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

Pierre André†

October 12, 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, schools with tough grade repetition policies show relatively successful academic outcomes. They 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, and Marco Manacorda 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 12\% of the pupils enrolled in Senegalese primary schools in 2004 were repeating their grades in 2005,\(^1\) whereas the African average is 15\% (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.\(^1,2\) 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. There are many reasons why high repetition rates in primary school can decrease national school attainment rates.

First, grade repetition is very expensive for the state and households alike since both private and public costs of schooling increase with the duration of schooling. For a given final grade, a grade repetition increases the time spent at school by one year, and increases accordingly all the costs of education. Whether the costs of grade repetition are compatible with universal primary education in developing countries is seriously debated in multilateral institutions. The average repetition rate is included in the “Comparative Performance Indicators” of the “Global Partnership for Education”,\(^3\) launched by the World Bank in collaboration with other donors.

Besides, grade repetition is probably 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. Economically, 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.

The psychologists and the pedagogical profession tend to share a widespread view that grade repetition does not improve learning achievement. Many studies in this branch of the literature measure the net effect of grade repetition on learning achievement. They usually control for pre-repetition test scores as a proxy for school ability and initial learning achievement at a given date (see Holmes (1989) for a meta analysis of many of those studies, and McCoy and Reynolds (1999) for a more recent study). However, these evaluations may suffer from endogeneity issues: conditional on pre-repetition test scores, grade repetition may be correlated with individual ability.

Jacob and Lefgren (2004) control for this potential bias using a discontinuity in school policy in Chicago. Pupils there took 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

\(^{1}\)Ministry of Education, Senegal (2005)
\(^{2}\)www.poledakar.org
\(^{3}\)The Global Partnership for Education is comprised of 46 developing countries, and more than 30 bilateral, regional, and international agencies, development banks (...) devoted to getting all children everywhere into school”. see http://www.globalpartnership.org
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 effect of grade repetition on school attainment in developing countries has been extensively studied with control-based identification strategies. However, conditional on a test-score prior to the grade repetition decision, teacher’s grade repetition decisions are probably not taken as random. Manacorda (2012) uses the Uruguayan 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.

The question of dropouts induced by grade-repetition is interesting 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, endogenous school dropouts may be economically inefficient under imperfect information. Two recent controlled experiments in developing countries (Nguyen, 2008 and Jensen, 2007) have shown that further information on the returns to schooling affect the school investment decisions. In the end, both political commitment and economic efficiency make it necessary to fight against endogenous primary school dropout.

This paper estimates the effect of grade repetition decision on immediate school dropout, i.e. on the probability to be enrolled at school the next school year. It controls for the potential correlation between the children’s unobservable characteristics and grade repetition with an original instrumental variables strategy, and controls for the selection of the information on grade repetition decisions. My instrumental strategy is based on the widespread idea that a child needs to reach a certain learning achievement to pass to the next grade. The grade repetition probability is strongly non-linear between pupils whose learning achievement are above and below this “target achievement”. This paper tries to exploit this non-linearity to identify the effect of grade repetition on school dropout.

The results reveal a negative effect of grade repetition on the probability of enrolment 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, whereas the average dropout rate in the sample is 2%. However, schools with tough grade repetition policies are relatively successful, which questions the validity of policies fighting against grade repetition.

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

PASEC and EBMS datasets both contain detailed information about schooling and are combined here to estimate the effect of grade repetition.

---

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

5 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.

6 The number of failed subjects being an integer, one can still discuss the validity of regression discontinuity designs in this context.
2.1 The PASEC panel

The PASEC conducted a panel survey among primary school pupils of 98 Senegalese schools between 1995 and 2001. 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. 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.

Whenever a child took a PASEC test in a given school year, the information includes his current grade. The information for grade repetition is 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 certain 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). However, the school information from EBMS does not give much information on grade repetitions. 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.

2.4 Selection on test participation

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.

---

7The tests were marked by the PASEC team. Consequently, test scores could not be influenced by teachers.
8Education et Bien-être des Ménages au Sénégal. This survey was designed by a team composed of Peter Glick, David Sahn, and Léopold Sarr (Cornell University, USA), and Christelle Dumas and Sylvie Lambert (LEA-INRA, France), and implemented in association with the Centre de Recherche en Economie Appliquée (Dakar, Senegal).
9It includes the number of grade repetitions in primary school and the number of grade repetitions in the last grade of primary school for each child living in a household surveyed by EBMS. This paper requires longitudinal information not included in the EBMS data.
Table 1: Number of children attending the tests during the panel, by grade and school year

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Sixth (CM2)</td>
<td>214</td>
<td>357</td>
<td>204</td>
<td>86</td>
<td>15</td>
<td>594</td>
</tr>
<tr>
<td>Fifth (CM1)</td>
<td>236</td>
<td>412</td>
<td>204</td>
<td>86</td>
<td>15</td>
<td>536</td>
</tr>
<tr>
<td>Fourth (CE2)</td>
<td>102</td>
<td>154</td>
<td>53</td>
<td>15</td>
<td></td>
<td>274</td>
</tr>
<tr>
<td>Third (CE1)</td>
<td>102</td>
<td>154</td>
<td>53</td>
<td>15</td>
<td></td>
<td>274</td>
</tr>
<tr>
<td>Second (CP)</td>
<td>102</td>
<td>154</td>
<td>53</td>
<td>15</td>
<td></td>
<td>274</td>
</tr>
<tr>
<td>Initial tests</td>
<td>817</td>
<td>817</td>
<td>696</td>
<td>566</td>
<td>614</td>
<td>551</td>
</tr>
</tbody>
</table>

Note: This table reports the attendance among the 921 children of PASEC sample resurveyed by EBMS 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. There is probably no reason dropouts could be more likely to take the tests.

If the difference between pupil’s learning achievement and teacher’s standards affects attendance to the tests, our estimates are likely to be biased. We believe that, if something, children lagging behind teacher’s standards are less likely to take the tests. For example, assume pupils have some information on their learning achievement relative to teacher’s standards before the tests. Pupils lagging behind teacher’s standards are probably less likely to take the tests than others: they may be discouraged or anticipate a dropout. Hence, the selection on unobservables is more drastic among them: only the most motivated of them take the tests. Hence, this bias would increase their dropout rate relative to others. We find that children lagging behind teacher’s standards are more likely to dropout in the selected sample, the difference may be even stronger in the full sample. Indeed, some children lagging behind teacher’s standards may be excluded from the sample because they have anticipated their dropout.

2.5 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. 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 \( t + 1 \).\(^\text{10}\) 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 decisions on school dropout: if grade repetition causes dropout, then it causes its own selection. However,\(^\text{10}\) The details and other cases are explained in appendix A
Figure 1: Sequence of the main events during the PASEC panel

Table 2: Observation of grade repetition decision

<table>
<thead>
<tr>
<th>date $t$</th>
<th>date $t+1$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Enrolled</td>
<td>Enrolled</td>
</tr>
<tr>
<td></td>
<td>Drops out</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Grade repetition decision</th>
<th>Grade repetition decision</th>
</tr>
</thead>
<tbody>
<tr>
<td>is observed</td>
<td>is not observed</td>
</tr>
</tbody>
</table>

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 (1):\(^{11}\)

$$
\begin{align*}
E_{ik,t+1} &= \mathbb{I}[\beta_{e1}S_{ik} + \beta_{e2}Z_s + X_{ik}\beta_{e3} + \gamma R_{ik} + u_{ik} > 0] \\
R_{ik} &= \mathbb{I}[S_{ik} + \beta_{r2}Z_R + X_{ik}\beta_{r3} + \epsilon_{ik} < 0] \\
\text{selection} &= \mathbb{I}[\beta_{s1}S_{ik} + \beta_{s2}Z_s + X_{ik}\beta_{s3} + \gamma_s R_{ik} + v_{ik} > 0]
\end{align*}
$$

(1)

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

- 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 (1) are 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. $\gamma$ and $\gamma_s$ 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 simpler 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,

\(^{11}\)The grade repetition of a child is denoted $R_{ik}$ for child $i$ in class $k$. His enrolment during the next school year is denoted $E_{ik,t+1}$ (at date $t + 1$). $selection$ takes value 1 if $R_{ik}$ is known, and 0 otherwise. $S_{ik}$ denotes the test score of child $i$ in class $k$. $Z_s$ is an instrument for the selection, which is discussed in section 3.3. $Z_R$ is an instrument for grade repetition which is discussed in section 3. The other determinants in each equation are a vector of covariates $X_{ik}$ and unobservables $u_{ik}, \epsilon_{ik}$ and $v_{ik}$.

\(^{12}\)Hypotheses about points where the distribution of $(\epsilon_{ik}, u_{ik}, v_{ik})$ should be positive and finite, and about the support of the distribution of the observables.
the derivative of the probability of grade repetition towards \( Z_R \) gives the sign of \( \alpha \) 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 enrolment towards \( Z_R \) have the same sign, and negative if they have opposite signs.

This paper does not intend to identify model (1) 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 the results in this paper do probably not only rely on the parametric structure of the models but also on the information from the data.

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 \( k \), at date \( t + 1 \)), which is the coefficient \( \gamma \) in the equation (2) below. The other determinants of dropout are test score \( S_{ik} \), and a vector of covariates \( X_{ik} \).

\[
E_{ik,t+1} = \mathbb{I} [\beta_{e1} S_{ik} + X_{ik} \beta_{e2} + \gamma R_{ik} + u_{ik} > 0] \tag{2}
\]

The main difficulty in identifying \( \gamma \) 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.

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:

\[
R_{ik} = \mathbb{I} [\beta_{r1} S_{ik} - \beta_{r2} t_k + f_r(S_{ik} - t_k) + X_{ik} \beta_{r4} + \epsilon_{ik} < 0] \tag{3}
\]

However, 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 \( LP_{ik} \):

\[
LP_{ik} = \min_{j \neq i, R_{jk} = 0} (S_{jk}) \tag{4}
\]

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 (5):

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

\(^{13}\)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.

\(^{14}\)This vector includes grade-year dummies, household wealth parents’ education, and group mean test score when not included in the model.
Model (5) controls for is 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. (the third equation controls for the selection issue described in section 2.5) The control for $S_{ik}$ ensures that the specification compares pupils with similar learning achievement, and potentially similar background in terms of unobservable characteristics. The control for $LP_{ik}$ controls for an homogeneous 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 (5) 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 (5).

