they are not collinear, so that the systems are fully identified in (X_0, Z_{10}, Z_{20}).

Assumption 4c ensures the other derivatives of the probability functions with respect to the covariates are not null in (X_0, Z_{10}, Z_{20}).

Assumption 5 ensures the support Θ is large enough to contain a pair of points with similar characteristics for s and e when the former has $r = 1$ and the latter has $r = 0$.

This proof has three steps: first, it is shown that the coefficients β and γ of the first two equations of model (5) are identified, second, it is shown that the coefficients β and γ of the last equation are identified, and finally, it is shown that the α are identified.

- **Identification of the first two equations of the model**

  Let us compute the derivatives of $\mathbb{P}(r = 1, s = 1|X, Z_1, Z_2)$. This probability and its derivatives can be estimated with the data in (X_0, Z_{10}, Z_{20}) if assumption 4a is true:

$$
\begin{aligned}
P^{(11)} &= \mathbb{P}(r = 1, s = 1|X, Z_1, Z_2) \\
&= \int_{-X\beta_r - \gamma_r Z_1}^{\infty} \int_{-X\beta_s - \gamma_s Z_2 - \alpha_s}^{\infty} \int_{\mathbb{R}} f(\varepsilon_r, \varepsilon_s, \varepsilon_e) d\varepsilon_r d\varepsilon_s d\varepsilon_e \\
&= F^{(11)}(-X\beta_r - \gamma_r Z_1, -X\beta_s - \gamma_s Z_2 - \alpha_s)
\end{aligned}
$$

  We note $F_1'^{(11)}$ and $F_2'^{(11)}$ the derivatives of $F^{(11)}$ with respect to its two arguments. The derivatives are:

$$
\frac{dP^{(11)}}{dX_1} = F_1'^{(11)} + F_2'^{(11)} \tag{B.2}
$$

$$
\frac{dP^{(11)}}{dX_i} = \beta_{ri} F_1'^{(11)} + \beta_{si} F_2'^{(11)} \; (\forall i \in \{1..K\}) \tag{B.3}
$$

$$
\frac{dP^{(11)}}{dZ_1} = \gamma_r F_1'^{(11)} \tag{B.4}
$$

$$
\frac{dP^{(11)}}{dZ_2} = \gamma_s F_2'^{(11)} \tag{B.5}
$$

  This is clearly not sufficient to identify β and γ. In fact, these four equations contain six unknown parameters, since $F_1'^{(11)}$ and $F_2'^{(11)}$ are unknown. So the derivatives of $\mathbb{P}(r = 0, o = 1|X, Z_1, Z_2)$ are necessary to identify γ and β.

$$\begin{aligned}
P^{(01)} &= \mathbb{P}(r=0, s=1|X, Z_1, Z_2) \\
&= \int_{-\infty}^{X\beta_r-\gamma_r Z_1} \int_{-X\beta_s-\gamma_s Z_2}^{\infty} \int_{\mathbb{R}} f(\varepsilon_r, \varepsilon_s, \varepsilon_e) d\varepsilon_r d\varepsilon_s d\varepsilon_e \\
&= F^{(01)}(-X\beta_r - \gamma_r Z_1, -X\beta_s - \gamma_s Z_2)
\end{aligned}$$

We note $F_1^{\prime(01)}$ and $F_2^{\prime(01)}$ the derivatives of $F^{(01)}$ towards its two arguments.

$$\frac{dP^{(01)}}{dX_1} = F_1^{\prime(01)} + F_2^{\prime(01)} \tag{B.6}$$

$$\frac{dP^{(01)}}{dX_i} = \beta_{ri} F_1^{\prime(01)} + \beta_{si} F_2^{\prime(01)} \tag{B.7}$$

$$\frac{dP^{(01)}}{dZ_1} = \gamma_r F_1^{\prime(01)} \tag{B.8}$$

$$\frac{dP^{(01)}}{dZ_2} = \gamma_s F_2^{\prime(01)} \tag{B.9}$$

From equation (B.2) rearranged with (B.4) and (B.5), and (B.6) rearranged with (B.8) and (B.9), we get the two equations system:

$$\left\{ \begin{array}{lcl} \frac{dP^{(11)}}{dX_1} & = & \frac{1}{\gamma_r}\frac{dP^{(11)}}{dZ_1} + \frac{1}{\gamma_s}\frac{dP^{(11)}}{dZ_2} \\ \frac{dP^{(01)}}{dX_1} & = & \frac{1}{\gamma_r}\frac{dP^{(01)}}{dZ_1} + \frac{1}{\gamma_s}\frac{dP^{(01)}}{dZ_2} \end{array} \right.$$

Under assumptions 4b and 2, this identifies γ_s and γ_r. We can then easily compute $F_1^{\prime(11)}$, $F_2^{\prime(11)}$, $F_1^{\prime(01)}$ and $F_2^{\prime(01)}$ with (B.4), (B.5), (B.8) and (B.9). The system:

$$\left\{ \begin{array}{lcl} \frac{dP^{(11)}}{dX_i} & = & \beta_{ri} F_1^{\prime(11)} + \beta_{si} F_2^{\prime(11)} \\ \frac{dP^{(01)}}{dX_i} & = & \beta_{ri} F_1^{\prime(01)} + \beta_{si} F_2^{\prime(01)} \end{array} \right.$$

identifies β_{ri} and β_{si}. In fact, assumption 2 ensures that $\begin{pmatrix} \gamma_r F_1^{\prime(11)} & \gamma_r F_1^{\prime(01)} \\ \gamma_s F_2^{\prime(11)} & \gamma_s F_2^{\prime(01)} \end{pmatrix}$ has full rank, that $\begin{pmatrix} F_1^{\prime(11)} & F_1^{\prime(01)} \\ F_2^{\prime(11)} & F_2^{\prime(01)} \end{pmatrix}$ has full rank.

- **Identification of the third equation**

  We compute the derivatives of $\mathbb{P}(e=1|X, Z_1, Z_2)$:

$$\begin{aligned}
P^{(1)} &= \mathbb{P}(e=1|X, Z_1, Z_2) \\
&= \int_{-X\beta_r-\gamma_r Z_1}^{\infty} \int_{\mathbb{R}} \int_{-X\beta_e-\gamma_e Z_2-\alpha_e}^{\infty} f(\varepsilon_r, \varepsilon_s, \varepsilon_e) d\varepsilon_r d\varepsilon_s d\varepsilon_e \\
&+ \int_{-\infty}^{X\beta_r-\gamma_r Z_1} \int_{\mathbb{R}} \int_{-X\beta_e-\gamma_e Z_2}^{\infty} f(\varepsilon_r, \varepsilon_s, \varepsilon_e) d\varepsilon_r d\varepsilon_s d\varepsilon_e \\
&= F^{(1)}(-X\beta_r - \gamma_r Z_1, -X\beta_e - \gamma_e Z_2, -\alpha_e)
\end{aligned}$$

