Batista de Oliveira, Luis Felipe; Soares, Sergei S. D.

Working Paper

The impact of the Programa Bolsa Família on grade repetition: Results from the single registry, attendance project and school census

Working Paper, No. 119

Provided in Cooperation with:
International Policy Centre for Inclusive Growth (IPC-IG)

Suggested Citation: Batista de Oliveira, Luis Felipe; Soares, Sergei S. D. (2013) : The impact of the Programa Bolsa Família on grade repetition: Results from the single registry, attendance project and school census, Working Paper, No. 119, International Policy Centre for Inclusive Growth (IPC-IG), Brasilia

This Version is available at:
https://hdl.handle.net/10419/119804

Working Paper

THE IMPACT OF THE PROGRAMA BOLSA FAMÍLIA ON GRADE REPETITION:

RESULTS FROM THE SINGLE REGISTRY, ATTENDANCE PROJECT AND SCHOOL CENSUS

Working Paper number 119 November, 2013

Luis Felipe Batista de Oliveira
Sergei S. D. Soares
Institute for Applied Economic Research (IPEA)

International Policy Centre for Inclusive Growth

International Policy Centre for Inclusive Growth (IPC - IG)
United Nations Development Programme

SBS, Quadra 1, Bloco J, Ed. BNDES, 13º andar
70076-900 Brasilia, DF - Brazil
Telephone: +55 61 21055000

E-mail: ipc@ipc-undp.org ▪ URL: www.ipc-undp.org

The International Policy Centre for Inclusive Growth is jointly supported by the United Nations Development Programme and the Government of Brazil.

Working Papers are available online at www.ipc-undp.org and subscriptions can be requested by email to ipc@ipc-undp.org

Print ISSN: 1812-108X

THE IMPACT OF THE PROGRAMA BOLSA FAMÍLIA ON GRADE REPETITION: RESULTS FROM THE SINGLE REGISTRY, ATTENDANCE PROJECT AND SCHOOL CENSUS

Luis Felipe Batista de Oliveira and Sergei S. D. Soares*

This text attempts to estimate the impacts of the Programa Bolsa Família on grade repetition, using matched data from the Single Registry (Cadastro Único, or CadÚnico), the Attendance Project and annual School Census. The authors use three approaches: i) comparing the results for poor children in the CadÚnico with and without the benefit; ii) estimating the treatment effect; and iii) estimating the impact of the percentage of beneficiary children in a given school on the average repetition level within the school. The results are far from impressive but do show a drop in the probability of grade repetition.

Keywords: Education and inequality; government policy; conditional cash transfers.

1 INTRODUCTION

One of the goals of targeted conditional cash transfer programmes is to increase the human capital of their beneficiaries. Were it not for that, the existence of conditionalities would not make sense. The rationale is clear: the transfer alleviates poverty today, and this condition sets families on the path towards future success. However, for all this to make sense, we need conditionalities to work.

It is not easy to assess the impact of the largest targeted conditional cash transfer programme in Brazil, the Programa Bolsa Família (PBF), on the academic performance of children of beneficiary families. Unlike (a few) other programmes, which were designed from the very beginning to be evaluated, the PBF was created first, and only afterwards was the issue of its impact evaluation considered.

* Researchers at Instituto de Pesquisa Econômica Aplicada (IPEA).

There is no control group, and, in fact, for a long time there was not even a single data source which monitored children from one year to the next, allowing their academic performance to be evaluated.

Although the PBF's design remains non-experimental, today we have information from the new School Census[1] and the Attendance Project that is individualised and longitudinal, which means that one can now monitor children from one school year to the next and observe their school history.

This study aims to estimate the effect of the PBF on academic performance, by examining three individual data sources: i) the Single Registry (CadÚnico), the administrative registry that unifies PBF; ii) the School Census, which collects information on each student; and iii) the Attendance Project, which is a database of records about the school attendance of children who benefit from the PBF.

2 THE DATABASES

To observe the relationship between the PBF and the probability of school failure, information from CadÚnico on beneficiaries and their families in the programme is needed, as is information about the schools they attend and their attendance. Thus, the databases from the School Census between 2008 and 2009 and the Attendance[2] will be important.