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 scores and dropout between classes with different requirements to pass the grade. We observe that students whose learning achievement lag behind teacher’s standards tend to repeat more often, and drop out more often. Model (5) 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 average test score and dropout
- The link between the relative position of individual test score to the last passer’s score and dropout

We can make clear only the latter matters in terms of dropout. (This is discussed in the last paragraph of section 3.3) This can probably rule out most endogeneity issues. However, if families observe the relative position of their children’s test score to the last passer’s score, this position probably affects dropout. We would identify an effect of the inability to meet teacher’s requirements on dropout in the general sense. In other words, this is an effect of school failure on dropout rather than an effect of grade repetition on dropout. The paper cannot disentangle the effects of the inability to meet teacher’s requirements from the effects of 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 that peer effects may lead to a “mirror effect”, where child $i$’s characteristics and $i$’s peer’s characteristic interact in both directions. In this paper, interactions between peers can probably affect our

---

15I mean linear in the latent variable
estimations mostly through $LP_{ik}$. If child $i$’s unobservable characteristics can affect $LP_{ik}$, $LP_{ik}$ 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 average 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, the correlation between test scores and dropout is probably lower: pupils with favorable unobservable characteristics rarely dropout even in adverse situations (the dropout probability is $\approx 2\%$ in the sample ). 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. Hence the correlation between $LP_{ik}$ and dropout is high. 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 inability 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.

This might nevertheless have general equilibrium 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 general equilibrium effects, measured by the direct effect of $LP_{ik}$ in model (5). In this paper, it is not possible to claim the variations in $LP_{ik}$ are due to a well identified and exogenous source. However, conditionnally on test scores, $LP_{ik}$ appears uncorrelated with individual and location characteristics. In appendix, Table C.1 shows that $LP_{ik}$ strongly depends on group average test score. The coefficient in the linear regression is 1, as if the grade repetition rates and group average learning achievement were independent. Besides, $LP_{ik}$ is not correlated with either observable community characteristics or observable household characteristics. Hence, conditionally on average test score, $LP_{ik}$ 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, $LP_{ik}$ may be one of these traits, and be correlated with other determinants of dropouts. It is hard to give the sign and magnitude of this effect. In the end, it is difficult to affirm the coefficients on $LP_{ik}$ in the estimations are causal.

### 3.3 First stage and reduced form estimates

This section presents the equivalent of the first stage and reduced form estimates corresponding to model (5):
Figure 2: Non-linear effect of the difference between test score and last passer’s score on grade repetition and enrollment

\[ E_{ik,t+1} = 1[ \beta_{e1}S_{ik} + \beta_{e2}LP_{ik} + f_e(S_{ik} - LP_{ik}) + X_{ik}\beta_{e4} + u_{ik} > 0] \]

\[ R_{ik} = 1[ \beta_{r1}S_{ik} - \beta_{r2}LP_{ik} + f_r(S_{ik} - LP_{ik}) + X_{ik}\beta_{r4} + \epsilon_{ik} < 0] \]

\[ \text{selection} = 1[ \beta_{s1}S_{ik} + \beta_{s2}LP_{ik} + \beta_{s3}Z_s + f_s(S_{ik} - LP_{ik}) + X_{ik}\beta_{s4} + v_{ik} > 0] \]

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 3 gives the estimation of model (6), and Figure 2 plots the coefficients of the difference between own 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 4 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** In the grade repetition equation, the effect of test score on dropout 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.
Table 3: Grade repetition and school dropouts as a function of a difference between own test score and the test score of the last passer ($S_{ik}^* - LP_{ik}$)

<table>
<thead>
<tr>
<th></th>
<th>repetition</th>
<th>enrolled$_t+1$</th>
<th>selection</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>(1)</td>
<td>(2)</td>
<td>(3)</td>
</tr>
<tr>
<td>Test score</td>
<td>-.231</td>
<td>-.200</td>
<td>.070</td>
</tr>
<tr>
<td></td>
<td>(.273)</td>
<td>(.197)</td>
<td>(.148)</td>
</tr>
<tr>
<td>Last passer’s test score</td>
<td>.083</td>
<td>.468</td>
<td>.047</td>
</tr>
<tr>
<td></td>
<td>(.242)</td>
<td>(.201)**</td>
<td>(.138)</td>
</tr>
<tr>
<td>$S_{ik}^* - LP_{ik} &lt; -1$</td>
<td>.763</td>
<td>-.409</td>
<td>-.157</td>
</tr>
<tr>
<td></td>
<td>(.481)</td>
<td>(.424)</td>
<td>(.324)</td>
</tr>
<tr>
<td>$-1 &lt; S_{ik}^* - LP_{ik} &lt; -0.75$</td>
<td>.966</td>
<td>-.979</td>
<td>-.610</td>
</tr>
<tr>
<td></td>
<td>(.399)**</td>
<td>(.372)**</td>
<td>(.270)**</td>
</tr>
<tr>
<td>$-0.75 &lt; S_{ik}^* - LP_{ik} &lt; -0.5$</td>
<td>.506</td>
<td>-.700</td>
<td>-.399</td>
</tr>
<tr>
<td></td>
<td>(.297)*</td>
<td>(.302)**</td>
<td>(.236)*</td>
</tr>
<tr>
<td>$-0.5 &lt; S_{ik}^* - LP_{ik} &lt; -0.25$</td>
<td>.339</td>
<td>-.471</td>
<td>-.404</td>
</tr>
<tr>
<td></td>
<td>(.224)</td>
<td>(.271)*</td>
<td>(.187)**</td>
</tr>
<tr>
<td>$-0.25 &lt; S_{ik}^* - LP_{ik} &lt; 0$</td>
<td>.113</td>
<td>.421</td>
<td>-.329</td>
</tr>
<tr>
<td></td>
<td>(.182)</td>
<td>(.380)</td>
<td>(.165)**</td>
</tr>
<tr>
<td>$0 &lt; S_{ik}^* - LP_{ik} &lt; 0.25$</td>
<td>Ref.</td>
<td>Ref.</td>
<td>Ref.</td>
</tr>
<tr>
<td>$0.25 &lt; S_{ik}^* - LP_{ik} &lt; 0.5$</td>
<td>-.412</td>
<td>.287</td>
<td>.205</td>
</tr>
<tr>
<td></td>
<td>(.177)**</td>
<td>(.323)</td>
<td>(.172)</td>
</tr>
<tr>
<td>$0.5 &lt; S_{ik}^* - LP_{ik} &lt; 0.75$</td>
<td>-.454</td>
<td>.190</td>
<td>.275</td>
</tr>
<tr>
<td></td>
<td>(.222)**</td>
<td>(.311)</td>
<td>(.200)</td>
</tr>
<tr>
<td>$0.75 &lt; S_{ik}^* - LP_{ik} &lt; 1$</td>
<td>-.554</td>
<td>.366</td>
<td>.181</td>
</tr>
<tr>
<td></td>
<td>(.258)**</td>
<td>(.309)</td>
<td>(.199)</td>
</tr>
<tr>
<td>$1 &lt; S_{ik}^* - LP_{ik} &lt; 1.5$</td>
<td>-.520</td>
<td>1.226</td>
<td>.273</td>
</tr>
<tr>
<td></td>
<td>(.323)</td>
<td>(.464)**</td>
<td>(.216)</td>
</tr>
<tr>
<td>$1.5 &lt; S_{ik}^* - LP_{ik}$</td>
<td>-.812</td>
<td>1.085</td>
<td>.469</td>
</tr>
<tr>
<td></td>
<td>(.484)*</td>
<td>(.445)**</td>
<td>(.318)</td>
</tr>
<tr>
<td>Group mean test score</td>
<td>.268</td>
<td>-.043</td>
<td>-.015</td>
</tr>
<tr>
<td></td>
<td>(.130)**</td>
<td>(.205)</td>
<td>(.116)</td>
</tr>
<tr>
<td>Negative shock on harvests</td>
<td>.256</td>
<td>.414</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(.195)</td>
<td>(.154)**</td>
<td></td>
</tr>
</tbody>
</table>

Household wealth and Parents’ education, Previous year’s test score
Grade*year dummies
Obs. 1818 1818 1818
log likelihood -1262.328 -1262.328 -1262.328
$\chi^2$ exclusion restriction
corresponding p value 7.211 .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”.

$S_{ik}^* - LP_{ik}$ stands for “Difference between own test score and last passer’s test score”.
Figure 3: Average grade repetition risk as a function of $S_{ik} - LP_{ik}$

The difference between own learning achievement and last passer’s score is strongly correlated with grade repetition. 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 estimates from Table 3 are plotted in Figure 2, in dark red. Figure 3 illustrate the quantitative magnitude of the effect. On average, changing of $S_{ik} - LP_{ik}$ can change the grade repetition risk from 50% to 5%.

Finally, for a given test score, grade repetition is more likely in good classes (in classes with a great mean test score). This is likely to mean that teacher’s standards are affected by the pupil’s learning achievement: the better the pupils do in class, the tougher the grade repetition policy will be.

**Determinants of enrollment** In the enrollment equation, the coefficient for test scores on dropout is not significant. On the other hand, the probability to be enrolled next year increases with the last passer’s test score. As seen above, many explanations can be given for this coefficient. However, should the interpretation of this coefficient be causal, this would mitigate the benefits of decreasing the last passer’s score on dropout.

The coefficients for the difference between own test score and last passer’s score is given in Table 3, and plotted in Figure 2, 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 2, the curve is grossly symmetric to the curve of the grade repetition equation, with two exceptions.

First, the best pupils ($S_{ik} - LP_{ik} > 1$) are the most likely to be enrolled the next year. They are not significantly less likely to repeat than their peers with $S_{ik} - LP_{ik} \in [0.5; 1]$. Very good pupils (relative to teacher’s standards) may dropout even less often than good pupils, even if neither are likely to repeat their grade. This is not a direct effect of distance to the teacher’s standards on dropout, which is not an effect of grade repetition.

Second, the curve of the coefficients for grade repetition is decreasing between $[-0.25; 0]$ and
[0.25; 0.5]. The curve of the same coefficients in the enrollment equation is flat. Close grade repetitions may be the least likely to generate dropouts because pupils are the most likely to make it to the next grade the next year. This could be interpreted as an heterogeneous effect of grade repetition.

