

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

**Working Paper**

## O impacto do programa bolsa família sobre a repetência: Resultados a partir do cadastro único, projeto frequência e censo escolar

Texto para Discussão, No. 1814

**Provided in Cooperation with:**

Institute of Applied Economic Research (ipea), Brasília

*Suggested Citation:* Batista de Oliveira, Luis Felipe; Soares, Sergei S. D. (2013) : O impacto do programa bolsa família sobre a repetência: Resultados a partir do cadastro único, projeto frequência e censo escolar, Texto para Discussão, No. 1814, Instituto de Pesquisa Econômica Aplicada (IPEA), Brasília

This Version is available at:

<https://hdl.handle.net/10419/91435>

**Standard-Nutzungsbedingungen:**

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

**Terms of use:**

*Documents in EconStor may be saved and copied for your personal and scholarly purposes.*

*You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.*

*If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.*

# 1814

TEXTO PARA DISCUSSÃO

**O IMPACTO DO PROGRAMA BOLSA FAMÍLIA SOBRE A REPETÊNCIA: RESULTADOS A PARTIR DO CADASTRO ÚNICO, PROJETO FREQUÊNCIA E CENSO ESCOLAR**

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

### **O IMPACTO DO PROGRAMA BOLSA FAMÍLIA SOBRE A REPETÊNCIA: RESULTADOS A PARTIR DO CADASTRO ÚNICO, PROJETO FREQUÊNCIA E CENSO ESCOLAR**

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

---

\* Técnicos de Planejamento e Pesquisa da Diretoria de Estudos e Políticas Sociais (Disoc) do Ipea.

## Governo Federal

**Secretaria de Assuntos Estratégicos da  
Presidência da República**  
Ministro Wellington Moreira Franco

**ipea** Instituto de Pesquisa  
Econômica Aplicada

Fundação pública vinculada à Secretaria de Assuntos Estratégicos da Presidência da República, o Ipea fornece suporte técnico e institucional às ações governamentais – possibilitando a formulação de inúmeras políticas públicas e programas de desenvolvimento brasileiro – e disponibiliza, para a sociedade, pesquisas e estudos realizados por seus técnicos.

### Presidente

Marcelo Côrtes Neri

**Diretor de Desenvolvimento Institucional**  
Luiz Cezar Loureiro de Azeredo

**Diretor de Estudos e Relações Econômicas  
e Políticas Internacionais**  
Renato Coelho Baumann das Neves

**Diretor de Estudos e Políticas do Estado,  
das Instituições e da Democracia**  
Alexandre de Ávila Gomide

**Diretor de Estudos e Políticas  
Macroeconômicas, Substituto**  
Cláudio Hamilton Matos dos Santos

**Diretor de Estudos e Políticas Regionais,  
Urbanas e Ambientais**  
Rogério Boueri Miranda

**Diretora de Estudos e Políticas Setoriais  
de Inovação, Regulação e Infraestrutura**  
Fernanda De Negri

**Diretor de Estudos e Políticas Sociais**  
Rafael Guerreiro Osorio

**Chefe de Gabinete**  
Sergei Suarez Dillon Soares

**Assessor-chefe de Imprensa e Comunicação**  
João Cláudio Garcia Rodrigues Lima

Ouvidoria: <http://www.ipea.gov.br/ouvidoria>  
URL: <http://www.ipea.gov.br>

## Texto para Discussão

Publicação cujo objetivo é divulgar resultados de estudos direta ou indiretamente desenvolvidos pelo Ipea, os quais, por sua relevância, levam informações para profissionais especializados e estabelecem um espaço para sugestões.

© Instituto de Pesquisa Econômica Aplicada – **ipea** 2013

Texto para discussão / Instituto de Pesquisa Econômica  
Aplicada.- Brasília : Rio de Janeiro : Ipea , 1990-

ISSN 1415-4765

1. Brasil. 2. Aspectos Econômicos. 3. Aspectos Sociais.  
I. Instituto de Pesquisa Econômica Aplicada.

CDD 330.908

As opiniões emitidas nesta publicação são de exclusiva e inteira responsabilidade do(s) autor(es), não exprimindo, necessariamente, o ponto de vista do Instituto de Pesquisa Econômica Aplicada ou da Secretaria de Assuntos Estratégicos da Presidência da República.

É permitida a reprodução deste texto e dos dados nele contidos, desde que citada a fonte. Reproduções para fins comerciais são proibidas.

JEL: I24, I28, I38

# SUMÁRIO

---

SINOPSE

ABSTRACT

1 INTRODUÇÃO .....7

2 AS BASES DE DADOS .....7

3 METODOLOGIA.....13

4 RESULTADOS.....16

5 CONSIDERAÇÕES FINAIS .....21

REFERÊNCIAS .....22



## SINOPSE

Este trabalho tenciona estimar o impacto do Programa Bolsa Família (PBF) sobre a repetência escolar usando dados do Cadastro Único (CadÚnico), Projeto Frequência e censos da educação básica. Três abordagens foram utilizadas: *i*) comparação de crianças com e sem benefício, inscritas no CadÚnico; *ii*) impacto do valor do benefício estimado mediante análise de regressão; e *iii*) impacto do percentual de crianças beneficiárias estimado usando escolas como unidade de análise. Os resultados são significativos e o PBF parece contribuir para alguma redução na probabilidade de repetência.

**Palavras-chave:** educação; desigualdade; políticas públicas; bolsa-família; transferência de renda.

## ABSTRACT<sup>i</sup>

This text attempts to estimate the impacts of *Bolsa Família* upon grade repetition using matched data from the Single Registry, the Attendance Project and the annual school censuses. Three approaches are used: *i*) comparison of results for poor children in the Single Registry with and without the benefit; *ii*) estimation of treatment effect; and *iii*) estimation of the impact of the percentage of beneficiary children in a given school upon the average repetition in the school. The results are far from impressive but do show a modest drop in the probability of grade repetition.