Many other studies (Klein and Ribeiro, 1991; Leon and Menezes-Filho, 2002; Duryea, 1998; Ribeiro, 1988) have been conducted to estimate repetition (failure) rates and their determinants, but all of them were conducted either at larger aggregation levels, such as for a state, or using databases such as the Monthly Employment Survey (PME), where there are no variables related to the PBF. From the CadÚnico, School Census and Attendance databases we expect to obtain estimates of PBF's impact by student or school.

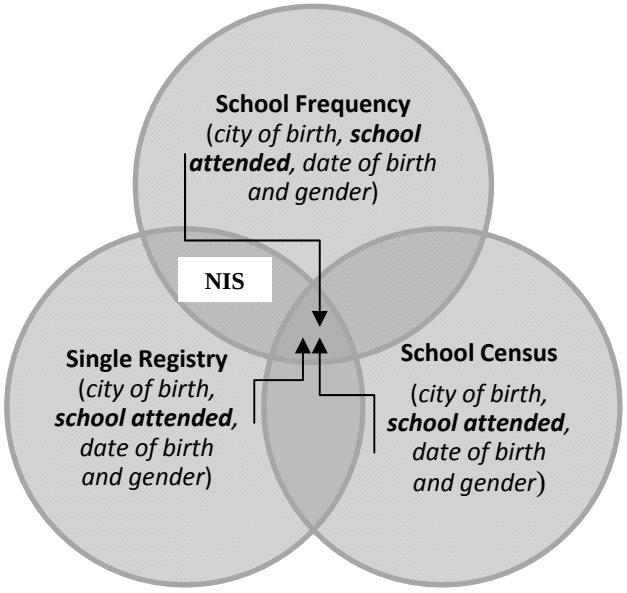
Unfortunately, integrating these databases at the level of the individual student is not a straightforward task. There is no unique identification key that locates, with perfection, the same child in the three databases. The pairing of the Attendance database and the 2008 CadÚnico[3] can be done with relative ease, using information from the recipient's Social Identification Number (NIS). However, adding any of these databases to the School Census[4] is harder to achieve. To that end, we built an 'INEP key' variable, composed of information regarding the city of birth, school code, date of birth and gender. Such information can be easily obtained from the three databases.

It is known that the variable 'school code' is not very well captured by CadÚnico. This is the main reason why we use of the Attendance database for the 2008 PBF, consolidated by the Department of Conditionalities in the Secretaria Nacional de Renda de Cidadania (SENARC — National Secretariat of Citizenship Income) in the Ministério do Desenvolvimento Social e Combate à Fome (MDS — Ministry of Social Development and Fight Against Hunger). That is, it is expected that this variable will be accurate in this database, and will link the database holding the information about the school environment of the student to the information about the student and his/her family, found within the CadÚnico (Figure 1). We chose the two-month period ending in May because, in the words of INEP (2008):

"The School Census is conducted annually. The data collection period is defined by Ordinance, and, in recent years, the beginning of the collection has been *the last Wednesday of May*, named as the National Day of School Census. This reference date was chosen to suit the school calendar from a country as great and diverse as Brazil. To gauge the academic progress of a student within the School Census, there is a second data collection process, also with a period defined by Ordinance, which generally occurs at the beginning of the year following the first-stage collection. In this collection, schools inform if the student has been approved, failed or did not attend school."

FIGURE 1

OVERLAPPING Information among the Databases and the Possible Common Key



It can be argued that the information above cannot generate a key that uniquely identifies the observations. There are, indeed, many duplicated keys. Given the large number of observations, the existence of more than one student per key is not a problem if students that are perfectly identified are not different from students who share a key. Students who share the same key are simply removed from the sample, leaving only those uniquely identified. Provided that the difference between the two types of student is random, this procedure is nothing more than the extraction of a sample.

In the last five years, the School Census has collected information on each individual student, as opposed to school or classroom averages used previously. In other words, it is the highest possible degree of disaggregation for this type of research. However, these important developments are followed by some computational difficulties arising from the size of the databases.[5]

The data are gathered through partnerships with state and municipal education secretariats, which collect data on school infrastructure, teaching staff, enrolment, length of the school day, school progress and movement, by level/stage and the teaching modality.

The Census is carried out in all public and private schools in the country, by means of a standardised questionnaire.