**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 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.

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 difference between own test score and last passer’s score is not treated as the other explanatory variables. Indeed, we assume that the effect of all the other explanatory is linear (in the latent variable equation), and the effect of the difference (between own test score and last passer’s score) is non-linear. 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.2 in appendix, the effect of some of the other explanatory variables is treated as non-linear. The results are very similar, and the non-linearities are not statistically significant (see Figure C.1).

### 3.4 Main results

Table 4 shows the estimation of model (5). This model is the structural equation corresponding to the reduced-form equation presented in Table 3.

**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 4. The corresponding variables are strongly significant, the \( \chi^2 \) test for the significance is about 30. Besides, the negative shocks on harvests are used as an exclusion restriction in the repetition equation. Like in Table 3, this coefficient is positive and significant in the selection equation.

**The effect of grade repetition on school dropout** Table 3 measured the effect of the difference between own test score and last passer’s score. This Table shows that when this difference is positive, the pupil is much more likely to be enrolled the next year. In Table 4, we consider that the only reason why lagging behind teacher’s standard affects dropout is the effect of grade repetition on dropout. Unsurprisingly given Table 3, Table 4 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%.
Table 4: Joint estimation of the determinants of grade repetition, selection and school dropouts corresponding to the model (5)

<table>
<thead>
<tr>
<th></th>
<th>repetition</th>
<th>enrolled_{t+1}</th>
<th>selection</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>(1)</td>
<td>(2)</td>
<td>(3)</td>
</tr>
<tr>
<td>Test score</td>
<td>-.223</td>
<td>-.023</td>
<td>-.238</td>
</tr>
<tr>
<td></td>
<td>(.223)</td>
<td>(.149)</td>
<td>(.123)*</td>
</tr>
<tr>
<td>Last passer’s test score</td>
<td>.074</td>
<td>.379</td>
<td>.305</td>
</tr>
<tr>
<td></td>
<td>(.207)</td>
<td>(.137)***</td>
<td>(.095)***</td>
</tr>
<tr>
<td>$S_{ik}^* - LP_{ik} &lt; -1$</td>
<td>.695</td>
<td>Ref.</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(.407)*</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$-1 &lt; S_{ik}^* - LP_{ik} &lt; -0.75$</td>
<td>1.131</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(.347)***</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$-0.75 &lt; S_{ik}^* - LP_{ik} &lt; -0.5$</td>
<td>.631</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(.262)**</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$-0.5 &lt; S_{ik}^* - LP_{ik} &lt; -0.25$</td>
<td>.485</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(.193)**</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$-0.25 &lt; S_{ik}^* - LP_{ik} &lt; 0$</td>
<td>.269</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(.155)*</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$0 &lt; S_{ik}^* - LP_{ik} &lt; 0.25$</td>
<td>Ref.</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$0.25 &lt; S_{ik}^* - LP_{ik} &lt; 0.5$</td>
<td>-.379</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(.159)**</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$0.5 &lt; S_{ik}^* - LP_{ik} &lt; 0.75$</td>
<td>-.451</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(.192)**</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$0.75 &lt; S_{ik}^* - LP_{ik} &lt; 1$</td>
<td>-.401</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(.226)*</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$1 &lt; S_{ik}^* - LP_{ik} &lt; 1.5$</td>
<td>-.492</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(.274)*</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$1.5 &lt; S_{ik}^* - LP_{ik}$</td>
<td>-.815</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(.451)*</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Negative shock on harvests</td>
<td>.150</td>
<td>.569</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(.254)</td>
<td>(.207)***</td>
<td></td>
</tr>
<tr>
<td>Grade repetition</td>
<td>-2.259</td>
<td>-2.740</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(.337)***</td>
<td>(.463)***</td>
<td></td>
</tr>
<tr>
<td>Average marginal effect of grade repetition</td>
<td>-.138</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(.051)***</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Household wealth and Parents’ education, Previous year’s test score, Group mean test score</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Obs.</td>
<td>1818</td>
<td>1818</td>
<td>1818</td>
</tr>
<tr>
<td>log likelihood</td>
<td>-1258.516</td>
<td>-1258.516</td>
<td>-1258.516</td>
</tr>
<tr>
<td>$\chi^2$ exclusion restrictions</td>
<td>30.290</td>
<td>7.575</td>
<td></td>
</tr>
<tr>
<td>corresponding p value</td>
<td>.0008</td>
<td>.06</td>
<td></td>
</tr>
</tbody>
</table>

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). 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.
This is not completely unlikely. All the dropouts we observe took place during primary school. School dropouts during school cycles are the most likely to be due to school failure. In addition, during the follow-up EMBM survey in 2003, household members told the reason of their dropout (when relevant). Most of the dropouts in our sample\textsuperscript{16} told us why they dropped out. According to them, 60% dropped out because of school failure, and 27% for “other reasons”. It is therefore credible that dropout within a school cycle is massively associated with school failure.

**Potential general equilibrium effects** The last passer’s score is positively correlated 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 4 illustrate this fact (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 groups. The “good pupils” \((0.25 < S_{ik} - LP_{ik})\) probably pass. The “leaning” the pupils benefit the most from a decrease of \(LP_{ik} (-1 < S_{ik} - LP_{ik} < 0.25)\). The “low-achievers” probably repeat anyway \((S_{ik} - LP_{ik} < -1)\).

With the estimates of Table 4, Figure 4 simulates the consequences of a decrease in \(LP_{ik}\) by 0.25, 0.5, 0.75 and 1 standard deviation of the test scores.\textsuperscript{17} This represents a fairly substantial decrease of grade repetition decisions, from 26% to 15%. As expected, this change is concentrated among “leaning” pupils, whose grade repetition probability decrease from 52% to 24%. (see Table D.1)

The dotted lines in Figure 4 show the consequences of a decrease of \(LP_{ik}\) if we neglect the potential general equilibrium effects. It means we model the consequences of a decrease of \(LP_{ik}\) on grade repetition, and assume the only consequences on dropout are due to the subsequent decrease of grade repetition rates.\textsuperscript{17} 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 4 show the consequences of a decrease of \(LP_{ik}\) if we assume the general equilibrium effects are measured by the coefficients of \(LP_{ik}\) in Table 4.\textsuperscript{17} It shows a totally different picture: dropout rates do not decrease when \(LP_{ik}\) decrease. The reasons are the following. For “leaning” pupils, decreasing \(LP_{ik}\) decrease the grade repetition rates and decrease dropout rates. On the other hand, for the “low-achievers”, decreasing \(LP_{ik}\) does not decrease much grade repetition risk. On the contrary, it increases dropout risk conditional on grade repetition. The dropout risk for this group increases from 7.1% to 10.7%.

## 4 Mid-term consequences of grade repetition

### 4.1 Full sample

This section assesses the mid-term consequences of grade repetition. Table 5 uses the same specifications as Table 4, changing the dependent variable.

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

(7)

In equation (7), \(S_{ik,t+\delta}\) is an outcome related to school enrollment of child \(ik\) at date \(t+\delta\), \(\delta\) years after the grade repetition decision. Table 5 estimates this model with 8 variables. \(enrolled_{t+1}\) is the same variable as in Table 4, and Table 5, column 1 recalls the main quantitative results of Table 4.

\textsuperscript{16}Those who lived in 2003 in the same household than during the PASEC panel starting in 1995

\textsuperscript{17}Details are given in appendix D.1.
Figure 4: Simulations: effect of a decrease of last passer’s score on dropout probability

Notes: Simulations based on the estimates of Table 4. 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 “general equilibrium effect”: 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 “general equilibrium effect”: 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
Table 5: Effect of grade repetition on short-term and mid-term outcomes - Full Sample

<table>
<thead>
<tr>
<th></th>
<th>enrolled&lt;sub&gt;t+1&lt;/sub&gt;</th>
<th>enrolled&lt;sub&gt;t+2&lt;/sub&gt;</th>
<th>enrolled&lt;sub&gt;t+3&lt;/sub&gt;</th>
<th>enrolled&lt;sub&gt;t+4&lt;/sub&gt;</th>
<th>Still enrolled (2003)</th>
<th>Last Grade &gt; 5</th>
<th>Last Grade &gt; 6</th>
<th>Last Grade &gt; 7</th>
</tr>
</thead>
<tbody>
<tr>
<td>Last passer’s score (enrollment equation)</td>
<td>.379</td>
<td>.128</td>
<td>.087</td>
<td>.135</td>
<td>.249</td>
<td>.176</td>
<td>.296</td>
<td>.246</td>
</tr>
<tr>
<td>Grade repetition (enrollment equation)</td>
<td>-.2.259</td>
<td>-.609</td>
<td>.193</td>
<td>.072</td>
<td>-.442</td>
<td>-.423</td>
<td>-.668</td>
<td>-.525</td>
</tr>
<tr>
<td>Average marginal effect of repetition (enrollment eq.)</td>
<td>-.138</td>
<td>-.055</td>
<td>.027</td>
<td>.013</td>
<td>-.156</td>
<td>-.118</td>
<td>-.189</td>
<td>-.119</td>
</tr>
<tr>
<td>Obs.</td>
<td>1818</td>
<td>1449</td>
<td>1449</td>
<td>1449</td>
<td>1789</td>
<td>1777</td>
<td>1777</td>
<td>1777</td>
</tr>
<tr>
<td>log-likelihood</td>
<td>-1258.516</td>
<td>-1117.741</td>
<td>-1267.688</td>
<td>-1382.891</td>
<td>-2197.15</td>
<td>-1942.983</td>
<td>-2023.004</td>
<td>-1889.06</td>
</tr>
<tr>
<td>Simulation: LP&lt;sub&gt;ik&lt;/sub&gt; decreases by 0.25 pts with “general equilibrium effect”</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>“low-achievers”</td>
<td>-.005</td>
<td>-.002</td>
<td>-.001</td>
<td>-.0006</td>
<td>-.006</td>
<td>-.005</td>
<td>-.007</td>
<td>-.004</td>
</tr>
<tr>
<td>“leaning”</td>
<td>.012</td>
<td>.007</td>
<td>-.004</td>
<td>-.002</td>
<td>.018</td>
<td>.015</td>
<td>.018</td>
<td>.009</td>
</tr>
<tr>
<td>“good pupils”</td>
<td>.002</td>
<td>.0008</td>
<td>-.0003</td>
<td>-.0002</td>
<td>.002</td>
<td>.002</td>
<td>.003</td>
<td>.002</td>
</tr>
<tr>
<td>Simulation: LP&lt;sub&gt;ik&lt;/sub&gt; decreases by 0.25 pts without “general equilibrium effect”</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>“low-achievers”</td>
<td>-6.06e-06</td>
<td>-.005</td>
<td>-.007</td>
<td>-.015</td>
<td>-.006</td>
<td>-.014</td>
<td>-.011</td>
<td></td>
</tr>
<tr>
<td>“leaning”</td>
<td>-.016</td>
<td>-.006</td>
<td>-.004</td>
<td>-.008</td>
<td>-.026</td>
<td>-.017</td>
<td>-.021</td>
<td>-.011</td>
</tr>
<tr>
<td>“good pupils”</td>
<td>-.0007</td>
<td>-.001</td>
<td>-.003</td>
<td>-.006</td>
<td>-.018</td>
<td>-.009</td>
<td>-.020</td>
<td>-.015</td>
</tr>
</tbody>
</table>