**Keywords:** education and inequality; government policy; conditional cash transfers.

---

<sup>i</sup>. *The versions in English of the abstracts of this series have not been edited by Ipea's editorial department.*  
As versões em língua inglesa das sinopses (*abstracts*) desta coleção não são objeto de revisão pelo Editorial do Ipea.



## 1 INTRODUÇÃO

Um dos objetivos dos programas condicionados de transferência de renda focalizados é o aumento do capital humano dos seus beneficiários. Se não fosse assim, a existência de condicionalidades não faria o menor sentido. A racionalidade é clara: a transferência ameniza a pobreza hoje e a condição faz com que as famílias andem na trilha da sua superação futura. No entanto, para que tudo isso faça sentido, é preciso que as condicionalidades funcionem.

Saber o impacto do maior programa condicionado de transferência de renda focalizado do Brasil, o Programa Bolsa Família (PBF), sobre o fluxo escolar de seus beneficiários, no entanto, não é fácil. Ao contrário de alguns (poucos) outros programas, cujo desenho foi feito com a avaliação já pensada, o PBF foi primeiro feito e depois pensou-se em como avaliá-lo. Não há grupo de controle e na verdade nem houve, por muito tempo, sequer uma única fonte de dados que seguisse as crianças de um ano para o próximo, permitindo o cálculo do fluxo escolar.

Apesar de o desenho do PBF continuar não experimental, hoje temos informações tanto no novo Censo Escolar<sup>1</sup> quanto no Projeto Frequência Escolar dos Beneficiários (Frequência) que são individualizadas e longitudinais, o que quer dizer que já se pode acompanhar as crianças de um ano ao próximo e saber, observacionalmente, sua trajetória.

Este trabalho visa estimar o efeito do PBF sobre o fluxo escolar mediante o uso de três fontes de dados individualizadas: *i*) o Cadastro Único (CadÚnico), que é o registro administrativo unificador do PBF; *ii*) o Censo Escolar, cujas informações são levantadas por aluno; e *iii*) o Projeto Frequência, que é uma base de registro da frequência escolar das crianças beneficiárias do programa.

## 2 AS BASES DE DADOS

Para uma observação das relações entre o PBF e as probabilidades de repetência, são necessárias tanto as informações dos beneficiários e suas famílias do programa, no

---

1. Serão utilizados os microdados do Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira (INEP) 2008.

CadÚnico, como aquelas sobre as escolas em que estudam e a frequência dos mesmos nesses estabelecimentos. Assim sendo, as bases do Censo Escolar entre 2008 e 2009 e da Frequência<sup>2</sup> serão instrumentos importantes.

Muitos outros estudos (Klein e Ribeiro, 1991; Leon e Menezes-Filho, 2002; Duryea, 1998; Fletcher e Ribeiro, 1988) já foram realizados com o objetivo de estimar as taxas de repetência e seus determinantes, mas todos foram feitos ou em níveis de agregação maiores, como um estado, ou utilizando bases de dados, como a Pesquisa Mensal de Emprego (PME), do Instituto Brasileiro de Geografia e Estatística (IBGE), nas quais não há variáveis relativas ao PBF. Com as bases mencionadas no parágrafo anterior, espera-se obter estimativas, por aluno ou escola, do impacto do PBF.

Infelizmente, a integração por aluno dessas bases não é trivial. Não há uma chave de identificação única de tal maneira que se localize, com perfeição, a mesma criança nas três bases. O pareamento do Projeto Frequência e o CadÚnico de 2008<sup>3</sup> pode ser feito com relativa facilidade usando a informação do Número de Identificação Social (NIS) do beneficiário. Entretanto, a junção de qualquer uma dessas bases ao Censo Escolar<sup>4</sup> é mais difícil. Para tanto, foi construída a variável chave-INEP composta pelas informações de *município de nascimento, código da escola, data de nascimento e sexo*. Tais informações podem ser facilmente obtidas nos três bancos de dados.

Sabe-se que a variável *código da escola* não é muito bem capturada no CadÚnico. Este é o principal motivo pelo qual é indicado o uso da base de Frequência do PBF de 2008, consolidada pelo Departamento de Condicionalidades da SENARC/MDS. Ou seja, espera-se que essa variável seja bem captada nessa base, sendo capaz de se tornar uma ponte entre a base que detém a informação do ambiente escolar do estudante com as informações do aluno e de sua família, encontradas no CadÚnico (figura 1). Escolheu-se o bimestre terminado em maio, pois nas palavras do INEP (2008):

---

2. Consolidada pelo Departamento de Condicionalidades da Secretaria Nacional de Renda de Cidadania (SENARC) do Ministério do Desenvolvimento Social (MDS).

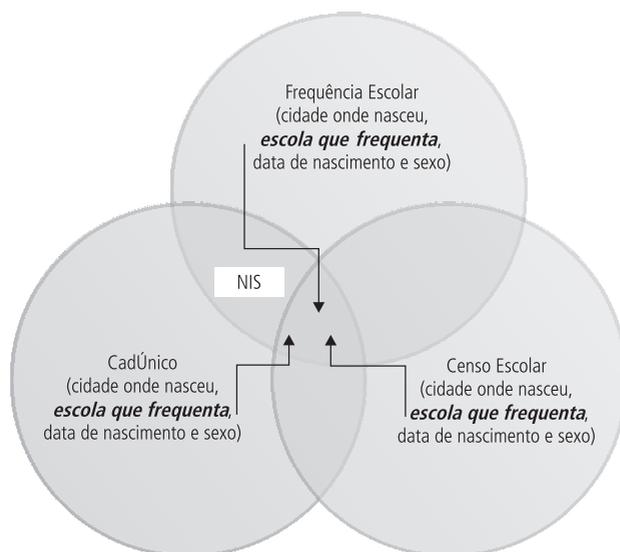
3. Foram selecionadas as famílias que atualizaram seus dados a partir de 1º de janeiro de 2008. Pretende-se, posteriormente, atualizar esse estudo com a versão 7 do CadÚnico. Para tanto, seria necessário um grande número de pessoas com informações atualizadas, algo que ainda não havia ocorrido no início desse trabalho.