Although the School Census does not contain information about family characteristics, some school environment variables are particularly important. Among them is information about teachers, classes and the school of the student. Table 1 shows the main controls of the proposed models, the source of which is the School Census.[6]

TABLE 1

Key Variables in the School Census

Controls	**Variables**
Student characteristics	Male; age-grade lag *(defasagem)*; uses public transport to school; has some kind of special need; attended year (first year as the base category); studying outside the state where the student was born *(migrante)*; failed the previous grade.
School characteristics	Infrastructure quality by principal component analysis (ACP);[7] school located in an urban area; administrative dependence (private, federal, state or municipal).
Class characteristics	Class length in minutes; length squared; enrolment in class; enrolment squared.
Teacher characteristics (per class)	Average year of graduation; percentage of teachers with pedagogy as their base degree; percentage of teachers that graduated from private institutions; percentage of teachers with post-graduate degrees; percentage of teachers with a Master's degree; percentage of teachers with a doctoral degree; number of teachers per class; percentage of teachers with higher education.

Source: School Census. Prepared by the authors.

2.1 ADJUSTMENTS

In the School Census, the databases with the information about enrolment, the classes, schools and teachers have been worked on by Oliveira and Soares (2012). Thus, only the regular elementary students were kept, and certain imperfections were removed, such as duplicate students; gender and/or date of birth changing from one year to another; students without an identification code; and schools that were not active.

In the Attendance database, students between the first and eighth grades in 2008 were kept. This allowed for harmonisation with the CadÚnico data for households that had updated their data for 1 January 2008. As an administrative record, this database required additional adjustments. Thus, we removed the households that did not respond to at least one of these questions:[8] occupancy status, number of rooms, type of construction, water supply, sanitary drainage and garbage collection.

In the CadÚnico database, people already marked as deleted from the system were removed from the sample. Furthermore, those with an unrealistic birth date,[9] without age information or with an adult missing in the education variable were removed from the sample.

2.2 ON THE INTERSECTION OF DATABASES

In the School Census, the set of variables that compose the INEP key uniquely identifies more than 96 per cent of the observations. This condition is also seen in the Attendance database.

After cross-referencing the CadÚnico database with the Attendance database, we can search in each of the components of the key for the number of coincidences and their relation to the total observations resulting from the cross-referencing.

Table 2 summarises the main results. It can be noted that, with the exception of the 'school code', almost absolute coincidences occur between the variables 'city of birth', 'date of birth' and 'gender' — something between 97 per cent and 99.9 per cent. If the 'school code' variable were modified by the 'family code' variable, the identification would be very efficient — i.e. there would be a key that was closer to the NIS in terms of connection efficiency between CadÚnico and the Attendance database, as this key would be able to identify 96 per cent of students (the 'student key' variable in Table 2).

However, the NIS is sufficient for this task, despite not being able to promote the integration of these two databases with the School Census.

TABLE 2

Quality of Data Check (2008)

Uniquely identified cases within each database (%)	
INEP key in the School Census	96.5
INEP key in the Attendance database	96.2
Coincidences between CadÚnico and the Attendance database variables (%)	
City of birth	99.0
Family code	97.4
Date of birth	99.9
Gender	99.9
School 1st two-month period	58.1
School 2nd two-month period	59.7
School 3rd two-month period	59.1
School 4th two-month period	57.6
School 5th two-month period	57.1
INEP key*	58.6
Student key**	96.1

Sources: School Census, CadÚnico and Attendance databases. Prepared by the authors.

Notes: * INEP key considers: city of birth; school code; date of birth; and gender.

** Student key considers: city of birth; family code; date of birth; and gender.

Thus, a solution to the partial integration of databases, which operates only with public School Census data and which does not use cumbersome integration methods, such as the phonetic study of names of students/relatives, would be done through the school identification variable. It is known that the Attendance database captures this variable well and, therefore, allows a good level of integration with INEP data. However, there is a loss of efficiency in the integration with CadÚnico.

With the 2008 CadÚnico data, it can be noted that the school information for almost 60 per cent of students in the CadÚnico coincides with that observed in the Attendance database in the second two-month period of the same year. It reaches the highest level of identification precisely in this two-month period. Coincidentally or not, at the end of May each year the schools submit their census information to INEP. The efficiency is lower in 2009 because it maintains the same CadÚnico for two years.