Notes: The Table reports the results of trivariate specifications as in Table 4 (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 4.
b: The simulations assess the consequences of decreasing LP<sub>ik</sub> 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<sub>ik</sub> on grade repetition. It measures the sum of the direct effect of LP<sub>ik</sub> measured in the enrollment equation and of the indirect effect on enrollment due to “the effect of grade repetition on enrollment”
Table 5 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 at given grade. Hence, if grade repetition does not affect dropout date in the mid-term, it may still affect the last grade attended. Table 5 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. It must nevertheless be mentioned that our sample includes only pupils being in grade 2 in 1995-1996. Those who repeat no grade should therefore have achieved grade 9 at the end of school year 2002-2003. This means the relevance of our estimates is limited, as some pupils may finish grade 7 and grade 8 after 2003. The data collection of the follow-up survey being in 2003, no information posterior to 2003 is available.

The specifications in Table 5 also measure whether the last passer’s score 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, this would mitigate the benefits of fighting against grade repetition policies.

The simulations in Table 5 estimate the consequences of decreasing the last passer’s score by 0.25 pt. This represents a decrease in the share of failing students from 26% to 22% (see Table D.1).

The first simulation shows the benefits of decreasing the last passer’s score due to the effect of grade repetition on achievement. 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. This effect is concentrated among leaning pupils, who are the most likely to escape grade repetition due when the last passer’s score decrease.

The first simulation shows the benefits of decreasing the last passer’s score due to the effect of grade repetition on achievement, and assuming the coefficients in the first row of . 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. This effect is concentrated among leaning pupils, who are the most likely to escape grade repetition due when the last passer’s score decrease.

The second simulation in Figure Table 5 shows the consequences of a decrease of $LP_{ik}$ if we assume the general equilibrium effects are measured by the coefficients of $LP_{ik}$ in the first row of this Table. The results are in the opposite direction than the results of the first simulation. Indeed, the simulations find that decreasing $LP_{ik}$ would decrease the share of pupils enrolled in 2003 by 1.5 percentage points, decrease the share of pupils reaching grades 7 and 8 by 1.4 and 1.1 percentage points. This effect is concentrated among “low-achievers” and “good pupils”. Indeed, their probability of grade repetition is not much affected by decreasing $LP_{ik}$, while they are affected by the general equilibrium effects.

4.2 Split sample

Table 6 shows the results of some of the specifications in Table 5 when the sample is split in 2 groups. The first group includes the observations in the first two years of the panel. The results are given in columns 1 to column 4. The second group includes the grade observations in the last two years The results are given in columns 5 to column 8.

Table 6 shows that grade repetition is the most likely to affect long-term achievement in the last years of the panel. Indeed, in this sample, a grade repetition decreases the likelihood to be enrolled for 3 more years by 30%, the likelihood to be enrolled in 2003 by 23%, and the likelihood to start secondary school by 18%. However, the simulations show that decreasing grade repetition rates can still deteriorate long-term outcomes. Indeed, the simulation “with general equilibrium effect” find that the probability to be enrolled in 2003 and the probability to start secondary school decrease by 1.5 and 2.4 percentage points if $LP_{ik}$ decreases by 0.25 pt.
Table 6: Effect of grade repetition on short-term and mid-term outcomes - Split sample

<table>
<thead>
<tr>
<th>Last passer’s score (enrollment equation)</th>
<th>First 2 years of the panel</th>
<th>Last 2 years of the panel</th>
<th>Last Grade enrolled (2003)</th>
<th>Last Grade enrolled (2003)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>enrolled</strong>$_{t+1}$</td>
<td>(1)</td>
<td>(5)</td>
<td>.232</td>
<td>.367</td>
</tr>
<tr>
<td><strong>enrolled</strong>$_{t+3}$</td>
<td>(2)</td>
<td>(6)</td>
<td>.300</td>
<td>.439</td>
</tr>
<tr>
<td>Still enrolled</td>
<td>(3)</td>
<td>(7)</td>
<td>(.127)***</td>
<td>(.083)***</td>
</tr>
<tr>
<td>Still enrolled (2003)</td>
<td>(4)</td>
<td>(8)</td>
<td>(.130)***</td>
<td></td>
</tr>
<tr>
<td>Grade repetition (enrollment equation)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>.067</td>
<td>(.643)</td>
<td>-2.017</td>
<td>(-2.017)</td>
<td></td>
</tr>
<tr>
<td>-1.38</td>
<td>(.353)</td>
<td>-1.487</td>
<td>(-1.487)</td>
<td></td>
</tr>
<tr>
<td>-.164</td>
<td>(.337)</td>
<td>-.932</td>
<td>(-.932)</td>
<td></td>
</tr>
<tr>
<td>-.932</td>
<td>(.380)**</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Average marginal effect of repetition (enrollment eq.)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>.002</td>
<td>(.015)</td>
<td>-.136</td>
<td>(-.136)</td>
<td></td>
</tr>
<tr>
<td>.016</td>
<td>(.043)</td>
<td>-.298</td>
<td>(-.298)</td>
<td></td>
</tr>
<tr>
<td>-.058</td>
<td>(.180)</td>
<td>-.227</td>
<td>(-.227)</td>
<td></td>
</tr>
<tr>
<td>-.267</td>
<td>(.107)**</td>
<td>-.181</td>
<td>(-.181)</td>
<td></td>
</tr>
<tr>
<td>Obs.</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1099</td>
<td>873</td>
<td>1080</td>
<td>1073</td>
<td></td>
</tr>
<tr>
<td>log-likelihood</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>-668.289</td>
<td>-695.834</td>
<td>-1278.873</td>
<td>-1155.264</td>
<td></td>
</tr>
<tr>
<td><strong>χ^2</strong> instruments</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>16.409</td>
<td>36.077</td>
<td>15.498</td>
<td>21.088</td>
<td></td>
</tr>
<tr>
<td>corresponding p value</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>.089</td>
<td>.00008</td>
<td>.003</td>
<td>.011</td>
<td></td>
</tr>
<tr>
<td>Grade repetition decreases by 0.25 pts</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>“low-achievers”</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5.79e-06</td>
<td>(.0009)</td>
<td>-.004</td>
<td>(-.004)</td>
<td></td>
</tr>
<tr>
<td>“leaning”</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>-0003</td>
<td>(.003)</td>
<td>-.014</td>
<td>(-.014)</td>
<td></td>
</tr>
<tr>
<td>“good pupils”</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>-1.00e-05</td>
<td>(.0001)</td>
<td>-.001</td>
<td>(-.001)</td>
<td></td>
</tr>
<tr>
<td>Simulation: (LP_{ik}) decreases by 0.25 pts <em>no general equilibrium effect</em></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Full Sample</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>-0.00009</td>
<td>(.00009)</td>
<td>.003</td>
<td>.011</td>
<td></td>
</tr>
<tr>
<td>Simulation: (LP_{ik}) decreases by 0.25 pts <em>with general equilibrium effect</em></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>“low-achievers”</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>-5.21e-06</td>
<td>(.0002)</td>
<td>-.007</td>
<td>(-.007)</td>
<td></td>
</tr>
<tr>
<td>“leaning”</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>-0.003</td>
<td>(.003)</td>
<td>.008</td>
<td>.029</td>
<td></td>
</tr>
<tr>
<td>“good pupils”</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>-1.00e-05</td>
<td>(.0001)</td>
<td>.0008</td>
<td>.005</td>
<td></td>
</tr>
<tr>
<td>Simulation: (LP_{ik}) decreases by 0.25 pts <em>no general equilibrium effect</em></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Full Sample</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>-0.0007</td>
<td>(.0002)</td>
<td>-.014</td>
<td>(-.014)</td>
<td></td>
</tr>
</tbody>
</table>
| Notes: The Table reports the results of trivariate specifications as in Table 4 (model (5)), with different dependant variables in the enrollment equation, and splitting the sample. *** , ** , 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: GHK simulator is run with 50 iterations for convergence reasons. 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”
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 are negatively affected by grade repetition.

However, grade repetition policies can indeed 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 effects. Schools with tough grade repetition policies emphasize similar or better 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 can improve global academic outcomes would be based on incomplete evidences.

References


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


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 Explanatory variables in main regressions

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. 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.

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.