We call $F_1'^{(1)}$, $F_2'^{(1)}$ and $F_3'^{(1)}$ the derivatives of $F^{(1)}$ with respect to its arguments. We compute the derivatives of $P^{(1)}$:

$$\begin{aligned}
\frac{dP^{(1)}}{dX_1} &= F_1'^{(1)} + F_2'^{(1)} && \text{(B.10)}\\
\frac{dP^{(1)}}{dX_i} &= \beta_{ri} F_1'^{(1)} + \beta_{si} F_2'^{(1)} && \text{(B.11)}\\
\frac{dP^{(1)}}{dZ_1} &= \gamma_r F_1'^{(1)} && \text{(B.12)}\\
\frac{dP^{(1)}}{dZ_2} &= \gamma_e F_2'^{(1)} && \text{(B.13)}
\end{aligned}$$

γ_r is known, so that $F_1'^{(1)}$ can be easily computed with (B.12). It is then possible to compute $F_2'^{(1)}$ with (B.10). Under assumption 4c, $F_2'^{(1)}$ is not null in $(X, Z_1, Z_2) \in \Theta$. That is why γ_e is identified by (B.13). Knowledge of β_{ri}, $F_1'^{(1)}$ and $F_2'^{(1)}$ identifies β_{si} in (B.11).

- **Identification of α_s.**

  Adapting Vytlacil and Yildiz (2007), it is easy to show that:

  If $\exists\,((X_a, Z_{1a}, Z_{2a}), (X_b, Z_{1b}, Z_{2b}), (X_c, Z_{1c}, Z_{2c}), (X_d, Z_{1d}, Z_{2d})) \in \Theta^4$ so that[16]

$$\left\{\begin{array}{l}
X_a\beta_r + \gamma_r Z_{1a} = X_b\beta_r + \gamma_r Z_{1b} = \kappa_{r1}\\
X_c\beta_r + \gamma_r Z_{1c} = X_d\beta_r + \gamma_r Z_{1d} = \kappa_{r2}\\
X_a\beta_s + \gamma_s Z_{2c} = X_c\beta_s + \gamma_s Z_{2c} = \kappa_{s1}\\
X_b\beta_s + \gamma_s Z_{2b} = X_d\beta_s + \gamma_s Z_{2d} = \kappa_{s2}
\end{array}\right. \Leftrightarrow \left\{\begin{array}{l}
\mathbb{P}(r|a) = \mathbb{P}(r|b)\\
\mathbb{P}(r|c) = \mathbb{P}(r|d)\\
\hat{\mathbb{P}}(s|a) = \hat{\mathbb{P}}(s|c)\\
\hat{\mathbb{P}}(s|b) = \hat{\mathbb{P}}(s|d)
\end{array}\right. \qquad \text{(B.14)}$$

  $0 < f(\varepsilon_r, \varepsilon_s, \varepsilon_e) < \infty$ in the neighborhood of a and of b and $\kappa_{r1} \neq \kappa_{r2}$.

  Then

$$\left(\begin{array}{rl}
 & \mathbb{P}(r=1, s=1|a) - \mathbb{P}(r=1, s=1|c)\\
= - & [\mathbb{P}(r=0, s=1|b) - \mathbb{P}(r=0, s=1|d)]
\end{array}\right) \Rightarrow \kappa_{s1} + \alpha_s = \kappa_{s2} \qquad \text{(B.15)}$$

  It is obvious that the converse is true. In fact, if $\kappa_{s1} + \alpha_s = \kappa_{s2}$, then:

$$\begin{aligned}
\mathbb{P}(r=1, s=1|a) + \mathbb{P}(r=0, s=1|b) &= \hat{\mathbb{P}}(s=1|b)\\
\mathbb{P}(r=1, s=1|c) + \mathbb{P}(r=0, s=1|d) &= \hat{\mathbb{P}}(s=1|d)
\end{aligned}$$

  because

[16] $\hat{\mathbb{P}}$ means that the probability is net of the effect of r on o.

$$
\begin{aligned}
\mathbb{P}(r=1,s=1|a)+\mathbb{P}(r=0,s=1|b) &= \int_{-\infty}^{\kappa_{r1}}\int_{-\kappa_{s1}-\alpha_s}^{\infty}\int_{\mathbb{R}} f(\varepsilon_r,\varepsilon_s,\varepsilon_e)d\varepsilon_r d\varepsilon_s d\varepsilon_e \\
&+ \int_{-\kappa_{r1}}^{\infty}\int_{-\kappa_{s2}}^{\infty}\int_{\mathbb{R}} f(\varepsilon_r,\varepsilon_s,\varepsilon_e)d\varepsilon_r d\varepsilon_s d\varepsilon_e \\
&= \int_{\mathbb{R}}\int_{-\kappa_{s2}}^{\infty}\int_{\mathbb{R}} f(\varepsilon_r,\varepsilon_s,\varepsilon_e)d\varepsilon_r d\varepsilon_s d\varepsilon_e \\
&= \hat{\mathbb{P}}(s=1|b)
\end{aligned}
$$

(B.14) ensures that $\hat{\mathbb{P}}(s=1|b)=\hat{\mathbb{P}}(s=1|d)$. Finally:

$$
\begin{aligned}
\mathbb{P}(r=1,s=1|a)+\mathbb{P}(r=0,s=1|b) &= \mathbb{P}(r=1,s=1|c)+\mathbb{P}(r=0,s=1|d) \\
\Leftrightarrow \mathbb{P}(r=1,s=1|a)-\mathbb{P}(r=1,s=1|c) &= -[\mathbb{P}(r=0,s=1|b)-\mathbb{P}(r=0,s=1|d)]
\end{aligned}
$$

Proof of equation (B.15):

We write the probabilities:

$$
\begin{aligned}
\mathbb{P}(r=1,s=1|\kappa_r,\kappa_s) &= \int_{-\kappa_r}^{\infty}\int_{-\kappa_s-\alpha_s}^{\infty}\int_{\mathbb{R}} f(\varepsilon_r,\varepsilon_s,\varepsilon_e)d\varepsilon_r d\varepsilon_s d\varepsilon_e \\
\mathbb{P}(r=0,s=1|\kappa_r,\kappa_s) &= \int_{-\infty}^{-\kappa_r}\int_{-\kappa_s}^{\infty}\int_{\mathbb{R}} f(\varepsilon_r,\varepsilon_s,\varepsilon_e)d\varepsilon_r d\varepsilon_s d\varepsilon_e
\end{aligned}
$$

Then we can easily compute the differences of (B.15):

$$
\begin{aligned}
\mathbb{P}(r=1,s=1|a)-\mathbb{P}(r=1,s=1|c) &= \int_{-\kappa_{r1}}^{-\kappa_{r2}}\int_{-\kappa_{s1}-\alpha_s}^{\infty}\int_{\mathbb{R}} f(\varepsilon_r,\varepsilon_s,\varepsilon_e)d\varepsilon_r d\varepsilon_s d\varepsilon_e \\
\mathbb{P}(r=0,s=1|b)-\mathbb{P}(r=0,s=1|d) &= \int_{-\kappa_{r2}}^{-\kappa_{r1}}\int_{-\kappa_{s2}}^{\infty}\int_{\mathbb{R}} f(\varepsilon_r,\varepsilon_s,\varepsilon_e)d\varepsilon_r d\varepsilon_s d\varepsilon_e
\end{aligned}
$$