Initially, it can be concluded that the proposed key is very efficient in connecting the Attendance and the School Census databases, since it uniquely identifies over 96 per cent of the observations in each database.[10] On the other hand, it boasts lower quality in terms of the integration between Attendance and CadÚnico (around 60 per cent). This loss in quality occurs because of poor capture of the school information in the CadÚnico. However, since this is solved by the Attendance database, there would be a good transition among the three databases in terms of the information for the students included in Attendance monitoring.

3 METHODOLOGY

In principle, an estimate made by regression analysis already offers a reasonable answer, if accompanied by a rich enough set of socio-economic variables, such as those found in CadÚnico for all students. However, as the previous section made clear, the data sources come with several limitations. Moreover, CadÚnico covers only part of the child population. The School Census contains no reasonable socio-economic information about students, and attendance information in the Attendance database covers only people who receive the PBF benefit.

That is, we just do not have enough data to build an equation of the type:

$$P(repetition)_{ik} = \mathrm{f}(\mathrm{X_{ik}}, \mathrm{Z_k}, \mathrm{BF_{ik}}) \quad (1)$$

where $P(repetition)_{ik}$ represents the probability of the child i in school k failing grade s in year t+1; X_{ik} represents the family variables; Z_k the school variables; and BF_{ik} shows whether the family of child i receives PBF benefits. Given these limitations, we will adopt three approaches.

3.1 APPROACH 1: CADÚNICO — LOW-INCOME POPULATION

The first approach will estimate this likelihood through a logistic regression, using only the CadÚnico universe. The disadvantage of this method is that variables are only valid for the poorest half of the children, not for the entire population. Nonetheless, this is not a major problem, since the information that is sought refers specifically to the impact of the PBF on the school progression of the poorest students.

Therefore, only two sources are used: the School Census, from which data about school progression (P(repetition)$_{ik}$) and the school, Z_k, are obtained, and the CadÚnico, which will provide information about the student's socio-economic background (X_{ik}). Unfortunately, data from the Attendance database will be lost, but the information from the School Census is fairly reliable when the adjustments tested in Oliveira and Soares (2012) are considered. These adjustments refer to the high level of discrepancies found in the School Census. Three consecutive years are required to perform these corrections. The greatest disadvantage is that the quality of the data obtained from the CadÚnico and the School Census is not very good. This implies that there will be errors in explanatory variables, with the consequent attenuation factor tending towards zero.

Despite these limitations, this is considered the most reliable approach: it is focused on the individual, it accurately identifies the person who receives the benefits, and it compares recipients to non-recipients.

3.2 APPROACH 2: RECIPIENTS ONLY — INTENSITY OF TREATMENT

It would be a pity not to use the available information about attendance, which is the basis of the Attendance Project. Unfortunately, since this data source covers only individuals who receive PBF benefits, it is simply impossible to measure impacts by comparing recipients to non-recipients. All the same, we can use the Attendance Project and the CadÚnico to predict the dose–response relationship of the PBF. The equation to be calculated is:

$$\mathrm{P(repetition)_{ik}} = \mathrm{f(X_{ik}, Z_k, F_{ik}, Y_{ik})} \quad (2)$$

where Y_{ik} represents the value per capita received by the family through the PBF, and F_{ik} indicates whether the child attended at least 85 per cent of days at school k. Possibly, families that receive higher amounts are more motivated to send their children to school on a regular basis. This forecast has obvious limitations. There is no clear indication that the mechanism used by the PBF is the volume of transfers.

3.3 APPROACH 3: SCHOOLS — WITH RECIPIENTS AND STUDENTS INCLUDED IN THE CADÚNICO

Lastly, instead of observing individual children, the study may focus on the schools. Using the CadÚnico, it is possible to calculate the percentage of poor children in a given school. Likewise, through the Attendance Project, the percentage of children who receive cash transfers from the PBF can be estimated. The following equation could be used:

$$P(repetition)_k = f(Z_k, TC_k, TBF_k) \quad (3)$$

where P_k represents the school repetition rate k; Z_k the school variables in the School Census; TC_k the proportion of children in the CadÚnico; and TBF_k the proportion of children in a school who receive cash transfers from the PBF. Once again, this approach has several shortcomings.

All calculations will use clustered errors for each municipality. The reason for this is that the data entered into the CadÚnico are variable and depend on the competence of the municipality that inputs the information. The model is a simple regression using the Ordinary Least Squares (OLS) method with clustered errors.