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

---

18The 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>
<thead>
<tr>
<th>Variable</th>
<th>N</th>
<th>mean</th>
<th>standard deviation</th>
<th>min.</th>
<th>max.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Enrolled next year</td>
<td>1823</td>
<td>.979</td>
<td></td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>(enrolled_{t+2})</td>
<td>1454</td>
<td>.959</td>
<td></td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>(enrolled_{t+3})</td>
<td>1454</td>
<td>.920</td>
<td></td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>(enrolled_{t+4})</td>
<td>1454</td>
<td>.882</td>
<td></td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>Still enrolled (2003)</td>
<td>1794</td>
<td>.673</td>
<td></td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>Last Grade &gt; 5</td>
<td>1782</td>
<td>.749</td>
<td></td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>Last Grade &gt; 6</td>
<td>1782</td>
<td>.437</td>
<td></td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>Last Grade &gt; 7</td>
<td>1782</td>
<td>.291</td>
<td></td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>Grade repetition</td>
<td>1823</td>
<td>.148</td>
<td></td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>Selection (on grade repetition observation)</td>
<td>1823</td>
<td>.867</td>
<td></td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>Negative shocks on harvests</td>
<td>1823</td>
<td>.087</td>
<td>.308</td>
<td>0</td>
<td>2</td>
</tr>
<tr>
<td>Test score</td>
<td>1823</td>
<td>-0.055</td>
<td>.953</td>
<td>-3.20</td>
<td>3.34</td>
</tr>
<tr>
<td>Last passer’s test score</td>
<td>1818</td>
<td>-0.770</td>
<td>.849</td>
<td>-3.20</td>
<td>2.63</td>
</tr>
<tr>
<td>Previous year’s test score</td>
<td>1823</td>
<td>.00481</td>
<td>1.01</td>
<td>-2.34</td>
<td>3.81</td>
</tr>
<tr>
<td>Group mean test score</td>
<td>1823</td>
<td>-0.0709</td>
<td>.536</td>
<td>-1.58</td>
<td>1.91</td>
</tr>
<tr>
<td>Household wealth</td>
<td>1823</td>
<td>-0.591</td>
<td>2.07</td>
<td>-3.12</td>
<td>4.38</td>
</tr>
<tr>
<td>Parent’s education</td>
<td>1823</td>
<td>2.05</td>
<td>1.47</td>
<td>1</td>
<td>8</td>
</tr>
<tr>
<td>Head is not Muslim (Christian or Animist)</td>
<td>1820</td>
<td>0.0335</td>
<td></td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>Ethnic group of the head: Wolof</td>
<td>1813</td>
<td>0.398</td>
<td></td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>Ethnic group of the head: Pulaar-Halpulaar</td>
<td>1813</td>
<td>0.187</td>
<td></td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>Ethnic group of the head: Serere</td>
<td>1813</td>
<td>0.250</td>
<td></td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>Ethnic group of the head: Dioola</td>
<td>1813</td>
<td>0.0281</td>
<td></td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>Ethnic group of the head: Mandingue-Sose</td>
<td>1813</td>
<td>0.102</td>
<td></td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>Ethnic group of the head: Others</td>
<td>1813</td>
<td>0.0342</td>
<td></td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>Community: mean asset index</td>
<td>1823</td>
<td>0.0884</td>
<td>1.68</td>
<td>-2.29</td>
<td>3.33</td>
</tr>
<tr>
<td>Community: mean education index</td>
<td>1823</td>
<td>2.20</td>
<td>.730</td>
<td>1.19</td>
<td>4</td>
</tr>
<tr>
<td>(\log(\text{Village or city population}))</td>
<td>1714</td>
<td>10.0</td>
<td>2.81</td>
<td>5.60</td>
<td>14.6</td>
</tr>
<tr>
<td>Community main activity: trade (ref:agri.)</td>
<td>1823</td>
<td>0.404</td>
<td></td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>Community has electricity</td>
<td>1823</td>
<td>0.813</td>
<td></td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>Rural</td>
<td>1823</td>
<td>0.535</td>
<td></td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>Distance to health center</td>
<td>1823</td>
<td>0.0730</td>
<td>.260</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>Distance to hospital</td>
<td>1823</td>
<td>1.63</td>
<td>1.26</td>
<td>0</td>
<td>3</td>
</tr>
</tbody>
</table>

Notes: Standard deviations are not reported for binary variables.
Table A.2: Grade attended during the PASEC panel for six imaginary cases

<table>
<thead>
<tr>
<th>case 1</th>
<th>case 2</th>
<th>case 3</th>
<th>case 4</th>
<th>case 5</th>
<th>case 6</th>
<th>school year</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>1995 - 1996</td>
</tr>
<tr>
<td>2</td>
<td>2,3</td>
<td>drop.</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>1996 - 1997</td>
</tr>
<tr>
<td>3</td>
<td>3</td>
<td>3,4</td>
<td>4</td>
<td>4</td>
<td>3</td>
<td>1997 - 1998</td>
</tr>
<tr>
<td>4</td>
<td>4</td>
<td>3,4,5</td>
<td>5</td>
<td>5</td>
<td>3</td>
<td>1998 - 1999</td>
</tr>
<tr>
<td>5</td>
<td>5</td>
<td>3,4,5,6</td>
<td>6</td>
<td>6</td>
<td>4</td>
<td>1999 - 2000</td>
</tr>
</tbody>
</table>

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

education was missing, it is replaced by the mean education of the other adults (aged more than 25 in 1995) in the household.

**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) However, for all the specifications presented, including a dummy for those cases did not change the effect of grade repetition on school dropouts.

A group is defined by all the children being in the same school and the same grade in a given school year. The peers of a child are the other pupils of his group than himself. Among the peers of a given child a given year, “passers” are those admitted to the next grade. Others must repeat their grade if they do not drop out and are called “repeaters”. Among the passers, the “last passer” is the passer with the lowest test score.

**Last passer’s test score** is another proxy for the teacher specific attitude to repetition. In fact, if the last passer’s score is high, a given child is expected to repeat more frequently.

**A.3 Other variables**

**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.

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

---

19 A group is an approximation of a class: there may be several classes per group in some cases. In fact, there may be several classes per grade in some schools. In that case, although all the pupils are in the same class in the first year of the panel, in the following years they may be in the same grade and in different classes.
Community has electricity 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

B Proofs for the semiparametric identification of model (1)

B.1 model (1)

This section proves that model (1) can be semiparametrically identified.

The model (1) is:

\[
\begin{align*}
  r &= \mathbb{I}(X\beta_r + \gamma_r Z_1 + \varepsilon_r > 0) \\
  s &= \mathbb{I}(X\beta_s + \gamma_s Z_2 + \alpha_s > 0) \\
  e &= \mathbb{I}(X\beta_e + \gamma_e Z_2 + \alpha_e r + \varepsilon_e > 0)
\end{align*}
\]  \hspace{1cm} (B.1)

(For simplicity r is repetition, s is selection, and e is enrolled, \(e\). 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 distribution function. This idea is used to show that all the parameters of model (1) are identified without any parametric assumption on \(f(\varepsilon_r, \varepsilon_s, \varepsilon_e)\).

\(\Theta\) 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 \((X_0, Z_{10}, Z_{20}), (X, Z_1, Z_2) \in \Theta\)
   
   \[
   \begin{pmatrix}
   \frac{dP(r=1,s=1)}{dz_1}(X_0, Z_{10}, Z_{20}) \\
   \frac{dP(r=1,s=1)}{dz_2}(X_0, Z_{10}, Z_{20}) \\
   \frac{dP(r=0,s=1)}{dz_1}(X_0, Z_{10}, Z_{20}) \\
   \frac{dP(r=0,s=1)}{dz_2}(X_0, Z_{10}, Z_{20})
   \end{pmatrix}
   \]

   has full rank

   (b) \(\forall (X, Z_1, Z_2)\) 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) \[
   \begin{align*}
   X_a\beta_r + \gamma_r Z_{1a} &= X_b\beta_r + \gamma_r Z_{1b} \\
   X_a\beta_s + \gamma_s Z_{2a} + \alpha_s &= X_b\beta_s + \gamma_s Z_{2b} \\
   X_a\beta_e + \gamma_e Z_{2a} + \alpha_e &= X_b\beta_e + \gamma_e Z_{2b}
   \end{align*}
   \]

   (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 (1), only the signs of the latent variables \((X\beta_r + \gamma_r Z_1 + \epsilon_r, X\beta_s + \gamma_s Z_2 + \alpha_s r + \epsilon_s, X\beta_e + \gamma_e Z_2 + \alpha_e r + \epsilon_e)\) 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 co-variates are not null in \((X_0, Z_{10}, Z_{20})\).

Assumption 5 ensures the support \(\Theta\) 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 \(\beta\) and \(\gamma\) of the first two equations of model (1) are identified, second, it is shown that the coefficients \(\beta\) and \(\gamma\) of the last equation are identified, and finally, it is shown that the \(\alpha\) are identified.

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

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

\[
P^{(11)} = P(r = 1, s = 1|X, Z_1, Z_2)
= \int_{-\infty}^{\infty} X\beta_r - \gamma_r Z_1 \int_{-\infty}^{\infty} \int_{\mathbb{R}} f(\epsilon_r, \epsilon_s, \epsilon_e) d\epsilon_r d\epsilon_s d\epsilon_e
= P^{(11)}(-X\beta_r - \gamma_r Z_1, -X\beta_s - \gamma_s Z_2 - \alpha_s)
\]

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

\[
\frac{dP^{(11)}}{dX_1} = F_1^{(11)} + F_2^{(11)} \quad (B.2)
\]

\[
\frac{dP^{(11)}}{dX_i} = \beta_{ri}F_1^{(11)} + \beta_{si}F_2^{(11)} \quad (\forall i \in \{1..K\}) \quad (B.3)
\]

\[
\frac{dP^{(11)}}{dZ_1} = \gamma_r F_1^{(11)} \quad (B.4)
\]

\[
\frac{dP^{(11)}}{dZ_2} = \gamma_s F_2^{(11)} \quad (B.5)
\]

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

\[
P^{(01)} = P(r = 0, s = 1|X, Z_1, Z_2)
= \int_{-\infty}^{\infty} X\beta_r - \gamma_r Z_1 \int_{-\infty}^{\infty} \int_{\mathbb{R}} f(\epsilon_r, \epsilon_s, \epsilon_e) d\epsilon_r d\epsilon_s d\epsilon_e
= P^{(01)}(-X\beta_r - \gamma_r Z_1, -X\beta_s - \gamma_s Z_2)
\]
We note $F_1^{(01)}$ and $F_2^{(01)}$ the derivatives of $F^{(01)}$ towards its two arguments.

\[
\begin{align*}
\frac{dP^{(01)}}{dX_1} &= F_1^{(01)} + F_2^{(01)} \\
\frac{dP^{(01)}}{dX_i} &= \beta_{ri} F_1^{(01)} + \beta_{si} F_2^{(01)} \\
\frac{dP^{(01)}}{dZ_1} &= \gamma_r F_1^{(01)} \\
\frac{dP^{(01)}}{dZ_2} &= \gamma_s F_2^{(01)}
\end{align*}
\]

(B.6)

(B.7)

(B.8)

(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:

\[
\begin{cases}
\frac{dP^{(11)}}{dX_1} = \frac{1}{\gamma_r} \frac{dP^{(11)}}{dZ_1} + \frac{1}{\gamma_s} \frac{dP^{(11)}}{dZ_2} \\
\frac{dP^{(01)}}{dX_1} = \frac{1}{\gamma_r} \frac{dP^{(01)}}{dZ_1} \end{cases}
\]