4. O código do aluno do Censo Escolar (*fk\_cod\_aluno*) não é encontrado no CadÚnico ou mesmo na base de Frequência.

O Censo Escolar é realizado anualmente. O período de coleta é definido por portaria, e, nos últimos anos, o início da coleta tem sido a última *quarta-feira do mês de maio*, nomeada como o Dia Nacional do Censo Escolar. Essa data de referência foi escolhida para se adequar ao calendário escolar de um país com a grandeza e a diversidade do Brasil. Para identificar a situação de rendimento escolar dos alunos informados pelo Censo Escolar, é feita uma segunda coleta, também com período definido por portaria, que, em geral, ocorre no início do ano seguinte à coleta da primeira etapa. Nessa coleta, as escolas informam se o aluno foi aprovado, reprovado ou deixou de frequentar a escola.

FIGURA 1

**Sobreposição de informações entre as bases e possível chave em comum**



Pode-se argumentar que as informações acima não são capazes de gerar uma chave que identifique unicamente as observações. Há, de fato, muitas chaves duplicadas. Dado o grande número de observações, a existência de mais de um aluno por chave não é um problema desde que alunos perfeitamente identificados não sejam diferentes de alunos que compartilham uma chave. Alunos que compartilham a mesma chave são simplesmente retirados da amostra, deixando apenas os unicamente identificados. Desde que a diferença entre os dois tipos de aluno seja aleatória, este procedimento nada mais é que a extração de uma amostra.

Há cinco anos, o INEP consolida as informações das escolas por meio de um levantamento por aluno<sup>5</sup> no Censo Escolar. Ou seja, trata-se do maior grau de desagregação possível para este tipo de pesquisa. Acompanhadas deste importante avanço, entretanto, estão algumas dificuldades computacionais decorrentes do tamanho das bases.<sup>6</sup>

Isso se dá por meio de parcerias com as Secretarias de Educação estaduais e municipais, coletando-se dados sobre a infraestrutura da escola, o pessoal docente, as matrículas, a jornada escolar, o rendimento e o movimento escolar, por nível/etapa e modalidade de ensino. É aplicado diretamente em todos os estabelecimentos de ensino públicos e privados do país, por meio de questionário padronizado.

Apesar de o Censo Escolar não conter informações sobre características familiares, algumas variáveis do ambiente escolar são particularmente importantes. Entre elas, as informações sobre professores, turma e escola do aluno. O quadro 1 aponta os principais controles dos modelos propostos adiante cuja fonte é o Censo Escolar.<sup>7</sup>

QUADRO 1  
Principais variáveis presentes no Censo Escolar

Controles	Variáveis
Características dos alunos	<i>Homem</i> ; defasagem idade-série ( <i>defasagem</i> ); utiliza transporte escolar público; possui algum tipo de necessidade especial; ano cursado (1 <sup>a</sup> ano como categoria-base); se estuda fora do estado que nasceu ( <i>migrante</i> ); repetiu no ano anterior.
Características da escola	Qualidade da infraestrutura por análise de componentes principais (ACP); <sup>1</sup> escola localizada em área urbana; dependência administrativa (privada, federal, estadual e municipal).
Características das turmas	Duração das aulas em minutos; duração ao quadrado; número de matrículas na turma; número de matrículas ao quadrado.
Características dos professores (em cada turma)	Média do ano de conclusão do curso superior; percentual de professores com licenciatura no curso de formação; percentual de professores formados em instituições privadas; percentual de professores com especialização; percentual de professores com mestrado; percentual de professores com doutorado; número de professores por turma; percentual de professores com curso superior.

Fonte: Censo Escolar. Elaboração dos autores.

Nota: <sup>1</sup> Soares e Sátyro (2010) resumem as informações de infraestrutura da escola em índice gerado pela ACP. Segundo os autores, este índice nada mais é do que o primeiro fator de uma ACP que engloba as existências de determinadas características das escolas. Nesse sentido, foram selecionadas: existência de água filtrada; abastecimento de água por rede pública; energia elétrica por rede pública; esgoto por rede pública; coleta periódica de lixo; sala de diretoria; sala de professores; laboratório de informática; laboratório de ciências; sala de recursos multifuncionais para atendimento educacional especializado (AEE); quadra de esportes; cozinha; biblioteca; parque infantil; sanitário dentro do prédio; sanitário adequado a alunos portadores de necessidades especiais (PNEs); dependências e vias adequadas a PNEs; televisão, videocassete; DVD; parabólica; copiadora; retroprojeto; impressora; computadores; acesso à internet; e alimentação escolar para os alunos.

5. Anteriormente, as informações se davam apenas no nível da escola.

6. Nos cinco anos de Censo Escolar, as bases em formato Stata (.dta) somam quase 40 *gigabytes* de informação.

7. Além destas, variáveis do CadÚnico e da base de Frequência também serão utilizadas no controle do *background* familiar e assiduidade dos alunos.

## 2.1 Adequações

No Censo Escolar, as bases com as informações das matrículas, das turmas, das escolas e dos docentes foram trabalhadas da maneira descrita por Oliveira e Soares (2012). Assim, foram mantidos apenas os alunos do ensino fundamental regular e retiradas algumas imperfeições como alunos duplicados; sexo e/ou data de nascimento trocados de um ano para o outro; alunos sem código de identificação, além das escolas que não estivessem em atividade.