We can now rewrite the first term of (B.15):

$$
\begin{aligned}
& \mathbb{P}(r=1,s=1|a)-\mathbb{P}(r=1,s=1|c) = -[\mathbb{P}(r=0,s=1|b)-\mathbb{P}(r=0,s=1|d)] \\
\Leftrightarrow\ & \int_{-\kappa_{r1}}^{-\kappa_{r2}}\left(\int_{-\kappa_{s1}-\alpha_s}^{\infty}\int_{\mathbb{R}} f(\varepsilon_r,\varepsilon_s,\varepsilon_e)d\varepsilon_s d\varepsilon_e - \int_{-\kappa_{s2}}^{\infty}\int_{\mathbb{R}} f(\varepsilon_r,\varepsilon_s,\varepsilon_e)d\varepsilon_s d\varepsilon_e\right)d\varepsilon_r = 0 \\
\Leftrightarrow\ & \int_{-\kappa_{r1}}^{-\kappa_{r2}}\int_{\mathbb{R}}\left(\int_{-\kappa_{s1}-\alpha_s}^{-\kappa_{s2}} f(\varepsilon_r,\varepsilon_s,\varepsilon_e)d\varepsilon_s\right)d\varepsilon_r d\varepsilon_e = 0
\end{aligned}
$$

$f(\varepsilon_r,\varepsilon_s,\varepsilon_e)>0$ in the neighborhood of a and b. As a consequence, it is strictly positive in a subset of the integration interval with a strictly positive Lebesgue measure if $\kappa_{s1}+\alpha_s\neq\kappa_{s2}$. So $\kappa_{s1}+\alpha_s=\kappa_{s2}$, QED.

Assumption 5 ensures that some points verifying (B.14) and (B.15) exist in Θ. In fact, points a and b in assumption 5 verify (B.14) and the second term of (B.15). c can be found in the neighborhood of a and d in the neighborhood of b: the hyperplanes $\hat{\mathbb{P}}(s|(X,Z_1,Z_2) = \hat{\mathbb{P}}(s|a)$ and $\hat{\mathbb{P}}(s|(X,Z_1,Z_2) = \hat{\mathbb{P}}(s|b)$ necessarily contain pairs of points that have the same $P(r)$, since $P(r|a) = P(r|b)$.

These points can be recognized because the validity of (B.14) and

$$\mathbb{P}(r=1,s=1|a) - \mathbb{P}(r=1,s=1|b) = -[\mathbb{P}(r=0,s=1|c) - \mathbb{P}(r=0,s=1|d)]$$

can be evaluated with the data and previous results.

- **Identification of α_e.**

If $\exists\,((X_a,Z_{1a},Z_{2a}),(X_b,Z_{1b},Z_{2b}),(X_c,Z_{1c},Z_{2c}),(X_d,Z_{1d},Z_{2d})) \in \Theta^4$ so that

$$\left\{\begin{array}{lcl} X_a\beta_r + \gamma_r Z_{1a} &=& X_b\beta_r + \gamma_r Z_{1b} = \kappa_{r1} \\ X_c\beta_r + \gamma_r Z_{1c} &=& X_d\beta_r + \gamma_r Z_{1d} = \kappa_{r2} \\ X_a\beta_s + \gamma_s Z_{1a} &=& X_c\beta_s + \gamma_s Z_{1c} = \kappa_{s1} \\ X_b\beta_s + \gamma_s Z_{1b} &=& X_d\beta_s + \gamma_s Z_{1d} = \kappa_{s2} \\ X_a\beta_e + \gamma_e Z_{2a} &=& X_c\beta_e + \gamma_e Z_{2c} = \kappa_{e1} \\ X_b\beta_e + \gamma_e Z_{2b} &=& X_d\beta_e + \gamma_e Z_{2d} = \kappa_{e2} \end{array}\right. \Leftrightarrow \left\{\begin{array}{lcl} \mathbb{P}(r|a) &=& \mathbb{P}(r|b) \\ \mathbb{P}(r|c) &=& \mathbb{P}(r|d) \\ \hat{\mathbb{P}}(s|a) &=& \hat{\mathbb{P}}(s|c) \\ \hat{\mathbb{P}}(s|b) &=& \hat{\mathbb{P}}(s|d) \\ \hat{\mathbb{P}}(e|a) &=& \hat{\mathbb{P}}(e|c) \\ \hat{\mathbb{P}}(e|b) &=& \hat{\mathbb{P}}(e|d) \end{array}\right. \tag{B.16}$$

and $\left\{\begin{array}{l} \kappa_{r1} \neq \kappa_{r2} \\ \kappa_{s1} + \alpha_s = \kappa_{r2} \end{array}\right.$ and $0 < f(\varepsilon_r,\varepsilon_s,\varepsilon_e) < \infty$ in the neighborhood of a and of b.

Then

$$\left(\begin{array}{ll} & \mathbb{P}(r=1,s=1,e=1|a) - \mathbb{P}(r=1,s=1,e=1|c) \\ = - & [\mathbb{P}(r=0,s=1,e=1|b) - \mathbb{P}(r=0,s=1,e=1|d)] \end{array}\right) \Rightarrow \kappa_{e1} + \alpha_e = \kappa_{e2} \tag{B.17}$$

For the same reason as for the identification of α_s, the converse of B.17 is true. In fact, if $\kappa_{e1} + \alpha_e = \kappa_{e2}$, then:

$$\begin{array}{lcl} \mathbb{P}(r=1,s=1,e=1|a) + \mathbb{P}(r=0,s=1,e=1|b) &=& \hat{\mathbb{P}}(s=1,c=1|b) \\ \mathbb{P}(r=1,s=1,e=1|c) + \mathbb{P}(r=0,s=1,e=1|d) &=& \hat{\mathbb{P}}(s=1,c=1|d) \end{array}$$

Proof of equation (B.17):