3.4 INTERPRETATION

The following coefficients are recorded as an odds ratio — that is, they denote how many times the probability of repetition is greater when a specific trait is present than when it is not present. Thus, if the coefficient of a binary variable (boys, for instance) is equal to 1.30, we may conclude that the given group has a weighted probability[11] of failing that is 30 per cent greater than the group in which this trait is not found (girls).

As a result, for coefficients close to or equal to 1 there is no difference between the groups.[12] On the other hand, coefficients that are smaller than 1 indicate that a person would have a lower probability of repeating than those who do not possess the given trait.[13] If the variable is continuous, the recorded values represent how much the probability of repeating marginally increases when the variable is increased by one unit.

It is important to highlight that only the effects estimated in the third approach are estimated using OLS. The interpretation of these coefficients is the usual: 0, and not 1, means no effect; negative coefficients mean a trait reduces repetition, and positive coefficients mean it increases the likelihood of repetition.

4 RESULTS

The results are presented in Table 3. In the first two approaches, the coefficients and directions are very similar. Thus, students who repeated the previous year[14] have a 46 per cent greater chance[15] of repeating than those who passed. Boys have a 70 per cent greater chance of repeating than girls, and students who are above the appropriate age range for the grade level (achievement gap) are also at a disadvantage.

An especially worrisome problem refers to the high level of repetition among students with special needs. This demonstrates that the school system has met with difficulties in dealing with these individuals. Data show that this group of students has an approximately 80 per cent greater chance of repeating than those who have no type of disability.

In terms of the school grade,[16] the first year is the one with the lowest probability of repetition. This is a positive outcome, because it is critically important to retain children at such an early stage of the learning process. The most noteworthy bottleneck affects students attending the sixth year (former fifth grade). In this second half of primary school, students have longer classes, more teachers and, probably, more responsibilities. Accordingly, teachers seem to use the repetition mechanism to establish, rather tactlessly, the beginning of this new stage.[17] Students in this year have a 70–80 per cent greater chance of repeating than students at the literacy level — a situation that also occurs among third-year students.

Apparently, students who migrate (who do not study in the state where they were born) have approximately 7 per small less chance of repeating than those who belong to families that have not moved. If, in terms of net values, migration is greater towards the states

where repetition mechanisms are less frequently used, such as the South-eastern region (which traditionally presents lower rates), this would explain these results.[18]

With regards to the school characteristics, we may conclude that the higher the quality of the infrastructure, the lower the probability of student failure. Urban schools have nearly 9 per cent more chance of registering repetitions than rural schools. Furthermore, municipal and state schools have 68–84 per cent more chance of having repeating students than private schools.

It is important to highlight that the longer the duration of the classes, the lower the chances of repetition. As expected, in terms of net values, the students who receive more information throughout their school day are more prepared for their exams. If class sizes are too large, there is also a greater chance of student failure. With regards to the cost–benefit ratio, it seems that extending the duration of classes is more effective than reducing the class size, as explained by Oliveira (2008).

In general, it is positive for teachers to have a certain level of specialisation or a Master's or Doctor's degree. However, we cannot say that better-qualified teachers are 'more demanding' and fail more students. On the contrary, these professionals would be better prepared to deal with their pupils and, therefore, would understand that the students are not the only ones to blame for school failure, and that they are actually both part of the problem and the solution.

Family characteristics are also relevant to ensure formal access to the labour market.[19] The larger the number of people in a household, the greater the probability of repetition. The poverty level of the family can also be assessed based on the number of rooms in the household: the larger the number of rooms, the lower the poverty level, and, concurrently, the lower the chances of repetition. Similarly, this is also true for households that have better infrastructure, such as water and sewage services and masonry structures. Nonetheless, the most important family variable is the parents' level of education. If a child's legal representative has completed at least primary school, this reduces the weighted probability of repetition by a little over 30 per cent.

We will now address the main objective of this paper, which is the effect of the PBF on repetition rates. In the first approach, which focuses only on the CadÚnico universe, the likelihood of repetition among students who are PBF recipients is 11 per cent lower than for other students. This means that the programme has a significant impact on student progression, although the results are relatively modest. By simulating the estimated probability of repetition, the result is 14.6 per cent for non-recipients who are included in the CadÚnico, and 13.2 per cent for recipients.