Na base de Frequência, foram mantidos os alunos entre a primeira e a oitava séries do ano de 2008. Isso serviu para harmonizar com os dados do CadÚnico para os domicílios que atualizaram seus dados a partir de 1º de janeiro de 2008. Por ser um registro administrativo, essa base necessitou de maiores adequações. Assim, foram removidos os domicílios que não responderam a, pelo menos, uma das perguntas sobre<sup>8</sup> situação de ocupação, número de cômodos, tipo de construção, abastecimento de água, escoamento sanitário ou coleta de lixo.

Já na base de pessoas do CadÚnico, as pessoas já marcadas como excluídas do sistema foram retiradas da amostra. Além disso, foram retirados aqueles com datas de nascimento irreais,<sup>9</sup> sem informação de idade ou com algum *missing* de adulto na variável de educação.

## 2.2 Sobre a intersecção das bases

No Censo Escolar, o conjunto das variáveis que compõe a Chave-INEP identifica, univocamente, mais de 96% das observações. Tal condição também é verificada na base de Frequência. Algum tipo de identificação duplicada é percebida, porém seus determinantes não foram observados até o presente momento.<sup>10</sup>

Após o cruzamento da base do CadÚnico com a base de Frequência, pôde-se buscar, em cada um dos componentes da chave, o número de coincidências e sua relação com o total de observações oriundas desse cruzamento.

---

8. Além disso, obviamente foram retirados os domicílios que não fossem ativos ou considerados excluídos pelo sistema.

9. Todos aqueles nascidos antes de 1910 foram considerados *missing*.

10. Marginalmente, uma pequena parte das observações duplicadas seria composta por gêmeos. Entretanto, isso ainda não explica o total de indivíduos que não são unicamente identificados.

A tabela 1 sintetiza os principais resultados. Nota-se, com exceção do *código da escola*, que ocorrem coincidências quase absolutas entre as variáveis *município de nascimento*, *data de nascimento* e *sexo* – algo entre 97% e 99,9%. Se a variável *código da escola* fosse trocada pela variável *código familiar*, a identificação seria muito eficiente. Ou seja, poder-se-ia ter uma chave muito próxima do NIS em termos de eficiência de ligação entre o CadÚnico e a base de Frequência, pois tal chave seria capaz de identificar 96% dos alunos (variável *chave-aluno* na tabela 1).

TABELA 1  
**Qualidade do batimento dos dados (2008)**

(Em %)

Casos unicamente identificados dentro de cada base	
Chave-INEP no Censo Escolar	96,5
Chave-INEP na base de Frequência	96,2
Coincidências entre as variáveis do CadÚnico e de Frequência	
Município de nascimento	99,0
Código familiar	97,4
Data de nascimento	99,9
Sexo	99,9
Escola 1º bimestre	58,1
Escola 2º bimestre	59,7
Escola 3º bimestre	59,1
Escola 4º bimestre	57,6
Escola 5º bimestre	57,1
Chave-INEP <sup>1</sup>	58,6
Chave-aluno <sup>2</sup>	96,1

Fontes: Censo Escolar, CadÚnico e Frequência.

Elaboração dos autores.

Notas: <sup>1</sup> Chave-INEP considera: município de nascimento, código da escola, data de nascimento e sexo.

<sup>2</sup> Chave-aluno considera: município de nascimento, código familiar, data de nascimento e sexo.

Contudo, o NIS é suficiente para essa tarefa, mesmo não sendo capaz de promover a integração dessas duas bases com o Censo Escolar. Tampouco seria eficiente, nesse papel, a variável *código familiar* – concatenada em *chave-aluno* – já que a mesma não é capturada no Censo Escolar.

Assim, uma saída para a integração parcial das bases, que opere apenas com os dados públicos do Censo Escolar e que não utilize métodos de integração trabalhosos como o estudo fonético dos nomes dos alunos/familiares, é por meio da variável de identificação da escola. Sabe-se que essa variável é bem capturada na Frequência e,

portanto, promove uma boa integração com os dados do INEP. Entretanto, perde-se eficiência na sua integração com o CadÚnico.

Note-se, pelos dados do CadÚnico de 2008, que quase 60% dos alunos do CadÚnico têm a informação da escola coincidente com a observada na base de Frequência no segundo bimestre do mesmo ano. Exatamente nesse bimestre, atinge-se o máximo de identificação. Por coincidência ou não, ao fim de maio de cada ano as escolas submetem suas informações censitárias ao INEP. Finalmente, a eficiência é menor em 2009, pois se mantém o mesmo CadÚnico em dois anos.

Inicialmente, pode-se concluir que a chave proposta é muito eficiente na junção das bases de Frequência e Censo Escolar, pois identifica unicamente mais de 96% das observações em cada base,<sup>11</sup> possuindo, por outro lado, menor qualidade na integração da Frequência com o CadÚnico (algo em torno de 60%). Tal perda se dá, substancialmente, pela má captura da informação da escola nesse último registro. Entretanto, como isso é contornado pela base de Frequência, pode-se afirmar que haveria uma boa transição entre as informações das três bases daqueles alunos que pertencem ao monitoramento de Frequência.

### 3 METODOLOGIA

Em princípio, uma estimação feita mediante análise de regressão já ofereceria uma resposta razoável, se acompanhada de um conjunto rico de variáveis socioeconômicas, como as encontradas no CadÚnico, para todos os alunos. No entanto, como a seção anterior deixou claro, as fontes de dados vêm com várias limitações. Além disso, o CadÚnico cobre apenas uma parte da população infantil (geralmente pobres). O Censo Escolar não contém nenhuma informação socioeconômica razoável dos alunos e a informação de alta frequência de qualidade do Projeto Presença cobre apenas as pessoas que recebem PBF.

---

11. No pior dos cenários, identificaria 92% da intersecção de informações de Frequência com Censo Escolar.

Ou seja, simplesmente não temos os dados para montar uma equação do tipo:

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

em que  $P(\text{repetência})_{ik}$  representa a probabilidade de a criança  $i$  na escola  $k$  repetir a série  $s$  no ano  $t + 1$ ;  $X_{ik}$  representa as variáveis familiares;  $Z_k$  as variáveis escolares; e  $BF_{ik}$  se a família da criança  $i$  recebe PBF. Diante dessas limitações, serão adotadas três abordagens.