We write the probabilities:

$$\begin{array}{lcl} \mathbb{P}(r=1,s=1,e=1|a) &=& \int_{-\kappa_{r1}}^{\infty}\int_{-\kappa_{s1}-\alpha_s}^{\infty}\int_{-\kappa_{e1}-\alpha_e}^{\infty} f(\varepsilon_r,\varepsilon_s,\varepsilon_e)d\varepsilon_r d\varepsilon_s d\varepsilon_e \\ \mathbb{P}(r=1,s=1,e=1|c) &=& \int_{-\kappa_{r2}}^{\infty}\int_{-\kappa_{s1}-\alpha_s}^{\infty}\int_{-\kappa_{e1}-\alpha_e}^{\infty} f(\varepsilon_r,\varepsilon_s,\varepsilon_e)d\varepsilon_r d\varepsilon_s d\varepsilon_e \\ \mathbb{P}(r=0,s=1,e=1|b) &=& \int_{-\infty}^{-\kappa_{r1}}\int_{-\kappa_{s2}}^{\infty}\int_{-\kappa_{e2}}^{\infty} f(\varepsilon_r,\varepsilon_s,\varepsilon_e)d\varepsilon_r d\varepsilon_s d\varepsilon_e \\ \mathbb{P}(r=0,s=1,e=1|d) &=& \int_{-\infty}^{-\kappa_{r2}}\int_{-\kappa_{s2}}^{\infty}\int_{-\kappa_{e2}}^{\infty} f(\varepsilon_r,\varepsilon_s,\varepsilon_e)d\varepsilon_r d\varepsilon_s d\varepsilon_e \end{array}$$

Then we can easily compute the differences of (B.17):

$$
\begin{aligned}
& \mathbb{P}(r=1,s=1,e=1|a) - \mathbb{P}(r=1,s=1,e=1|c) \\
= & \int_{-\kappa_{r1}}^{-\kappa_{r2}} \int_{-\kappa_{s1}-\alpha_s}^{\infty} \int_{-\kappa_{e1}-\alpha_e}^{\infty} f(\varepsilon_r,\varepsilon_s,\varepsilon_e) d\varepsilon_r d\varepsilon_s d\varepsilon_e \\
& \mathbb{P}(r=0,s=1,e=1|b) - \mathbb{P}(r=0,s=1,e=1|d) \\
= & \int_{-\kappa_{r2}}^{-\kappa_{r1}} \int_{-\kappa_{s2}}^{\infty} \int_{-\kappa_{e2}}^{\infty} f(\varepsilon_r,\varepsilon_s,\varepsilon_e) d\varepsilon_r d\varepsilon_s d\varepsilon_e
\end{aligned}
$$

We can now rewrite the first term of (B.15):

$$
\begin{aligned}
& \mathbb{P}(r=1,s=1|a) - \mathbb{P}(r=1,s=1|c) = -[\mathbb{P}(r=0,s=1|b) - \mathbb{P}(r=0,s=1|d)] \\
\Leftrightarrow & \int_{-\kappa_{r1}}^{-\kappa_{r2}} \int_{-\kappa_{s2}}^{\infty} \int_{-\kappa_{e1}-\alpha_e}^{-\kappa_{e2}} f(\varepsilon_r,\varepsilon_s,\varepsilon_e) d\varepsilon_r d\varepsilon_s d\varepsilon_e = 0
\end{aligned}
$$

$f(\varepsilon_r,\varepsilon_s,\varepsilon_e) > 0$ in the neighborhood of any point of Θ (assumption 4c). As a consequence, it is strictly positive in a subset of the integration interval with a strictly positive Lebesgue measure if $\kappa_{e1} + \alpha_e \neq \kappa_{e2}$. That is why $\kappa_{e1} + \alpha_s = \kappa_{e2}$. Assumption 5 ensures that those points exist, so α_e can be identified.

B.2 Model (5) without Z_2

This appendix proves that Z_2 is unnecessary for identifying the sign of α_e. Accordingly, it is theoretically not necessary to control for selection to identify the sign of α_e semiparametrically. The corresponding model is:

$$
\left\{
\begin{array}{lcllll}
r & = & \mathbb{1}(X\beta_r + \gamma_r Z & & +\varepsilon_r > 0) \\
s & = & \mathbb{1}(X\beta_s & +\alpha_s r & +\varepsilon_s > 0) \\
e & = & \mathbb{1}(X\beta_e & +\alpha_e r & +\varepsilon_e > 0)
\end{array}
\right. \tag{B.18}
$$

(For simplicity r is *repetition*, s is *selection*, and e is *enrolled*$_{t+1}$. For the same reason, the equations have been written in a simple form $X\beta + \gamma Z + \varepsilon$)

Let us recall that r is observed if and only if $s = 1$. $f(\varepsilon_r,\varepsilon_s,\varepsilon_e)$ is the distribution function of $(\varepsilon_r,\varepsilon_s,\varepsilon_e)$. Manski (1988) shows that in the one-dimensional binary model case, the parameters are identified by the derivatives of the probability function of the dependent variable. This idea is used to show that the sign of α_e is identified in model (B.18) without any parametric assumption on $f(\varepsilon_r,\varepsilon_s,\varepsilon_e)$. Θ is the support of (X,Z). We make the following assumptions:

1. The distribution of $(\varepsilon_r,\varepsilon_s,\varepsilon_e)$ is independent of (X,Z).
2. $\gamma_r \neq 0$
3. $\exists (X_0, Z_0) \in \Theta$ verifying :
   (a) In the neighborhood of (X_0, Z_0), $(X,Z) \in \Theta$
   (b) $\int_{\mathbb{R}} \int_{\mathbb{R}} f(-X_0\beta_r - \gamma_r Z_0, \varepsilon_s, \varepsilon_e) d\varepsilon_s d\varepsilon_e < \infty$

(c) $f(\varepsilon_r, \varepsilon_s, \varepsilon_e) > 0$ in the neighborhood of $(-X_0\beta_r - \gamma_r Z_0, -X_0\beta_s - \alpha_s, -X_0\beta_s - \alpha_e)$, called Γ

Assumption 1 is necessary in Manski (1988) and is still necessary in this case. It ensures that the derivatives of the probability functions with respect to X or Z are not caused by variations of $f(\varepsilon_r, \varepsilon_s, \varepsilon_e)$.

Assumption 2 ensures that the instrument has a causal effect on r.

Assumption 3a ensures that it is possible to compute the derivatives of the probability functions with the data since the points in the neighborhood of (X_0, Z_0) are in the support of (X, Z). It is certainly possible to extend the identification result in the case where X contains some binary variables.

Assumption 3b ensures that the density of ε_r in $-X_0\beta_r - \gamma_r Z_0$ is finite, so that the derivatives of the probabilities with respect to Z are finite.

Assumption 3c ensures that the derivatives of the probability functions with respect to Z are not null.