Observing the dose–response relationship exclusively for programme recipients, we may conclude that the second approach does not indicate that an increase in cash transfers would generate higher or lower repetition rates, given that the probability ratio is equal to 1.00.

The school attendance conditionality dummy demonstrates that students who fulfil the programme's attendance conditionality have 40 per cent less chance of repeating than those who do not. Nevertheless, it would be rash to say that this is a direct impact of the requirement, given that 98 per cent of the students in this model[20] abide by the rules, and, in the absence of other instrumental variables capable of explaining this issue, we cannot overlook the fact that the attendance rates and cash transfers are endogenously related.

TABLE 3

Results of the Three Approaches

Variables	Odds ratio		OLS
	Approach 1	Approach 2	Approach 3
Repeated the previous year	1.458***	1.457***	0.0662***
	(0.0320)	(0.0319)	(0.00141)
Male	1.707***	1.720***	0.0505***
	(0.0155)	(0.0153)	(0.00429)
Achievement gap	1.236***	1.230***	0.0102***
	(0.00790)	(0.00793)	(0.000295)
Student transport	1.006	1.012	–0.0119***
	(0.0197)	(0.0205)	(0.00118)
Special needs	1.757***	1.791***	–0.00788
	(0.0560)	(0.0611)	(0.0114)
Second year	1.318***	1.348***	0.0736***
	(0.0782)	(0.0807)	(0.00294)
Third year	1.742***	1.803***	0.0915***
	(0.106)	(0.114)	(0.00296)
Fourth year	1.044	1.074	0.0558***
	(0.0692)	(0.0721)	(0.00428)
Fifth year	1.027	1.065	0.0192***
	(0.0621)	(0.0650)	(0.00404)
Sixth year	1.743***	1.778***	0.0870***
	(0.0878)	(0.0919)	(0.00491)
Seventh year	1.234***	1.256***	0.0170**
	(0.0604)	(0.0631)	(0.00697)
Eighth year	1.013	1.025	–0.0467***
	(0.0501)	(0.0524)	(0.00711)
Ninth year	1.009	1.020	0.0542***
	(0.0479)	(0.0485)	(0.00566)
Migrant	0.930***	0.932***	–0.0200***
	(0.0258)	(0.0251)	(0.00114)
Infrastructure	0.988*	0.988*	–0.00491***
	(0.00623)	(0.00653)	(0.000218)
Urban school	1.093***	1.087***	–0.00193*
	(0.0256)	(0.0262)	(0.00105)
Federal dependency	1.920*	1.084	0.0171**
	(0.644)	(0.319)	(0.00794)
State dependency	1.844***	1.792***	0.0569***
	(0.131)	(0.129)	(0.00119)
Municipal dependency	1.747***	1.682***	0.0424***
	(0.128)	(0.126)	(0.00121)
Class duration	0.996**	0.996**	0.000224***
	(0.00148)	(0.00156)	(4.69e-05) →

Duration^2	1.000**	1.000**	–3.49e-07***
	(1.87e-06)	(1.97e-06)	(6.10e-08)
Students in class	1.026***	1.026***	0.000876***
	(0.00321)	(0.00315)	(6.23e-05)
Students in class*	1.000***	1.000***	–2.62e-06***
	(5.77e-05)	(5.51e-05)	(5.28e-07)
Teachers per class	1.045***	1.046***	0.00573***
	(0.00681)	(0.00711)	(0.000166)
% of teachers with academic specialisation	0.956*	0.960	–0.0113***
	(0.0257)	(0.0259)	(0.00102)
% of teachers with Master's degrees	0.844**	0.837*	–0.00993*
	(0.0683)	(0.0758)	(0.00580)
% of teachers with Doctor's degrees	0.723	0.783	0.0439**
	(0.185)	(0.229)	(0.0210)
% of teachers with Bachelor's degrees	1.162***	1.157***	–0.00945***
	(0.0396)	(0.0408)	(0.00132)
Year of graduation — teacher	1.005	1.004	–0.000282***
	(0.00380)	(0.00390)	(7.50e-05)
% of teachers with licentiate degrees	0.915*	0.892**	–0.0128***
	(0.0434)	(0.0433)	(0.00229)
% of qualified teachers — private schools	1.034	1.047	0.00101
	(0.0336)	(0.0362)	(0.000911)
Legal representative a formal worker	1.054***	1.096***	
	(0.0109)	(0.0133)	
People in household	1.061***	1.015***	
	(0.00258)	(0.00456)	
Number of rooms	0.946***	0.951***	
	(0.00297)	(0.00304)	
Masonry	0.935***	0.943***	
	(0.0144)	(0.0151)	
Water	0.977**	0.982*	
	(0.00957)	(0.0102)	
Sewage	0.988**	0.988*	
	(0.00610)	(0.00636)	
Waste	1.023	1.025	
	(0.0176)	(0.0190)	
Legal representative with primary school level of education	0.678***	0.690***	
	(0.00747)	(0.00816)	
Fulfilled attendance requirements		0.594***	
		(0.0183)	
Total PBF benefits received by family		1.000***	
		(2.18e-05)	
Receives PBF benefits	0.891***		
	(0.0136)		→