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

\[
\begin{cases}
\frac{dP^{(11)}}{dX_1} = \beta_{ri} F_1^{(11)} + \beta_{si} F_2^{(11)} \\
\frac{dP^{(01)}}{dX_1} = \beta_{ri} F_1^{(01)} + \beta_{si} F_2^{(01)}
\end{cases}
\]

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

\begin{itemize}
  \item \textbf{Identification of the third equation}
  
  We compute the derivatives of $\mathbb{P}(e = 1|X, Z_1, Z_2)$:

  \[
  P^{(1)} = \mathbb{P}(e = 1|X, Z_1, Z_2) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f(\varepsilon_r, \varepsilon_s, \varepsilon_e) d\varepsilon_r d\varepsilon_s d\varepsilon_e
  
  + \int_{-\infty}^{\infty} \int_{-\infty}^{\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_r - \gamma_r Z_2 - \alpha_e)
  \]

  We call $F_1^{(11)}$, $F_2^{(11)}$ and $F_3^{(1)}$ the derivatives of $F^{(1)}$ with respect to its arguments. We compute the derivatives of $P^{(1)}$:}

\end{itemize}
\[
\frac{dP^{(1)}}{dX_1} = F'_1(1) + F'_2(1) \quad (B.10)
\]
\[
\frac{dP^{(1)}}{dX_i} = \beta_{ri} F'_1(1) + \beta_{si} F'_2(1) \quad (B.11)
\]
\[
\frac{dP^{(1)}}{dZ_1} = \gamma_r F'_1(1) \quad (B.12)
\]
\[
\frac{dP^{(1)}}{dZ_2} = \gamma_e F'_2(1) \quad (B.13)
\]

\(\gamma_e\) 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 \(\gamma_e\) is identified by (B.13). Knowledge of \(\beta_{ri}, F'_1(1)\) and \(F'_2(1)\) identifies \(\beta_{si}\) in (B.11).

- **Identification of \(\alpha_s\).**

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\(^{20}\)

\[
\begin{aligned}
X_a \beta_r + \gamma_r Z_{1a} &= X_b \beta_r + \gamma_r Z_{1b} = \kappa_{r1} \\
X_c \beta_r + \gamma_r Z_{1c} &= X_d \beta_r + \gamma_r Z_{1d} = \kappa_{r2} \\
X_a \beta_s + \gamma_s Z_{2c} &= X_c \beta_s + \gamma_s Z_{2c} = \kappa_{s1} \\
X_b \beta_s + \gamma_s Z_{2b} &= X_d \beta_s + \gamma_s Z_{2d} = \kappa_{s2}
\end{aligned}
\]

\[
\begin{aligned}
\mathbb{P}(r=1, s=1 | a) &= \mathbb{P}(r=1, s=1 | b) \\
\mathbb{P}(r=0, s=1 | c) &= \mathbb{P}(r=0, s=1 | d) \quad \Rightarrow \quad \mathbb{P}(r=1, s=1 | b) - \mathbb{P}(r=0, s=1 | d) = \mathbb{P}(s=1 | b) - \mathbb{P}(s=1 | d) \\
\Rightarrow \quad \kappa_{s1} + \alpha_s = \kappa_{s2}
\end{aligned}
\]

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

Then

\[
\begin{aligned}
\mathbb{P}(r=1, s=1 | a) - \mathbb{P}(r=1, s=1 | c) &= \mathbb{P}(r=0, s=1 | b) - \mathbb{P}(r=0, s=1 | d) \\
\Rightarrow \quad \kappa_{s1} + \alpha_s = \kappa_{s2}
\end{aligned}
\]

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

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

because

\(^{20}\)\(\hat{\mathbb{P}}\) means that the probability is net of the effect of \(r\) on \(o\).
\[ \mathbb{P}(r = 1, s = 1|a) + \mathbb{P}(r = 0, s = 1|b) = \int_{-\infty}^{\kappa_{s2}} \int_{-\kappa_{s1} - \alpha_s}^{\kappa_{s2}} \int_{-\kappa_{s1}}^{\infty} f(\varepsilon_r, \varepsilon_s, \varepsilon_e) d\varepsilon_r d\varepsilon_s d\varepsilon_e \\
+ \int_{-\kappa_{s1}}^{\kappa_{s1} - \alpha_s} f(\varepsilon_r, \varepsilon_s, \varepsilon_e) d\varepsilon_r d\varepsilon_s d\varepsilon_e \\
= \int_{\mathbb{R}} \int_{-\kappa_{s1}}^{\kappa_{s2}} \int_{-\kappa_{s1} - \alpha_s}^{\infty} f(\varepsilon_r, \varepsilon_s, \varepsilon_e) d\varepsilon_r d\varepsilon_s d\varepsilon_e \\
= \mathbb{P}(s = 1|b) \]

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

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

**Proof of equation (B.15):**

We write the probabilities:

\[ \mathbb{P}(r = 1, s = 1|\kappa_r, \kappa_s) = \int_{-\kappa_r}^{\kappa_{s2}} \int_{-\kappa_{s1} - \alpha_s}^{\kappa_{s2}} \int_{-\kappa_{s1}}^{\infty} f(\varepsilon_r, \varepsilon_s, \varepsilon_e) d\varepsilon_r d\varepsilon_s d\varepsilon_e \]

\[ \mathbb{P}(r = 0, s = 1|\kappa_r, \kappa_s) = \int_{-\kappa_r}^{\kappa_{s2}} \int_{-\kappa_{s1} - \alpha_s}^{\kappa_{s2}} \int_{-\kappa_{s1}}^{\infty} f(\varepsilon_r, \varepsilon_s, \varepsilon_e) d\varepsilon_r d\varepsilon_s d\varepsilon_e \]

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

\[ \mathbb{P}(r = 1, s = 1|a) - \mathbb{P}(r = 1, s = 1|c) = \int_{-\kappa_r}^{\kappa_{s2}} \int_{\kappa_{s1} - \alpha_s}^{\kappa_{s1}} \int_{-\kappa_{s1}}^{\infty} f(\varepsilon_r, \varepsilon_s, \varepsilon_e) d\varepsilon_r d\varepsilon_s d\varepsilon_e \]

\[ \mathbb{P}(r = 0, s = 1|b) - \mathbb{P}(r = 0, s = 1|d) = \int_{-\kappa_r}^{\kappa_{s2}} \int_{\kappa_{s1} - \alpha_s}^{\kappa_{s2}} \int_{-\kappa_{s1}}^{\infty} f(\varepsilon_r, \varepsilon_s, \varepsilon_e) d\varepsilon_r d\varepsilon_s d\varepsilon_e \]

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

\[ \mathbb{P}(r = 1, s = 1|a) - \mathbb{P}(r = 1, s = 1|c) = -[\mathbb{P}(r = 0, s = 1|b) - \mathbb{P}(r = 0, s = 1|d)] \]
\[ \Leftrightarrow \int_{-\kappa_r}^{\kappa_{s2}} \int_{-\kappa_{s1} - \alpha_s}^{\kappa_{s1}} \int_{\kappa_{s1} - \alpha_s}^{\kappa_{s2}} f(\varepsilon_r, \varepsilon_s, \varepsilon_e) d\varepsilon_s d\varepsilon_e = 0 \]
\[ \Leftrightarrow \int_{-\kappa_r}^{\kappa_{s2}} \int_{\kappa_{s1} - \alpha_s}^{\kappa_{s2}} f(\varepsilon_r, \varepsilon_s, \varepsilon_e) d\varepsilon_s d\varepsilon_e = 0 \]

\( f(\varepsilon_r, \varepsilon_s, \varepsilon_e) > 0 \) 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 $\Theta$. 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 $\alpha_e$.**

  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

  $$\begin{align*}
  X_a \beta_r + \gamma_r Z_{1a} &= X_b \beta_r + \gamma_r Z_{1b} = \kappa_{r1} \\
  X_c \beta_r + \gamma_r Z_{1c} &= X_d \beta_r + \gamma_r Z_{1d} = \kappa_{r2} \\
  X_a \beta_s + \gamma_s Z_{1a} &= X_c \beta_s + \gamma_s Z_{1c} = \kappa_{s1} \\
  X_b \beta_s + \gamma_s Z_{1b} &= X_d \beta_s + \gamma_s Z_{1d} = \kappa_{s2} \\
  X_a \beta_e + \gamma_e Z_{2a} &= X_c \beta_e + \gamma_e Z_{2c} = \kappa_{e1} \\
  X_b \beta_e + \gamma_e Z_{2b} &= X_d \beta_e + \gamma_e Z_{2d} = \kappa_{e2}
  \end{align*}$$

  \hspace{1cm} (B.16)

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

  Then

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

  \hspace{1cm} (B.17)

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

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

  \hspace{1cm} Proof of equation (B.17):