– **Proof that the sign of γ_r is identified**

We write $\mathbb{P}(r = 1, s = 1, e = 1|X, Z)$, which is identified by the data in (X_0, Z_0) because of assumption 3a:

$$\begin{aligned}
\mathbb{P}(r = 1, s = 1, e = 1|X, Z) &= \int_{-X\beta_r-\gamma_r Z}^{\infty} \int_{-X\beta_s-\alpha_s}^{\infty} \int_{-X\beta_e-\alpha_e}^{\infty} f(\varepsilon_r, \varepsilon_s, \varepsilon_e) d\varepsilon_r d\varepsilon_s d\varepsilon_e \\
\Rightarrow d\mathbb{P}(r = 1, s = 1, s = 1|X, Z)/dZ &= \gamma_r \int_{-X\beta_s-\alpha_s}^{\infty} \int_{-X\beta_e-\alpha_e}^{\infty} f(-X\beta_r - \gamma_r Z, \varepsilon_s, \varepsilon_e) d\varepsilon_s d\varepsilon_e \\
0 &\leq \int_{-X\beta_s-\alpha_s}^{\infty} \int_{-X\beta_e-\alpha_e}^{\infty} f(-X\beta_r - \gamma_r Z, \varepsilon_s, \varepsilon_e) d\varepsilon_s d\varepsilon_e
\end{aligned}$$

Assumption 3b ensures that:

$$\int_{-X_0\beta_s-\alpha_s}^{\infty} \int_{-X_0\beta_e-\alpha_e}^{\infty} f(-X_0\beta_r - \gamma_r Z_0, \varepsilon_s, \varepsilon_e) d\varepsilon_s d\varepsilon_e \leq \int_{\mathbb{R}} \int_{\mathbb{R}} f(-X_0\beta_r - \gamma_r Z_0, \varepsilon_s, \varepsilon_e) d\varepsilon_s d\varepsilon_e < \infty$$

And assumption 3c ensures that:

$$\begin{aligned}
&\int_{[-X_0\beta_s-\alpha_s,\infty]\times[-X_0\beta_e-\alpha_e,\infty]} f(-X_0\beta_r - \gamma_r Z_0, \varepsilon_s, \varepsilon_e) d\varepsilon_s d\varepsilon_e \\
\geq &\int_{([-X_0\beta_s-\alpha_s,\infty]\times[-X_0\beta_e-\alpha_e,\infty])\cap\Gamma} f(-X_0\beta_r - \gamma_r Z_0, \varepsilon_s, \varepsilon_e) d\varepsilon_s d\varepsilon_e > 0
\end{aligned}$$

That is why

$$0 < \int_{-X_0\beta_s-\alpha_s}^{\infty} \int_{-X_0\beta_e-\alpha_e}^{\infty} f(-X_0\beta_r - \gamma_r Z_0, \varepsilon_s, \varepsilon_e) d\varepsilon_s d\varepsilon_e < \infty$$

so that $\frac{d\mathbb{P}(r=1,s=1,e=1|X,Z)}{dZ}(X_0, Z_0)$ has the same sign as γ_r.

- **Proof that the sign of α_e is identified**
  Now, let us focus on $\mathbb{P}(e=1|X,Z)$:

$$
\begin{aligned}
\mathbb{P}(e=1|X,Z) &= \mathbb{P}(e=1,r=1|X,Z)+\mathbb{P}(e=1,r=0|X,Z) \\
&= \int_{-X\beta_r-\gamma_r Z}^{\infty}\int_{\mathbb{R}}\int_{-X\beta_e-\alpha_e}^{\infty} f(\varepsilon_r,\varepsilon_s,\varepsilon_e)d\varepsilon_r d\varepsilon_s d\varepsilon_e \\
&+ \int_{-\infty}^{-X\beta_r-\gamma_r Z}\int_{\mathbb{R}}\int_{-X\beta_e}^{\infty} f(\varepsilon_r,\varepsilon_s,\varepsilon_e)d\varepsilon_r d\varepsilon_s d\varepsilon_e \\
&= \int_{\mathbb{R}}\int_{\mathbb{R}}\int_{-X\beta_e}^{\infty} f(\varepsilon_r,\varepsilon_s,\varepsilon_e)d\varepsilon_r d\varepsilon_s d\varepsilon_e \\
&+ \int_{-X\beta_r-\gamma_r Z}^{\infty}\int_{\mathbb{R}}\int_{-X\beta_e-\alpha_e}^{-X\beta_e} f(\varepsilon_r,\varepsilon_s,\varepsilon_e)d\varepsilon_r d\varepsilon_s d\varepsilon_e \\
\Rightarrow d\mathbb{P}(e=1|X,Z)/dZ &= \gamma_r \int_{\mathbb{R}}\int_{-X\beta_e-\alpha_e}^{-X\beta_e} f(-X\beta_r-\gamma_r Z,\varepsilon_s,\varepsilon_e)d\varepsilon_s d\varepsilon_e
\end{aligned}
$$

  Again, if $\alpha_e>0$, then $0<\int_{\mathbb{R}}\int_{-X_0\beta_e-\alpha_e}^{-X_0\beta_e} f(-X_0\beta_r-\gamma_r Z_0,\varepsilon_s,\varepsilon_e)d\varepsilon_s d\varepsilon_e<\infty$, because of hypotheses 3b and 3c. For the same reasons, if $\alpha_e<0$, then $-\infty<\int_{\mathbb{R}}\int_{-X_0\beta_e-\alpha_e}^{-X_0\beta_e} f(-X_0\beta_r-\gamma_r Z_0,\varepsilon_s,\varepsilon_e)d\varepsilon_s d\varepsilon_e<0$. This shows that $d\mathbb{P}(e=1|X,Z)/dZ$ and $\alpha_e\gamma_r$ have the same sign. The sign of γ_r is identified, so the sign of α_e is identified.