### 3.1 Abordagem 1: CadÚnico – público pobre

A primeira é estimar essa probabilidade por meio de regressão logística, usando apenas o universo do CadÚnico. A desvantagem é que as variáveis em questão valem apenas para a metade mais pobre da população infantil e não para todo mundo. Mas, dado que se deseja saber o impacto do PBF sobre o fluxo escolar justamente dos mais pobres, isto não é um problema grave.

Para tanto, são utilizadas apenas duas bases: o censo, de onde virão os dados sobre fluxo ( $P(\text{repetência})_{ik}$ ) e a escola  $Z_k$ , e o CadÚnico, de onde virão os dados sobre origem socioeconômica ( $X_{ik}$ ). Infelizmente, os dados de alta qualidade e frequência do Projeto Presença serão perdidos, restando os dados do Censo Escolar, com os devidos ajustes testados em Oliveira e Soares (2012). Estes ajustes basicamente dizem respeito ao alto grau de atrito no censo e requerem três anos consecutivos para serem feitos. A maior desvantagem é que nem os dados do CadÚnico nem aqueles do censo são de alta qualidade, o que sugere que haverá considerável erro nas variáveis explicativas, com todo o viés de atenuação em direção ao zero que isso implica.

A despeito das limitações, acredita-se ser esta a abordagem mais confiável: é feita em nível do indivíduo, identifica exatamente quem recebe o benefício e compara beneficiários com não beneficiários.

### 3.2 Abordagem 2: apenas beneficiários – intensidade do tratamento

Seria uma pena não usar a base de alta frequência e alta qualidade que se tem disponível, que é a base do Projeto Presença. Infelizmente, como esta base cobre apenas indivíduos que são beneficiários, simplesmente não há como medir impactos comparando

beneficiários e não beneficiários. O que se pode fazer é usar o Projeto Presença e o CadÚnico para estimar o efeito dosagem do PBF. A equação a ser estimada é:

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

Em que  $Y_{ik}$  representa o valor *per capita* recebido do Bolsa Família e  $F_{ik}$  a sua frequência à escola  $k$ . Possivelmente, famílias que recebem valores maiores têm mais incentivo de enviar suas crianças à escola com assiduidade. Trata-se de uma estimação com limitações óbvias. Não é nada claro que o canal pelo qual o PBF atue seja o volume da transferência.

### 3.3 Abordagem 3 – escolas: com beneficiários e pertencentes ao cadastro

Finalmente, é possível abrir mão de observar crianças individuais e escolas. Usando o CadÚnico, pode-se estimar a porcentagem das crianças em uma dada escola que é pobre. Usando o Projeto Presença, pode-se saber a porcentagem que recebe PBF. Seria possível, assim, estimar a seguinte equação:

$$P(\text{repetência})_k = f(Z_k, TC_k, TBF_k) \quad (3)$$

em que  $P_k$  representa a taxa de repetência na escola  $k$ ;  $Z_k$  as variáveis escolares no Censo Escolar;  $TC_k$  a proporção de crianças no CadÚnico; e  $TBF_k$  a proporção de crianças em escolas que recebem PBF. Mais uma vez, é uma abordagem que sofre de limitações. A principal é que  $TC_k$  é uma variável muito limitada de *status* socioeconômico da escola.

Em todas as estimações serão usados erros agrupados<sup>12</sup> por município, pois o preenchimento do CadÚnico é variável e depende da competência do município que o está cadastrando. Assim, o modelo torna-se uma regressão por Mínimos Quadrados Ordinários (MQO) comum.

### 3.4 Interpretação

Os coeficientes da seção seguinte estão reportados em termos da razão entre chances, ou seja, denotam quantas vezes a probabilidade de repetência é maior, para aquela

12. Tradução livre de *clustered errors*.

característica observada, do que quando não se observa essa característica. Assim se o coeficiente de uma variável dicotômica (por exemplo, meninos) possuir um valor igual a 1,30, diz-se que aquele grupo possui uma probabilidade ponderada<sup>13</sup> de reprovar que é 30% maior do que a do grupo sem a característica (meninas).

Logo, para coeficientes iguais a 1 ou próximos, não há distinção entre os grupos.<sup>14</sup> De forma oposta, coeficientes menores do que 1 indicam menores chances de repetência do que no caso das pessoas que não possuam a característica em questão.<sup>15</sup> Caso a variável seja contínua, deve-se observar que os valores reportados representam o quanto aumentam as chances de repetência com o incremento de apenas uma unidade da variável em questão.

Ressalta-se, novamente, que apenas os efeitos estimados na terceira abordagem são em termos de um modelo MQO. Logo, devem ser interpretados da maneira usual, com os dados agrupados por escola.

## 4 RESULTADOS

Os resultados encontram-se na tabela 2. Nas duas primeiras abordagens, notam-se coeficientes e direções muito similares. Assim, alunos que repetiram no ano anterior<sup>16</sup> possuem 46% mais chances<sup>17</sup> de repetir do que aqueles que foram aprovados. Os meninos possuem 70% mais chances de repetir do que as meninas e os estudantes com idade além da que seria correta (defasagem) para o ano em que estão também são bastante prejudicados.

---

13. Ponderada, pois se a probabilidade de repetência para as pessoas que possuem determinada característica for  $P_r$ , a probabilidade de não repetência seria  $1 - P_r$ . De tal maneira que uma chance ponderada é vista como  $P_r/(1 - P_r)$ .

14. Alguns coeficientes reportados possuem a seguinte representação "1,000\*\*\*". Ou seja, seu efeito é estatisticamente significativo, porém marginalmente positivo em alguma casa decimal da quarta em diante. Isso porque, para fins de facilidade de leitura, os valores dos coeficientes são reportados apenas até a terceira casa decimal.