  We write the probabilities:

  $$\begin{align*}
  \mathbb{P}(r = 1, s = 1, e = 1|a) &= \int_{-\kappa_{r1}}^{\infty} \int_{-\kappa_{s1} - \alpha_s}^{\infty} \int_{-\kappa_{e1} - \alpha_e}^{\infty} f(\varepsilon_r, \varepsilon_s, \varepsilon_e) d\varepsilon_r d\varepsilon_s d\varepsilon_e \\
  \mathbb{P}(r = 1, s = 1, e = 1|c) &= \int_{-\kappa_{r2}}^{\infty} \int_{-\kappa_{s1} - \alpha_s}^{\infty} \int_{-\kappa_{e1} - \alpha_e}^{\infty} f(\varepsilon_r, \varepsilon_s, \varepsilon_e) d\varepsilon_r d\varepsilon_s d\varepsilon_e \\
  \mathbb{P}(r = 0, s = 1, e = 1|b) &= \int_{-\kappa_{r1}}^{\infty} \int_{-\kappa_{s1} - \alpha_s}^{\infty} \int_{-\kappa_{e1} - \alpha_e}^{\infty} f(\varepsilon_r, \varepsilon_s, \varepsilon_e) d\varepsilon_r d\varepsilon_s d\varepsilon_e \\
  \mathbb{P}(r = 0, s = 1, e = 1|d) &= \int_{-\kappa_{r2}}^{\infty} \int_{-\kappa_{s1} - \alpha_s}^{\infty} \int_{-\kappa_{e1} - \alpha_e}^{\infty} f(\varepsilon_r, \varepsilon_s, \varepsilon_e) d\varepsilon_r d\varepsilon_s d\varepsilon_e
  \end{align*}$$
Then we can easily compute the differences of (B.17):

\[
\mathbb{P}(r = 1, s = 1, e = 1 | a) - \mathbb{P}(r = 1, s = 1, e = 1 | c) = \int_{-\kappa_{r1}}^{\kappa_{r2}} \int_{-\kappa_{s1} - \alpha_s}^{\kappa_{s2} - \alpha_s} \int_{-\kappa_{e1} - \alpha_e}^{\kappa_{e2} + \alpha_e} f(\varepsilon_r, \varepsilon_s, \varepsilon_e) d\varepsilon_r d\varepsilon_s d\varepsilon_e
\]

\[
\mathbb{P}(r = 0, s = 1, e = 1 | b) - \mathbb{P}(r = 0, s = 1, e = 1 | d) = \int_{-\kappa_{r1}}^{\kappa_{r2}} \int_{-\kappa_{s2}}^{\kappa_{s2} - \alpha_s} \int_{-\kappa_{e2}}^{\kappa_{e2} - \alpha_e} f(\varepsilon_r, \varepsilon_s, \varepsilon_e) d\varepsilon_r d\varepsilon_s d\varepsilon_e
\]

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

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

\[
\Leftrightarrow \int_{-\kappa_{r1}}^{\kappa_{r2}} \int_{-\kappa_{s1} - \alpha_s}^{\kappa_{s2} - \alpha_s} \int_{-\kappa_{e1} - \alpha_e}^{\kappa_{e2} + \alpha_e} f(\varepsilon_r, \varepsilon_s, \varepsilon_e) d\varepsilon_r d\varepsilon_s d\varepsilon_e = 0
\]

\(f(\varepsilon_r, \varepsilon_s, \varepsilon_e) > 0\) in the neighborhood of any point of \(\Theta\) (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 \(\alpha_e\) can be identified.

**B.2 Model (1) without \(Z_2\)**

This appendix proves that \(Z_2\) is unnecessary for identifying the sign of \(\alpha_e\). Accordingly, it is theoretically not necessary to control for selection to identify the sign of \(\alpha_e\) semiparametrically.

The corresponding model is:

\[
\begin{align*}
 r & = \mathbb{I}(X\beta_r + \gamma_r Z + \varepsilon_r > 0) \\
 s & = \mathbb{I}(X\beta_s + \alpha_s r + \varepsilon_s > 0) \\
 e & = \mathbb{I}(X\beta_e + \alpha_e r + \varepsilon_e > 0)
\end{align*}
\]

(For simplicity \(r\) is repetition, \(s\) is selection, and \(e\) is enrolled \(+1\). 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 \(\alpha_e\) is identified in model (B.18) without any parametric assumption on \(f(\varepsilon_r, \varepsilon_s, \varepsilon_e)\). \(\Theta\) 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) \( f(e_r, e_s, e_e) > 0 \) 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 \( \Gamma \)

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 \( f(e_r, e_s, e_e) \).

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 \( e_e \) 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 \( \gamma_r \) 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:

\[
\mathbb{P}(r = 1, s = 1, e = 1|X, Z) = \int_{-X_0\beta_s - \alpha_s}^{\infty} \int_{-X_0\beta_e - \alpha_e}^{\infty} f(e_r, e_s, e_e) d\epsilon_r d\epsilon_s d\epsilon_e
\]

\[
\Rightarrow d\mathbb{P}(r = 1, s = 1, e = 1|X, Z)/dZ = \gamma_r \int_{-X_0\beta_s - \alpha_s}^{\infty} \int_{-X_0\beta_e - \alpha_e}^{\infty} f(-X\beta_r - \gamma_r Z, e_s, e_e) d\epsilon_s d\epsilon_e
\]

Assumption 3b ensures that:

\[
\int_{-X_0\beta_s - \alpha_s}^{\infty} \int_{-X_0\beta_e - \alpha_e}^{\infty} f(-X_0\beta_r - \gamma_r Z_0, e_s, e_e) d\epsilon_s d\epsilon_e \leq \int_{\mathbb{R}} \int_{\mathbb{R}} f(-X_0\beta_r - \gamma_r Z_0, e_s, e_e) d\epsilon_s d\epsilon_e < \infty
\]

And assumption 3c ensures that:

\[
\int_{-X_0\beta_s - \alpha_s, \infty}^{\infty} \int_{-X_0\beta_e - \alpha_e, \infty}^{\infty} f(-X_0\beta_r - \gamma_r Z_0, e_s, e_e) d\epsilon_s d\epsilon_e \\
\geq \int_{([-X_0\beta_s - \alpha_s, \infty] \times [-X_0\beta_e - \alpha_e, \infty]) \cap \Gamma} f(-X_0\beta_r - \gamma_r Z_0, e_s, e_e) d\epsilon_s d\epsilon_e > 0
\]

That is why

\[
0 < \int_{-X_0\beta_s - \alpha_s}^{\infty} \int_{-X_0\beta_e - \alpha_e}^{\infty} f(-X_0\beta_r - \gamma_r Z_0, e_s, e_e) d\epsilon_s d\epsilon_e < \infty
\]

so that \( d\mathbb{P}(r = 1, s = 1, e = 1|X, Z) \) (\( X_0, Z_0 \)) has the same sign as \( \gamma_r \).
– Proof that the sign of $\alpha_e$ is identified

Now, let us focus on $\mathbb{P}(e = 1|X, Z)$:

$$\mathbb{P}(e = 1|X, Z) = \mathbb{P}(e = 1, r = 1|X, Z) + \mathbb{P}(e = 1, r = 0|X, Z)$$

$$= \int_{-\infty}^{\infty} -X\beta_e - \gamma_r Z \int_{\mathbb{R}} \int_{-X\beta_e - \alpha_e} \int_{\mathbb{R}} \int_{-X\beta_e} f(\varepsilon_r, \varepsilon_s, \varepsilon_e) d\varepsilon_r d\varepsilon_s d\varepsilon_e$$

$$+ \int_{-\infty}^{\infty} -X\beta_e - \gamma_r Z \int_{\mathbb{R}} \int_{-X\beta_e} f(\varepsilon_r, \varepsilon_s, \varepsilon_e) d\varepsilon_r d\varepsilon_s d\varepsilon_e$$

$$= \int_{\mathbb{R}} \int_{\mathbb{R}} \int_{-X\beta_e} f(\varepsilon_r, \varepsilon_s, \varepsilon_e) d\varepsilon_r d\varepsilon_s d\varepsilon_e$$

$$+ \int_{-\infty}^{\infty} -X\beta_e - \gamma_r Z \int_{\mathbb{R}} \int_{-X\beta_e - \alpha_e} f(\varepsilon_r, \varepsilon_s, \varepsilon_e) d\varepsilon_r d\varepsilon_s d\varepsilon_e$$

$$\Rightarrow d\mathbb{P}(e = 1|X, Z)/dZ = \gamma_r \int_{\mathbb{R}} \int_{-X\beta_e - \alpha_e} f(-X\beta_r - \gamma_r Z, \varepsilon_s, \varepsilon_e) d\varepsilon_r d\varepsilon_s d\varepsilon_e$$

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

#### C.1 Determinants of $LP_{ik}$

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

Additional covariates: 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 community.
C.2 Reduced form with non-linear controls

Table C.2 and Figure C.1 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.2 are similar to Table 3. Neither of the dummies set is jointly significant. If something, Figure C.1 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.1: Comparison between the estimates of Table 3 and Table C.2