C Additional tables

C.1 Determinants of LP_{-ik}

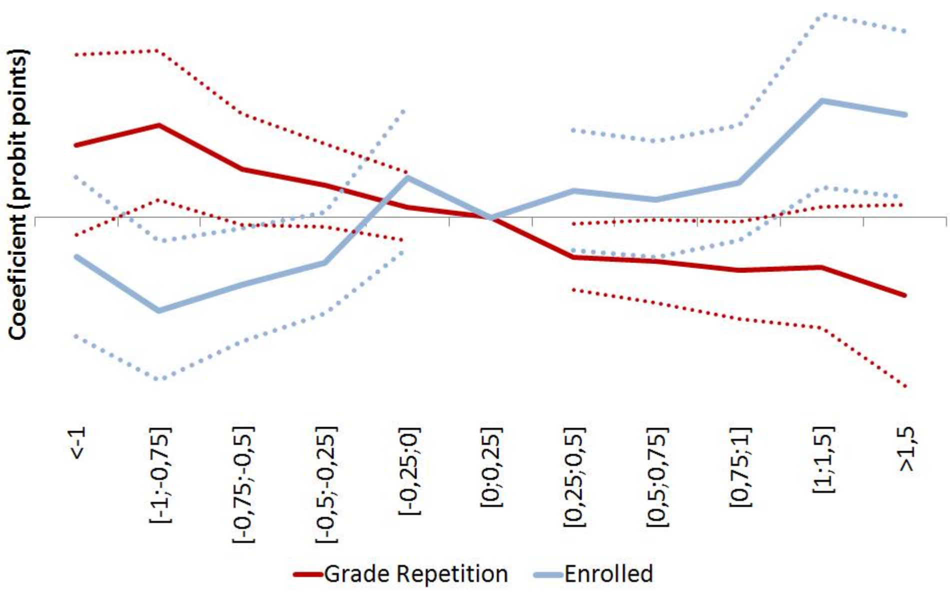
Table C.1: Determinants of LP_{-ik}

	Community characteristics	Household characteristics
	(1)	(2)
Group[a] mean test score	1.028 (.092)***	1.033 (.084)***
Community mean asset index	-.110 (.083)	
Community mean educations index	.143 (.146)	
ln(population)	.036 (.034)	
Community main occupation: trade (ref: agriculture)	.115 (.143)	
Electricity in the community	.022 (.163)	
Rural	.041 (.192)	
Distance to the next health center	.201 (.175)	
Distance to the next hospital	-.023 (.045)	
Asset index		.011 (.021)
Parent's education		.006 (.016)
Household head: non-muslim		.081 (.131)
Household head: Pulaar, halpulaar (ref: wolof)		.097 (.062)
Household head: Serere (ref: wolof)		.017 (.094)
Household head: Dioola (ref: wolof)		.101 (.094)
Household head: Mandingue-Sose (ref: wolof)		.029 (.086)
Household head: others (ref: wolof)		.031 (.072)
Grade*year dummies	Yes	Yes
Obs.	1709	1805
R^2	.545	.517
Joint significance community or hh. variables	1.012	.553
P-value	.439	.811

Note: ***, ** and * mean respectively that the coefficient is significantly different from 0 at the 1%, 5% and 10% level. The standard deviations of the estimators are corrected for the correlation of the residuals between different observations of the same community.

a: A group is composed of all the observations from the same school, the same year and the same grade. This is an approximation of a class, since in some schools, there are several classes per grade.

Figure C.1: Non-linear effect of the difference between test score and last passer's score on grade repetition and enrollment



Notes: Plot of the estimates of Table C.2. Dotted lines give the confidence intervals at the 5% level.

C.2 First stage and reduced form estimates

This section presents the equivalent of the first stage and reduced form estimates corresponding to model (5):

$$\begin{cases} E_{ik,t+1} = \mathbb{1}[\ \beta_{e1}S_{ik} + \beta_{e2a}LP_{-ik} + \beta_{e3}Z_s + f_e(S_{ik} - LP_{-ik}) + X_{ik}\beta_{e4} + u_{ik} > 0] \\ R_{ik} = \mathbb{1}[\ \beta_{r1}S_{ik} - \beta_{r2}LP_{-ik} + f_r(S_{ik} - LP_{-ik}) + X_{ik}\beta_{r4} + \epsilon_{ik} < 0] \\ selection = \mathbb{1}[\ \beta_{s1}S_{ik} + \beta_{s2}LP_{-ik} + \beta_{s3}Z_s + f_s(S_{ik} - LP_{-ik}) + X_{ik}\beta_{s4} + v_{ik} > 0] \end{cases} \tag{C.1}$$

In this specification, we do not measure the effect of grade repetition on dropout. Instead, we measure the effect of the position relative to the target achievement on grade repetition, and on dropout. The effect of the position relative to the target achievement on dropout is assumed to be an indirect effect of grade repetition on dropout in our main estimations.

Table C.2 gives the estimation of model (C.1), and Figure C.1 plots the coefficients of the difference between individual test score and last passer's score. The model is estimated with a maximum likelihood method, as a "trivariate probit" specification. The distribution of the error terms follow a trivariate normal distribution, simulated with a GHK simulator. The three columns of Table 3 correspond to the model's three equations. The data are pooled for the various grades and years. Each specification includes grade-year dummies in each equation.

Determinants of grade repetition and of selection Similarly to Table 3, pupils are much less likely to repeat the grade when their test score is higher than the last passer's score in Table C.2. The corresponding coefficients are plotted in Figure C.1, in dark red. Besides, the negative shocks on harvests are used as an exclusion restriction in the repetition equation. Like in Table C.2, this coefficient is positive and significant in the selection equation.

Table C.2: Grade repetition and school dropouts as a function of a difference between own test score and the test score of the last passer ($S_{ik} - LP_{-ik}$)

	repetition	$enrolled_{t+1}$	*selection*
	(1)	(2)	(3)
Test score	-.231 (.273)	-.200 (.197)	.070 (.148)
LP_{-ik}	.083 (.242)	.468 (.201)**	.047 (.138)
$S_{ik} - LP_{-ik} < -1$	.763 (.481)	-.409 (.424)	-.157 (.324)
$-1 < S_{ik} - LP_{-ik} < -0.75$	.966 (.399)**	-.979 (.372)***	-.610 (.270)**
$-0.75 < S_{ik} - LP_{-ik} < -0.5$	.506 (.297)*	-.700 (.302)**	-.399 (.236)*
$-0.5 < S_{ik} - LP_{-ik} < -0.25$	.339 (.224)	-.471 (.271)*	-.404 (.187)**
$-0.25 < S_{ik} - LP_{-ik} < 0$	.113 (.182)	.421 (.380)	-.329 (.165)**
$0 < S_{ik} - LP_{-ik} < 0.25$	Ref.	Ref.	Ref.
$0.25 < S_{ik} - LP_{-ik} < 0.5$	-.412 (.177)**	.287 (.323)	.205 (.172)
$0.5 < S_{ik} - LP_{-ik} < 0.75$	-.454 (.222)**	.190 (.311)	.275 (.200)
$0.75 < S_{ik} - LP_{-ik} < 1$	-.554 (.258)**	.366 (.309)	.181 (.199)
$1 < S_{ik} - LP_{-ik} < 1.5$	-.520 (.323)	1.226 (.464)***	.273 (.216)
$1.5 < S_{ik} - LP_{-ik}$	-.812 (.484)*	1.085 (.445)**	.469 (.318)
Group[a] mean test score	.268 (.130)**	-.043 (.205)	-.015 (.116)
Negative shock on harvests		.256 (.195)	.414 (.154)***
Household wealth and Parents' education, Previous year's test score	Yes	Yes	Yes
Grade*year dummies	Yes	Yes	Yes
Obs.	1818	1818	1818
log likelihood	-1262.328	-1262.328	-1262.328
χ^2 exclusion restriction			7.211
corresponding p value			.007

Notes: The model is estimated with a maximum likelihood method, as a "trivariate probit" specification. The distribution of the error terms follow a trivariate normal distribution, simulated with a GHK simulator (25 iterations). ***, ** and * mean respectively that the coefficient is significantly different from 0 at the 1%, 5% and 10% level. Standard errors clustered between different observations of the same child.

$S_{ik} - LP_{-ik}$ stands for "Difference between own test score and last passer's test score".