Percentage of recipients in school			0.126***
			(0.0150)
Percentage registered in school			–0.0688***
			(0.0142)
Constant	2.47e-06*	8.37e-06	0.494***
	(1.84e-05)	(6.40e-05)	(0.151)
Observations	1.219.916	1.053.169	66.645
Pseudo-R^2 and R^2 (Approach 3)	0.0611	0.0618	0.315

Sources: School Census, CadÚnico and Attendance Project. Created by authors.

Notes: The control variables used for the states were removed from this presentation to make space for other data. Significant standard errors are shown in brackets.

*** p < 0.01, ** p < 0.05, * p < 0.1.

Based on the identified variables, we may conclude that the variable 'percentage registered in school' could indicate a lower level of repetition in schools that have a higher percentage of students who are included in the CadÚnico. On the other hand, the 'percentage of recipients' has a more relevant effect on repetition rates than all other variables used in Approach 3. Thus, from the student's perspective, it is possible to infer that the PBF reduces the chances of repetition. On the other hand, from the viewpoint of the schools, the percentage of recipients may depict the socio-economic status of the schools, which are generally not well prepared and are capable of punishing students by means of the repetition mechanism, blaming them for poor learning outcomes that are actually not their fault.

5 FINAL CONSIDERATIONS

Among the main factors that are associated with the school repetition of low-income students, there are a few that are worthy of note. The high retention rates are recorded among pupils with special needs, boys, students who belong to families with a lower level of education, and those who study in schools that are ill equipped and that are administratively dependent on the municipalities and on the states. It is also clear that the poorer the household, the worse the results obtained by their children. Despite this, the PBF has helped these families to guarantee better conditions for their children.

Considering that the administrative records are subject to operational errors and quality issues, the data must be considered in a broader sense and interpreted more in terms of the trends they reveal, rather than the actual values they convey. The continuous improvement of the CadÚnico (such as version 7) and of the School Census is expected to produce more dependable information. This would enable the development of future studies, with more reliable estimates.

Nevertheless, the main conclusion is that there is evidence that suggests that the PBF reduces school repetition among its recipients.

REFERENCES

Duryea, S. (1998). 'Children's advancement through school in Brazil: the role of transitory shocks to household income'. SSRN eLibrary.

INEP (2008). 'Microdados do censo escolar – manual do usuário', *Coordenação-Geral de Informações e Indicadores Educacionais* [S.l.]. Brasília, Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira.

Klein, R. and S.C.O. Ribeiro (1991). 'O censo educacional e o modelo de fluxo: o problema da repetência', *Revista brasileira de estatística*, Vol. 1, No. 1: 5–45.

Leon, F. L. L. and N.A. Menezes-Filho (2002). 'Reprovação, avanço e evasão escolar no Brasil', *Pesquisa e planejamento econômico*, Vol. 32, No. 3.

Menezes-Filho, N. et al. (2009). 'Avaliando o impacto da progressão continuada nas taxas de rendimento e desempenho escolar do Brasil', Mimeo: 26.

Oliveira, J.M. (2008). 'Custo-efetividade de políticas de redução do tamanho da classe e ampliação da jornada escolar: uma aplicação de estimadores de matching'. São Paulo, FEA/USP, 31º Prêmio BNDES de Economia (2º lugar),.