15. Exemplo: um valor de 0,85 indicaria que as pessoas que possuem determinada característica têm uma chance de repetência 15% menor ( $0,85 - 1$ ) do que as que não a possuem.

16. Ou seja, que já haviam sofrido repetência em 2007.

17. Sempre em termos ponderados.

**TABELA 2**  
**Razão entre chances nas três abordagens**

Variáveis	Razão entre chances		MQO
	Abordagem 1	Abordagem 2	Abordagem 3
Repetiu no ano anterior	1.458*** (0.0320)	1.457*** (0.0319)	0.0662*** (0.00141)
Homem	1.707*** (0.0155)	1.720*** (0.0153)	0.0505*** (0.00429)
Defasagem	1.236*** (0.00790)	1.230*** (0.00793)	0.0102*** (0.000295)
Transporte escolar	1.006 (0.0197)	1.012 (0.0205)	-0.0119*** (0.00118)
Necessidade especial	1.757*** (0.0560)	1.791*** (0.0611)	-0.00788 (0.0114)
Segundo ano	1.318*** (0.0782)	1.348*** (0.0807)	0.0736*** (0.00294)
Terceiro ano	1.742*** (0.106)	1.803*** (0.114)	0.0915*** (0.00296)
Quarto ano	1.044 (0.0692)	1.074 (0.0721)	0.0558*** (0.00428)
Quinto ano	1.027 (0.0621)	1.065 (0.0650)	0.0192*** (0.00404)
Sexto ano	1.743*** (0.0878)	1.778*** (0.0919)	0.0870*** (0.00491)
Sétimo ano	1.234*** (0.0604)	1.256*** (0.0631)	0.0170** (0.00697)
Oitavo ano	1.013 (0.0501)	1.025 (0.0524)	-0.0467*** (0.00711)
Nono ano	1.009 (0.0479)	1.020 (0.0485)	0.0542*** (0.00566)
Migrante	0.930*** (0.0258)	0.932*** (0.0251)	-0.0200*** (0.00114)
Infraestrutura	0.988* (0.00623)	0.988* (0.00653)	-0.00491*** (0.000218)
Escola urbana	1.093*** (0.0256)	1.087*** (0.0262)	-0.00193* (0.00105)
Dependência federal	1.920* (0.644)	1.084 (0.319)	0.0171** (0.00794)
Dependência estadual	1.844*** (0.131)	1.792*** (0.129)	0.0569*** (0.00119)
Dependência municipal	1.747*** (0.128)	1.682*** (0.126)	0.0424*** (0.00121)
Duração da aula	0.996** (0.00148)	0.996** (0.00156)	0.000224*** (4.69e-05)
Duração <sup>2</sup>	1.000** (1.87e-06)	1.000** (1.97e-06)	-3.49e-07*** (6.10e-08)
Alunos na turma	1.026*** (0.00321)	1.026*** (0.00315)	0.000876*** (6.23e-05)
Alunos na turma <sup>2</sup>	1.000*** (5.77e-05)	1.000*** (5.51e-05)	-2.62e-06*** (5.28e-07)

(Continua)

(Continuação)

Variáveis	Razão entre chances		MQO
	Abordagem 1	Abordagem 2	Abordagem 3
Professores por turma	1.045*** (0.00681)	1.046*** (0.00711)	0.00573*** (0.000166)
% professores com especialização	0.956* (0.0257)	0.960 (0.0259)	-0.0113*** (0.00102)
% professores com mestrado	0.844** (0.0683)	0.837* (0.0758)	-0.00993* (0.00580)
% professores com doutorado	0.723 (0.185)	0.783 (0.229)	0.0439** (0.0210)
% professores com curso superior	1.162*** (0.0396)	1.157*** (0.0408)	-0.00945*** (0.00132)
Ano de conclusão do curso – professor	1.005 (0.00380)	1.004 (0.00390)	-0.000282*** (7.50e-05)
% professores com licenciatura	0.915* (0.0434)	0.892** (0.0433)	-0.0128*** (0.00229)
% professores formados – rede privada	1.034 (0.0336)	1.047 (0.0362)	0.00101 (0.000911)
Formalização do responsável legal	1.054*** (0.0109)	1.096*** (0.0133)	
Pessoas no domicílio	1.061*** (0.00258)	1.015*** (0.00456)	
Número de cômodos	0.946*** (0.00297)	0.951*** (0.00304)	
Tijolo/alvenaria	0.935*** (0.0144)	0.943*** (0.0151)	
Água	0.977** (0.00957)	0.982* (0.0102)	
Esgoto	0.988** (0.00610)	0.988* (0.00636)	
Lixo	1.023 (0.0176)	1.025 (0.0190)	
Responsável legal com ensino fundamental	0.678*** (0.00747)	0.690*** (0.00816)	
Cumpriu a condicionalidade		0.594*** (0.0183)	
Benefício total na família		1.000*** (2.18e-05)	
Recebe benefício	0.891*** (0.0136)		
Percentual de beneficiários na escola			0.126*** (0.0150)
Percentual de cadastrados na escola			-0.0688*** (0.0142)
Constante	2.47e-06* (1.84e-05)	8.37e-06 (6.40e-05)	0.494*** (0.151)
Observações	1,219,916	1,053,169	66,645
Pseudo- $R^2$ e $R^2$ (abordagem 3)	0.0611	0.0618	0.315

Fontes: Censo Escolar, CadÚnico e Frequência.

Elaboração dos autores.

Obs.: 1. Os controles utilizados para as UFs foram suprimidos dessa apresentação para preservar o espaço.

2. Erros-padrão robustos entre parênteses.

3. \*\*\*  $p < 0,01$ , \*\*  $p < 0,05$ , \*  $p < 0,1$ .