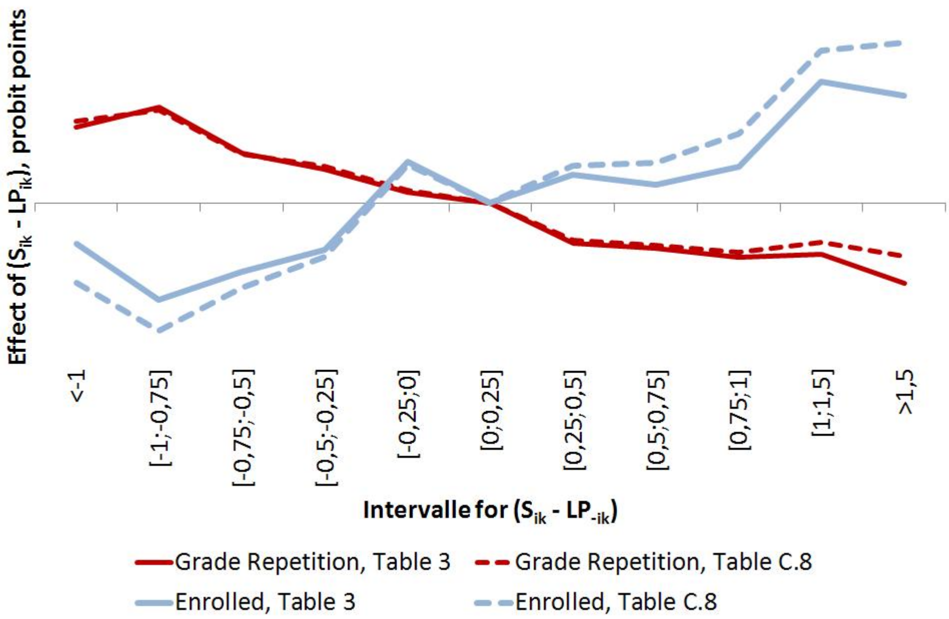
a: A group is composed of all the observations from the same school, the same year and the same grade. This is an approximation of a class, since in some schools, there are several classes per grade.

Determinants of enrollment In the enrollment equation, the coefficients for the difference between own test score and last passer's score is given in Table C.2, and plotted in Figure C.1, in light blue. They tend to mirror the coefficients of the repetition equation: pupils with a learning achievement greater than teacher's standards are the least likely to drop out. In Figure C.1, the curve is grossly symmetric to the curve of the grade repetition equation.

C.3 Reduced form with non-linear controls

Table C.3 and Figure C.2 gives compares the estimates when several variables based on test-scores are treated non-linearly. In sum, we add in the specification dummies for levels of own test score, of group mean test score, of difference to group mean, and of last passer's score. Hence these variables are treated non-linearly. The results in Table C.3 are similar to Table C.2. Neither of the dummies set is jointly significant. If something, Figure C.2 shows the effect of test own score relative to the target achievement on dropout is a bit greater with the dummies set control.

Figure C.2: Comparison between the estimates of Table C.2 and Table C.3



Notes: Plot of the estimates of Table C.2 and Table C.3.

Table C.3: Modification of Table C.2 with non-linear treatment of the control variables based on test scores

	repetition	$enrolled_{t+1}$	*selection*
	(1)	(2)	(3)
Test score	.093	-.282	.072
	(.495)	(.311)	(.255)
LP_{-ik}	-.098	.613	.110
	(.235)	(.269)**	(.188)
$S_{ik} - LP_{-ik} < -1$	.823	-.804	-.189
	(.492)*	(.547)	(.346)
$-1 < S_{ik} - LP_{-ik} < -0.75$	.938	-1.290	-.595
	(.400)**	(.447)***	(.289)**
$-0.75 < S_{ik} - LP_{-ik} < -0.5$	.495	-.858	-.402
	(.309)	(.355)**	(.247)
$-0.5 < S_{ik} - LP_{-ik} < -0.25$	.391	-.320	
	(.184)	(.386)	(.166)*
$-0.25 < S_{ik} - LP_{-ik} < 0$	.127	.391	-.320
	(.184)	(.386)	(.166)*
$0 < S_{ik} - LP_{-ik} < 0.25$	Ref.	Ref.	Ref.
$0.25 < S_{ik} - LP_{-ik} < 0.5$	-.381	.375	.230
	(.185)**	(.326)	(.175)
$0.5 < S_{ik} - LP_{-ik} < 0.75$	-.431	.403	.303
	(.231)*	(.287)	(.205)
$0.75 < S_{ik} - LP_{-ik} < 1$	-.500	.696	.222
	(.275)*	(.323)**	(.219)
$1 < S_{ik} - LP_{-ik} < 1.5$	-.400	1.541	.307
	(.333)	(.527)***	(.242)
$1.5 < S_{ik} - LP_{-ik}$	-.537	1.619	.457
	(.507)	(.546)***	(.343)
Group[e] mean test score	.419	-.671	-.240
	(.378)	(.478)	(.314)
Negative shock on harvests		.156	.396
		(.196)	(.147)***
χ^2 test score dummies[a]	6.982	7.742	.592
corresponding p value	.137	.101	.964
χ^2 difference to group[e] mean dummies[b]	1.416	5.245	1.412
corresponding p value	.841	.263	.842
χ^2 group[e] mean dummies[c]	5.695	6.128	2.189
corresponding p value	.127	.106	.534
χ^2 last passer's score dummies[d]	7.539	1.984	7.539
corresponding p value	.184	.851	.184
Household wealth and Parents' education, Previous year's test score	Yes	Yes	Yes
Grade*year dummies	Yes	Yes	Yes
Obs.	1818	1818	1818
log likelihood	-1237.971	-1237.971	-1237.971
χ^2 grade year dummies	4.958	29.208	9.239
corresponding p value	.292	7.09e-06	.055
χ^2 exclusion restriction			7.281
corresponding p value			.007

Additional covariates in each equation: test score, group mean test score previous year's test score, household wealth, parents' education, grade-year dummies.
Note: ***, ** and * mean respectively that the coefficient is significantly different from 0 at the 1%, 5% and 10% level. The standard deviations of the estimators are corrected for the correlation of the residuals between different observations of the same child.
a: Dummies for: test score<-1, -1<test score<-0.5, 0<test score<0.5, 0.5<test score<1, 1<test score. -0.5<test score<0 omitted
b: Difference to group mean: difference between own test score and the group mean. Dummies for: difference<-1, -1<difference<-0.5, -0.5<difference<0, 0.5<difference. 0<difference<0.5 omitted
c: Dummies for: group mean<-0.5, -0.5<group mean<0, 0.5<group mean. 0<group mean<0.5 omitted
d: Dummies for: last passer's score<-1.5, -1<last passer's score<-0.5, -0.5<last passer's score<0, 0<last passer's score<0.5, 0.5<last passer's score. -1.5<group mean<-1 omitted
e: A group is composed of all the observations from the same school, the same year and the same grade. This is an approximation of a class, since in some schools, there are several classes per grade.