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

<table>
<thead>
<tr>
<th></th>
<th>repetition</th>
<th>enrolled</th>
<th>selection</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>(1)</td>
<td>(2)</td>
<td>(3)</td>
</tr>
<tr>
<td>Test score</td>
<td>.093</td>
<td>-.282</td>
<td>.072</td>
</tr>
<tr>
<td></td>
<td>(.495)</td>
<td>(.311)</td>
<td>(.255)</td>
</tr>
<tr>
<td>Last passer’s test score</td>
<td>-.098</td>
<td>.613</td>
<td>.110</td>
</tr>
<tr>
<td></td>
<td>(.235)</td>
<td>(.269)**</td>
<td>(.188)</td>
</tr>
<tr>
<td>$S^*<em>{ik} - LP</em>{ik} &lt; -1$</td>
<td>.823</td>
<td>-.804</td>
<td>-.189</td>
</tr>
<tr>
<td></td>
<td>(.492)*</td>
<td>(.547)</td>
<td>(.346)</td>
</tr>
<tr>
<td>$-1 &lt; S^*<em>{ik} - LP</em>{ik} &lt; -0.75$</td>
<td>.938</td>
<td>-.290</td>
<td>-.595</td>
</tr>
<tr>
<td></td>
<td>(.400)**</td>
<td>(.447)**</td>
<td>(.289)**</td>
</tr>
<tr>
<td>$-0.75 &lt; S^*<em>{ik} - LP</em>{ik} &lt; -0.5$</td>
<td>.495</td>
<td>-.858</td>
<td>-.402</td>
</tr>
<tr>
<td></td>
<td>(.309)</td>
<td>(.355)**</td>
<td>(.247)</td>
</tr>
<tr>
<td>$-0.5 &lt; S^*<em>{ik} - LP</em>{ik} &lt; -0.25$</td>
<td>.391</td>
<td>-.320</td>
<td>(.166)*</td>
</tr>
<tr>
<td></td>
<td>(.184)</td>
<td>(.386)</td>
<td></td>
</tr>
<tr>
<td>$-0.25 &lt; S^*<em>{ik} - LP</em>{ik} &lt; 0$</td>
<td>.127</td>
<td>.391</td>
<td>-.320</td>
</tr>
<tr>
<td></td>
<td>(.184)</td>
<td>(.386)</td>
<td>(.166)*</td>
</tr>
<tr>
<td>$0 &lt; S^*<em>{ik} - LP</em>{ik} &lt; 0.25$</td>
<td>Ref.</td>
<td>Ref.</td>
<td>Ref.</td>
</tr>
<tr>
<td>$0.25 &lt; S^*<em>{ik} - LP</em>{ik} &lt; 0.5$</td>
<td>-.381</td>
<td>.375</td>
<td>.230</td>
</tr>
<tr>
<td></td>
<td>(.185)**</td>
<td>(.326)</td>
<td>(.175)</td>
</tr>
<tr>
<td>$0.5 &lt; S^*<em>{ik} - LP</em>{ik} &lt; 0.75$</td>
<td>-.431</td>
<td>.403</td>
<td>.303</td>
</tr>
<tr>
<td></td>
<td>(.231)*</td>
<td>(.287)</td>
<td>(.205)</td>
</tr>
<tr>
<td>$0.75 &lt; S^*<em>{ik} - LP</em>{ik} &lt; 1$</td>
<td>-.500</td>
<td>.696</td>
<td>.222</td>
</tr>
<tr>
<td></td>
<td>(.275)*</td>
<td>(.323)**</td>
<td>(.219)</td>
</tr>
<tr>
<td>$1 &lt; S^*<em>{ik} - LP</em>{ik} &lt; 1.5$</td>
<td>-.400</td>
<td>1.541</td>
<td>.307</td>
</tr>
<tr>
<td></td>
<td>(.333)</td>
<td>(.527)**</td>
<td>(.242)</td>
</tr>
<tr>
<td>$1.5 &lt; S^*<em>{ik} - LP</em>{ik}$</td>
<td>-.537</td>
<td>1.619</td>
<td>.457</td>
</tr>
<tr>
<td></td>
<td>(.507)</td>
<td>(.546)**</td>
<td>(.343)</td>
</tr>
<tr>
<td>Group mean test score</td>
<td>.419</td>
<td>-.671</td>
<td>-.240</td>
</tr>
<tr>
<td></td>
<td>(.378)</td>
<td>(.478)</td>
<td>(.314)</td>
</tr>
<tr>
<td>Negative shock on harvests</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>.156</td>
<td>.396</td>
<td>(.196)</td>
</tr>
<tr>
<td></td>
<td>(.147)***</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\chi^2$ test score dummies$^a$</td>
<td>6.982</td>
<td>7.742</td>
<td>.592</td>
</tr>
<tr>
<td>corresponding p value</td>
<td>.137</td>
<td>.101</td>
<td>.964</td>
</tr>
<tr>
<td>$\chi^2$ difference to group mean dummies$^b$</td>
<td>1.416</td>
<td>5.245</td>
<td>1.412</td>
</tr>
<tr>
<td>corresponding p value</td>
<td>.841</td>
<td>.263</td>
<td>.842</td>
</tr>
<tr>
<td>$\chi^2$ group mean dummies$^c$</td>
<td>5.695</td>
<td>6.128</td>
<td>2.189</td>
</tr>
<tr>
<td>corresponding p value</td>
<td>.127</td>
<td>.106</td>
<td>.534</td>
</tr>
<tr>
<td>$\chi^2$ last passer’s score dummies$^d$</td>
<td>7.539</td>
<td>1.984</td>
<td>7.539</td>
</tr>
<tr>
<td>corresponding p value</td>
<td>.184</td>
<td>.851</td>
<td>.184</td>
</tr>
<tr>
<td>Household wealth and Parents’ education, Previous year’s test score</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td></td>
</tr>
<tr>
<td>Grade*year dummies</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Obs.</td>
<td>1818</td>
<td>1818</td>
<td>1818</td>
</tr>
<tr>
<td>log likelihood</td>
<td>-1237.971</td>
<td>-1237.971</td>
<td>-1237.971</td>
</tr>
<tr>
<td>$\chi^2$ grade year dummies</td>
<td>4.958</td>
<td>29.208</td>
<td>9.239</td>
</tr>
<tr>
<td>corresponding p value</td>
<td>.292</td>
<td>7.09e-06</td>
<td>.055</td>
</tr>
<tr>
<td>$\chi^2$ exclusion restriction</td>
<td></td>
<td></td>
<td>7.281</td>
</tr>
<tr>
<td>corresponding p value</td>
<td></td>
<td></td>
<td>.007</td>
</tr>
</tbody>
</table>

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 $<-1$, $-1<difference<-0.5$, $-0.5<difference<0$, $0<difference<0.5$ difference. $0<difference<0.5$ omitted

$^c$: Dummies for: group mean $<-0.5$, $-0.5<group mean<0$, $0<group mean$.

$^d$: 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
D Simulations

D.1 Formulas

The model of interest is:

\[
\begin{align*}
E_{ik,t+1} &= 1\left[ \beta_{e1}S_{ik} + \beta_{e2}LP_{ik} + \gamma R_{ik} + X_{ik}\beta_{e4} + u_{ik} > 0 \right] \\
R_{ik} &= 1\left[ \beta_{r1}S_{ik} - \beta_{r2}LP_{ik} + f_r(S_{ik} - LP_{ik}) + X_{ik}\beta_{r4} + \epsilon_{ik} < 0 \right] \\
selection &= 1\left[ \beta_{s1}S_{ik} + \beta_{s2}LP_{ik} + \beta_{s3}Z_s + \gamma R_{ik} + X_{ik}\beta_{s4} + v_{ik} > 0 \right]
\end{align*}
\]

(D.1)

Grade repetition risk

For each observation, it is possible to compute the grade repetition risk:

\[
P_{\text{red}} = \Phi(\beta_{r1}S_{ik} - \beta_{r2}LP_{ik} + f_r(S_{ik} - LP_{ik}) + X_{ik}\beta_{r4}).
\]

This risk can easily be adapted to speculative situations with different \(LP_{ik}\).

\[
\tilde{P}_{\text{red}} = \Phi(\beta_{r1}S_{ik} - \beta_{r2}\tilde{LP}_{ik} + f_r(S_{ik} - \tilde{LP}_{ik}) + X_{ik}\beta_{r4}).
\]

gives individual grade repetition risks. The simulations presented here give their average on different groups.

Dropout risk, no “general equilibrium effect”

To simplify the algebra, \(E_{ik,t+1}\) and \(R_{ik}\) are assumed independent. Hence the probability of \(E_{ik,t+1}\) writes \(P_{\text{enr}} = P_{\text{red}}\Phi(\beta_{e1}S_{ik} + \beta_{e2}LP_{ik} + \gamma + X_{ik}\beta_{e4}) + (1 - P_{\text{red}})\Phi(S_{ik} + \beta_{e2}LP_{ik} + X_{ik}\beta_{e4}).\)

The simulations compute the consequences of a speculative change in \(LP_{ik}\) on \(P_{\text{red}}\). They change dropout risk through the change in \(P_{\text{red}}\), but not through \(\beta_{e2}LP_{ik}\). The new probability of \(E_{ik,t+1}\) writes \(\tilde{P}_{\text{enr}} = \tilde{P}_{\text{red}}\Phi(\beta_{e1}S_{ik} + \beta_{e2}\tilde{LP}_{ik} + \gamma + X_{ik}\beta_{e4}) + (1 - \tilde{P}_{\text{red}})\Phi(S_{ik} + \beta_{e2}\tilde{LP}_{ik} + X_{ik}\beta_{e4}),\) where \(\tilde{P}_{\text{red}}\) is the new \(P_{\text{red}}\).

Dropout risk, with “general equilibrium effect”

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

Precision of the estimates

When presented in the paper, they derive from a Delta-method not presented here.
## D.2 Additional Tables

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

<table>
<thead>
<tr>
<th>Grade repetition</th>
<th>LP(i_k) decreases by 0.25 pt</th>
<th>LP(i_k) decreases by 0.5 pt</th>
<th>LP(i_k) decreases by 0.75 pt</th>
<th>LP(i_k) decreases by 1 pt</th>
</tr>
</thead>
<tbody>
<tr>
<td>Full sample</td>
<td>26.4%</td>
<td>19.3%</td>
<td>16.7%</td>
<td>14.7%</td>
</tr>
<tr>
<td>“low-achievers”</td>
<td>74.1%</td>
<td>74.7%</td>
<td>71.2%</td>
<td>65.5%</td>
</tr>
<tr>
<td>“leaning”</td>
<td>51.8%</td>
<td>32.5%</td>
<td>27.4%</td>
<td>23.8%</td>
</tr>
<tr>
<td>“good pupils”</td>
<td>12.7%</td>
<td>10.3%</td>
<td>8.7%</td>
<td>7.7%</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Drop-out risk</th>
<th>LP(i_k) decreases by 0.25 pt</th>
<th>LP(i_k) decreases by 0.5 pt</th>
<th>LP(i_k) decreases by 0.75 pt</th>
<th>LP(i_k) decreases by 1 pt</th>
</tr>
</thead>
<tbody>
<tr>
<td>No “general equilibrium effect”</td>
<td>3.2%</td>
<td>2.4%</td>
<td>2.1%</td>
<td>1.9%</td>
</tr>
<tr>
<td>Full sample</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>“low-achievers”</td>
<td>7.1%</td>
<td>7.1%</td>
<td>6.7%</td>
<td>6.1%</td>
</tr>
<tr>
<td>“leaning”</td>
<td>5.7%</td>
<td>3.6%</td>
<td>3.0%</td>
<td>2.6%</td>
</tr>
<tr>
<td>“good pupils”</td>
<td>1.9%</td>
<td>1.6%</td>
<td>1.4%</td>
<td>1.2%</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>With “general equilibrium effect”</th>
<th>LP(i_k) decreases by 0.25 pt</th>
<th>LP(i_k) decreases by 0.5 pt</th>
<th>LP(i_k) decreases by 0.75 pt</th>
<th>LP(i_k) decreases by 1 pt</th>
</tr>
</thead>
<tbody>
<tr>
<td>Full sample</td>
<td>3.2%</td>
<td>3.1%</td>
<td>3.1%</td>
<td>3.2%</td>
</tr>
<tr>
<td>“low-achievers”</td>
<td>7.1%</td>
<td>9.5%</td>
<td>10.3%</td>
<td>10.7%</td>
</tr>
<tr>
<td>“leaning”</td>
<td>5.7%</td>
<td>4.7%</td>
<td>4.6%</td>
<td>4.5%</td>
</tr>
<tr>
<td>“good pupils”</td>
<td>1.9%</td>
<td>2.1%</td>
<td>2.1%</td>
<td>2.2%</td>
</tr>
</tbody>
</table>

Notes: Simulations based on the estimates of Table 4. 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”: \(0.25 < S_{ik} - LP_{ik}\)