Um dos dados mais preocupantes é o alto índice de repetência entre alunos que possuem algum tipo de necessidade especial. Isso demonstra uma dificuldade do sistema escolar em lidar com essas pessoas. Os dados indicam que esse público possui chances, aproximadamente 80%, maiores de falha na progressão do que aqueles que não possuem nenhuma limitação.

Ao controlar pelo ano<sup>18</sup> cursado, nota-se que não há nenhum ano com chances de repetência menores do que as do primeiro ano. Isso é bom, pelo fato de não haver uma decisão tão forte como a de reter a criança em fase tão precoce do aprendizado. O gargalo mais notável ocorre entre os alunos do sexto ano (antiga quinta série). Nessa segunda metade do ensino fundamental, os alunos passam a ter aulas com maior duração, um aumento do número de professores é, provavelmente, mais exigido. Assim, os professores parecem utilizar o instrumento da repetência para delimitar, de maneira pouco sutil, o início dessa nova etapa.<sup>19</sup> Os alunos nessa etapa possuem entre 70% e 80% mais chances de repetência do que os alunos da classe de alfabetização, algo que também é notado entre os alunos do terceiro ano.

Aparentemente, alunos que migraram (não estudam no estado onde nasceram) possuem chances de repetência, aproximadamente, 7% menores do que aqueles cujas famílias não tomaram essa decisão. Se, em termos líquidos, a migração possuir incidência maior em direção aos estados onde o instrumento de repetência é utilizado com menor intensidade, como no Sudeste (tradicionalmente com taxas mais baixas), tem-se a explicação desse valor.<sup>20</sup>

Com respeito às características das escolas, nota-se que quanto melhor a infraestrutura, menores as chances de seus alunos serem reprovados. As escolas urbanas possuem cerca de 9% mais chances de reprovação do que as escolas do campo. Além disso, as escolas municipais e estaduais possuem entre 68% e 84% mais chances de reprovação do que as escolas privadas.

---

18. Antes do Projeto de Lei (PL) nº 3.675/2004, os alunos do ensino fundamental eram distribuídos entre a primeira e a oitava série. Com a inclusão da classe de alfabetização como primeiro ano deste período, a educação básica passou a considerar a antiga primeira série como o segundo ano do ensino fundamental – e assim por diante até o nono ano.

19. Sobre a pedagogia da repetência e os seus efeitos, ver Ribeiro (1991).

20. Apesar de não reportados na tabela, foram adicionados controles para as Unidades da Federação (UFs). No entanto, nenhuma possui chances de repetência estatisticamente significantes menores do que São Paulo. Isso pode ser explicado pelas políticas de progressão continuada como descrito em Menezes-Filho *et al.* (2009).

Nota-se que quanto maior a duração das aulas, menores as chances de repetência. Como esperado, os alunos que, em termos líquidos, recebem maior conteúdo ao longo de seu dia na escola estão mais preparados para suas avaliações. Já turmas muito grandes apontam para maiores chances de reprovação. Em termos de custo-benefício, uma política de ampliação da duração parece ser mais eficiente do que de redução da classe, como visto em Oliveira (2008).

De maneira geral, parece ser excelente que os professores possuam algum grau de especialização/mestrado/doutorado. Assim, não se pode dizer que professores mais preparados são “mais exigentes” e reprovam mais. Pelo contrário, esses profissionais estariam mais preparados para lidar com seu público e, portanto, não culpariam exclusivamente o aluno pelo fracasso na escola, sendo este parte tanto do problema como da solução.

Quanto às características das famílias, percebe-se alguma importância para a inserção formal no mercado de trabalho.<sup>21</sup> Quanto maior o número de pessoas no domicílio, maiores as chances de repetência. Um maior número de cômodos apontaria para um nível de pobreza mais baixo e, de fato, há uma menor chance de repetência. Da mesma maneira, isso também é verificado em domicílios com melhor infraestrutura, tal como acesso a água, esgoto e ter sido construído com tijolo/alvenaria. A variável mais importante na família refere-se à educação dos pais. Nota-se que os responsáveis legais que possuem ao menos o ensino fundamental completo reduzem as chances ponderadas de repetência em um pouco mais do que 30%.

Tratemos agora do objetivo principal deste trabalho: mensurar os efeitos do PBF na repetência. Na primeira abordagem, apenas com alunos do CadÚnico, nota-se que o aluno participante do programa possui uma probabilidade de repetência cerca de 11% menor do que os demais. Ou seja, o programa tem um impacto significativo sobre a progressão das crianças, embora o mesmo seja relativamente modesto. Ao simular a probabilidade de repetência predita, encontra-se 14,6% para não beneficiários pertencentes ao CadÚnico, contra 13,2% para beneficiários.

---

21. Foram classificados como “formais” os responsáveis legais que eram assalariados com carteira ou aposentados ou pensionistas.

Ao se observar o efeito-dosagem do programa, apenas entre beneficiários, a segunda abordagem não indica que aumentos nas quantias transferidas geram maiores ou menores taxas de repetência, com razão entre chances igual a 1,00.

Outra variável considerada, a frequência à escola, indica que alunos que cumprem a condicionalidade de frequência possuem chances até 40% menores de repetência do que aqueles que não cumprem. Entretanto, afirmar que esse é um impacto exato da condicionalidade pode ser precipitado, já que cerca de 98% dos alunos, nesse modelo,<sup>22</sup> cumprem o requisito e não é possível descartar endogeneidade entre frequência e valor recebido, na ausência de alguma variável instrumental capaz de sanar a questão.

Com relação à última abordagem, percebe-se uma persistência da direção dos efeitos na maioria das variáveis, mas há uma diferença considerável em termos da magnitude. Isto se explica pelo fato de as variáveis estarem em termos relativos (percentual na escola, ou média dos alunos na escola) em uma regressão simples por MQO.