Oliveira, L.F.B. and S.S.D. Soares (2012). 'Determinantes da repetência escolar no Brasil: uma análise de painel dos censos escolares entre 2007 e 2010', *Texto para Discussão*, No. 1706. Brasília, Instituto de Pesquisa Econômica Aplicada, <http://www.ipea.gov.br/portal/images/stories/PDFs/TDs/td_1706.pdf> (accessed 28 October 2013).

Ribeiro, S.C. (1988). *PNADs em foco: anos 80*. Belo Horizonte, Associação Brasileira de Estudos Populacionais.

Ribeiro, S.C. (1991). 'A pedagogia da repetência', *Estudos avançados*, Vol. 5, No. 12, August.

Soares, S. and N.G.D. Sátyro (2010). 'Infraestrutura das escolas brasileiras e desempenho escolar' in M. Costa and M.P. Morais (Org.), *Infraestrutura social e urbana no Brasil: subsídios para uma agenda de pesquisa e formulação de políticas públicas* (*Eixos Estratégicos do Desenvolvimento Brasileiro*, Vol. 2: 151–192). Brasília, Instituto de Pesquisa Econômica Aplicada.

NOTES

1. Using 2008 microdata from the Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira (INEP).

2. Consolidated by the Department of Conditionalities of SENARC in the MDS.

3. Families were selected who had their data updated as of 1 January 2008. It is intended for this study to be updated to version 7 of CadÚnico. To do so, a large number of people would be required, with updated information — something that had not yet occurred at the beginning of this work.

4. The 'student code' from the School Census ('fk_cod_aluno') is not found in CadÚnico or in the Attendance database.

5. In the five years of the School Census, the databases in Stata format (.dta) amount to almost 40 gigabytes of information.

6. Besides these variables, CadÚnico and the Attendance database will also be used to control family background and student attendance.

7. Soares and Sátyro (2010) summarise the information on school infrastructure in an index generated by the ACP. According to the authors, this index is nothing more than the first factor of an ACP, which includes the existence of certain characteristics of the schools. The following were selected: the existence of filtered water; public water supply; public electric power; public sewer; periodic collection of garbage; boardroom; staff room; computer lab; science lab; multifunction resources for specialised educational services (AEE); sports court; kitchen; library; playground; toilet inside the building; toilets suitable for students with special needs; facilities and suitable routes for students with special needs; television; VCR; DVD player; satellite dish; photocopiers; overhead projector; printer; computer; internet access; and school meals for students.

8. Moreover, households that were not considered active or that were deleted by the system were, obviously, removed.

9. All those born before 1910 were considered 'missing'.

10. In the worst-case scenario, it would identify 92 per cent of the information intersection between the Attendance and the School Census.

11. Weighted because if the probability of repetition for people who have a specific trait is P_i, the likelihood of non-repetition would be $1—P_i$. Consequently, a weighted probability would be illustrated as $P_i/(1 — P_i)$.

12. Some recorded coefficients are represented as '1.000***'. This means that their effect is statistically significant, but marginally positive from the fourth decimal place on. The coefficients are only recorded up to the third decimal place to make them legible.

13. Example: 0.85 would indicate that people who possess a certain trait have 15 per cent less chance of repeating (1 – 0.85) than those who do not.

14. Meaning those who had repeated in 2007.

15. Always in weighted terms.

16. Before the approval of Bill of Law 3675/2004, primary school students attended grades 1 to 8. With the new legislation, literacy classes were included as part of the curriculum for the first year of schooling. In terms of basic education, the former first 'grade' became the second 'year' of primary school, which now progresses until the ninth year.

17. See Ribeiro (1991) for more information about the pedagogy of school repetition and its effects.

18. Although this information is absent from the table, control variables were added for the states (UFs). Nevertheless, no state has registered statistically significant lower chances of repetition than São Paulo. This can be explained by the continuous progression policies described in Menezes-Filho et al. (2009).

19. Individuals who are formally employed, retirees and pensioners were classified as 'formal workers'.

20. That is, they are not classified as 'missing' due to a lack of information in any of the listed variables.

International Policy Centre for Inclusive Growth (IPC - IG)
United Nations Development Programme

SBS, Quadra 1, Bloco J, Ed. BNDES, 13º andar
70076-900 Brasilia, DF - Brazil
Telephone: +55 61 2105 5000

E-mail: ipc@ipc-undp.org ▪ URL: www.ipc-undp.org