D Simulations

D.1 Formulas

The model of interest is:

$$\begin{cases} E_{ik,t+1} & = \mathbb{1}[\ \beta_{e1}S_{ik} \quad +\beta_{e2}LP_{-ik} \qquad\qquad\qquad\qquad +\gamma R_{ik} \quad +X_{ik}\beta_{e4} + u_{ik} > 0] \\ R_{ik} & = \mathbb{1}[\ \beta_{r1}S_{ik} \quad -\beta_{r2}LP_{-ik} \quad +f_r(S_{ik} - LP_{-ik}) \qquad\qquad +X_{ik}\beta_{r4} + \epsilon_{ik} < 0] \\ selection & = \mathbb{1}[\ \beta_{s1}S_{ik} \quad +\beta_{s2}LP_{-ik} \quad +\beta_{s3}Z_s \qquad\qquad +\gamma_s R_{ik} \quad +X_{ik}\beta_{s4} + \upsilon_{ik} > 0] \end{cases} \tag{D.1}$$

Grade repetition risk For each observation, it is possible to compute the grade repetition risk: $P_{red} = \Phi\left(\beta_{r1}S_{ik} - \beta_{r2}LP_{-ik} + f_r(S_{ik} - LP_{-ik}) + X_{ik}\beta_{r4}\right)$. This risk can easily be adapted to speculative situations with different LP_{-ik}. $\tilde{P}_{red} = \Phi\left(\beta_{r1}S_{ik} - \beta_{r2}\widetilde{LP}_{ik} + f_r(S_{ik} - \widetilde{LP}_{ik}) + X_{ik}\beta_{r4}\right)$ gives individual grade repetition risks. The simulations presented here give their sample average (and sub-sample averages).

Dropout risk, no "direct effect of grade repetition policy" To simplify the algebra, $E_{ik,t+1}$ and R_{ik} are assumed independent. Hence the probability of $E_{ik,t+1}$ writes
$P_{enr} = P_{red}\Phi\left(\beta_{e1}S_{ik} + \beta_{e2}LP_{-ik} + \gamma + X_{ik}\beta_{e4}\right) + (1 - P_{red})\Phi\left(S_{ik} + \beta_{e2}LP_{-ik} + X_{ik}\beta_{e4}\right)$. The simulations compute the consequences of a speculative change in LP_{-ik} on P_{red}. They change dropout risk through the change in P_{red}, but not through $\beta_{e2}LP_{-ik}$. The new probability of $E_{ik,t+1}$ writes $\tilde{P}_{enr} = \tilde{P}_{red}\Phi\left(\beta_{e1}S_{ik} + \beta_{e2}LP_{-ik} + \gamma + X_{ik}\beta_{e4}\right) + (1 - \tilde{P}_{red})\Phi\left(S_{ik} + \beta_{e2}LP_{-ik} + X_{ik}\beta_{e4}\right)$, where $\tilde{P}_{red}$ is the new P_{red}.

Dropout risk, with "direct effect of grade repetition policy" $E_{ik,t+1}$ and R_{ik} are still assumed independent. The simulations compute the consequences of a speculative change in LP_{-ik} on P_{red}. They change dropout risk through the change in P_{red}, and through $\beta_{e2}LP_{-ik}$. The new probability of $E_{ik,t+1}$ writes $\tilde{P}_{enr} = \tilde{P}_{red}\Phi\left(\beta_{e1}S_{ik} + \beta_{e2}\widetilde{LP}_{ik} + \gamma + X_{ik}\beta_{e4}\right) + (1 - \tilde{P}_{red})\Phi\left(S_{ik} + \beta_{e2}\widetilde{LP}_{ik} + X_{ik}\beta_{e4}\right)$, where $\tilde{P}_{red}$ is the new P_{red}, and $\widetilde{LP}_{ik}$ is the speculative LP_{-ik}.

Precision of the estimates When presented in the paper, they derive from a Delta-method not detailed here.

Table C.4: Estimation of the link between last passer's test score and school outcomes in reduced form (probit models)

	$enrolled_{t+1}$	$enrolled_{t+2}$	$enrolled_{t+3}$	$enrolled_{t+4}$	*Still enrolled* (2003)	*Last Grade* > 5	*Last Grade* > 6	*Last Grade* > 7
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Test score	.275	.162	.346	.432	.377	.492	.759	.781
	(.116)**	(.099)	(.084)***	(.076)***	(.055)***	(.061)***	(.062)***	(.071)***
LP_{-ik}	-.009	.053	.152	.165	.184	.102	.200	.177
	(.084)	(.090)	(.086)*	(.079)**	(.062)***	(.069)	(.063)***	(.070)**
Group[a] mean test score	-.160	-.048	-.281	-.369	-.453	-.322	-.747	-.690
	(.192)	(.196)	(.173)	(.159)**	(.123)***	(.131)**	(.138)***	(.153)***
Negative shock on harvests	.180	.355	.191	.130	.166	-.053	.082	-.006
	(.211)	(.244)	(.220)	(.208)	(.151)	(.144)	(.146)	(.162)
Household wealth and Parents' education, Previous year's test score	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Grade*year dummies	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Obs.	1818	1449	1449	1449	1789	1777	1777	1777
log-likelihood	-158.914	-224.825	-371.157	-487.188	-1060.973	-833.129	-892.445	-756.374

Additional covariates in each equation: previous year's test score, household wealth, parents' education, grade-year dummies.
Note: ***, ** and * mean respectively that the coefficient is significantly different from 0 at the 1%, 5% and 10% level. The standard deviations of the estimators are corrected for the correlation of the residuals between different observations of the same child.
a: A group is composed of all the observations from the same school, the same year and the same grade. This is an approximation of a class, since in some schools, there are several classes per grade.

D.2 Additional Tables

Table D.1: Simulations: effect of a decrease of last passer's score

			Reference	LP_{-ik} decreases by 0.25 pt	LP_{-ik} decreases by 0.5 pt	LP_{-ik} decreases by 0.75 pt	LP_{-ik} decreases by 1 pt
Grade repetition		Full sample	26,4%	22,5%	19,3%	16,7%	14,7%
		"low-achievers"	74,1%	77,9%	74,7%	71,2%	65,5%
		"leaning"	51,8%	40,8%	32,5%	27,4%	23,8%
		"good pupils"	12,7%	11,4%	10,3%	8,7%	7,7%
Drop-out risk	No "direct effect of g.r. policy"	Full sample	3,2%	2,7%	2,4%	2,1%	1,9%
		"low-achievers"	7,1%	7,5%	7,1%	6,7%	6,1%
		"leaning"	5,7%	4,5%	3,6%	3,0%	2,6%
		"good pupils"	1,9%	1,7%	1,6%	1,4%	1,2%
	With "direct effect of g.r. policy"	Full sample	3,2%	3,2%	3,1%	3,1%	3,2%
		"low-achievers"	7,1%	8,7%	9,5%	10,3%	10,7%
		"leaning"	5,7%	5,1%	4,7%	4,6%	4,5%
		"good pupils"	1,9%	2,0%	2,1%	2,1%	2,2%

Notes: Simulations based on the estimates of Table 3. The unit for test scores is the standard deviation of distribution of the test for the year-grade.
"low-achievers": $S_{ik} - LP_{-ik} < -1$
"leaning": $-1 < S_{ik} - LP_{-ik} < 0.25$
"good pupils": $0.25 < S_{ik} - LP_{-ik}$