Nas variáveis de interesse, percebe-se que a variável *percentual de cadastrados na escola* poderia apontar um menor nível de repetência em escolas com maior percentual de alunos pertencentes ao CadÚnico. Por outro lado, o *percentual de beneficiários* é mais incisivo no efeito sobre a incidência de repetência do que todas as variáveis da abordagem 3. Ou seja, no nível do aluno, percebe-se que o PBF traz uma redução na probabilidade de repetência, enquanto no nível da escola, o percentual de beneficiários pode ser um retrato do *status* socioeconômico da escola: em geral, mal preparada e que é capaz de punir os alunos com o instrumento da repetência, culpando-os por um aprendizado ruim que não vem a ser culpa deles.

## 5 CONSIDERAÇÕES FINAIS

Entre os principais fatores associados à repetência dos alunos pobres, destaca-se a forte incidência da retenção nos alunos PNEs, nos meninos, naqueles que estão em famílias menos escolarizadas, nas escolas menos estruturadas e que dependem administrativamente dos municípios e estados. Percebe-se, ainda, que os domicílios

---

22. Ou seja, que não são classificados como *missing* por falta de informação em qualquer uma das variáveis listadas.

menos favorecidos trazem os piores retornos para suas crianças, mas que o PBF tem ajudado essas famílias a garantir melhores condições para seus filhos.

Por se tratar de registros administrativos sujeitos a falhas de preenchimento e de qualidade, em um sentido amplo, os dados devem ser muito mais interpretados em termos de direção do que de suas magnitudes. Ou seja, espera-se que o contínuo aperfeiçoamento do CadÚnico (como a versão 7) e do Censo Escolar ensejaria estimativas mais confiáveis. Isso traria possibilidades de estudos futuros, com maior confiabilidade nas estimativas.

Fica, no entanto, a mensagem principal: há evidências de que o PBF reduz a repetência de quem o recebe.

## REFERÊNCIAS

DURYEA, S. **Children's advancement through school in Brazil: the role of transitory shocks to household income.** SSRN e Library, 1998.

FLETCHER, P. R.; RIBEIRO, S. C. A educação na estatística nacional. *In*: SAWYER, D. O. (Org.). **PNADs em foco: anos 80.** Nova Friburgo: ABEP, 1988. p. 13-32.

INEP – INSTITUTO NACIONAL DE ESTUDOS E PESQUISAS EDUCACIONAIS ANÍSIO TEIXEIRA. **Microdados do censo escolar** – manual do usuário. Coordenação-Geral de Informações e Indicadores Educacionais. [s.l.], 2008.

KLEIN, R.; RIBEIRO, S. C. O censo educacional e o modelo de fluxo: o problema da repetência. **Revista brasileira de estatística**, v. 1, n. 1, p. 5-45, 1991.

LEON, F. L. L.; MENEZES-FILHO, N. A. Reprovação, avanço e evasão escolar no Brasil. **Pesquisa e planejamento econômico**, v. 32, n. 3, 2002.

MENEZES-FILHO, N. *et al.* **Avaliando o impacto da progressão continuada nas taxas de rendimento e desempenho escolar do Brasil.** 2009. p. 26. Mimeografado.

OLIVEIRA, J. M. **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), 2008.

OLIVEIRA, L. F. B.; SOARES, S. S. D. **Determinantes da repetência escolar no Brasil: uma análise de painel dos censos escolares entre 2007 e 2010.** Brasília: Ipea, 2012 (Texto para Discussão, n. 1.706). Disponível em: <[http://www.ipea.gov.br/sites/000/2/publicacoes/tds/TD\\_1706.pdf](http://www.ipea.gov.br/sites/000/2/publicacoes/tds/TD_1706.pdf)>.

RIBEIRO, S. C. **PNADs em foco**: anos 80. Belo Horizonte: ABEP, 1988.

\_\_\_\_\_. A pedagogia da repetência. **Estudos avançados**, v. 5, n. 12, ago. 1991.

SOARES, S.; SÁTYRO, N. G. D. Infraestrutura das escolas brasileiras e desempenho escolar. *In*: COSTA, M.; MORAIS, M. P. (Org.). **Infraestrutura social e urbana no Brasil**: subsídios para uma agenda de pesquisa e formulação de políticas públicas. Brasília: Ipea, 2010 (Série Eixos Estratégicos do Desenvolvimento Brasileiro, v. 2, p. 151-192).

## **EDITORIAL**

### **Coordenação**

Cláudio Passos de Oliveira

### **Supervisão**

Andrea Bossle de Abreu

### **Revisão**

Cristina Celia Alcantara Possidente  
Edylene Daniel Severiano (estagiária)  
Eliezer Moreira  
Elisabete de Carvalho Soares  
Lucia Duarte Moreira  
Luciana Nogueira Duarte  
Míriam Nunes da Fonseca

### **Editoração eletrônica**

Roberto das Chagas Campos  
Aeromilson Mesquita  
Aline Cristine Torres da Silva Martins  
Carlos Henrique Santos Vianna  
Hayra Cardozo Manhães (estagiária)  
Maria Hosana Carneiro Cunha

### **Capa**

Luís Cláudio Cardoso da Silva

### **Projeto Gráfico**

Renato Rodrigues Bueno

### **Livraria do Ipea**

SBS – Quadra 1 – Bloco J – Ed. BNDES, Térreo.

70076-900 – Brasília – DF

Fone: (61) 3315-5336

Correio eletrônico: [livraria@ipea.gov.br](mailto:livraria@ipea.gov.br)



---

Composto em Adobe Garamond Pro 12/16 (texto)  
Frutiger 67 Bold Condensed (títulos, gráficos e tabelas)  
Impresso em offset 90g/m<sup>2</sup>  
Cartão supremo 250g/m<sup>2</sup> (capa)  
Rio de Janeiro-RJ

---

### **Missão do Ipea**

Produzir, articular e disseminar conhecimento para aperfeiçoar as políticas públicas e contribuir para o planejamento do desenvolvimento brasileiro.

