

Educação, pobreza e desigualdade social: discussões e proposições metodológicas para pesquisa

Gabriela Schneider

Universidade Federal do Paraná, Brasil

 <https://orcid.org/0000-0002-6346-2849>

gabrielaschneider@ufpr.br

Domingos Abreu

Universidade Federal do Ceará, Brasil

 <https://orcid.org/0000-0002-1252-4579>

domingos_edu@yahoo.com.br

Ana Lorena de Oliveira Briel

Universidade Federal do Paraná, Brasil

 <https://orcid.org/0000-0001-7978-5805>

analorena@ufpr.br

Alexandre Jeronimo Correia Lima

Universidade Federal do Ceará, Brasil

 <https://orcid.org/0000-0001-6437-8318>

alexandrejeronimo@ufc.br

Polyana Lunelli

Universidade Federal do Paraná, Brasil

 <https://orcid.org/0000-0003-1227-5207>

plunelli85@gmail.com

Irapuan Peixoto Lima Filho

Universidade Federal do Ceará, Brasil

 <https://orcid.org/0000-0002-6669-2471>

irapuan.peixoto@ufc.br

Suellem Raquel de Freitas Bonfim

Secretaria Municipal de Educação de Curitiba

 <https://orcid.org/0000-0002-7422-9022>

sususel@yahoo.com.br

Introdução

No Brasil é notório famílias em situação de pobreza e de extrema pobreza enfrentarem dificuldades para que seus filhos tenham acesso à escola e nela permaneçam até concluir a Educação Básica; por isso, é fundamental à ciência olhar para os estudantes e as políticas educacionais. Com o objetivo de cotejar essas discussões se consolidou um grupo de pesquisa que busca discutir as relações entre educação, pobreza e desigualdade social,¹ mesclando olhares de diferentes Ci-

¹Trata-se de uma parceria interinstitucional de pesquisadores da área de Educação da Universidade Federal do Paraná e da área de Ciências Sociais (Sociologia e Ciência Política) da Universidade Federal do Ceará que discutem o tema da educação, da pobreza e da desigualdade social. A formação desse grupo (que conta com estudantes de iniciação científica e de mestrado de ambas as universidades), começou a ser desenhado em 2014 quando o Ministério da Educação por meio da antiga Secretaria de Educação Continuada, Alfabetização, Diversidade e Inclusão/Ministério da Educação (SECADI/MEC) atualmente Secretaria de Modalidades Especializadas, lançou o Programa Educação, Pobreza e Desigualdade Social

ências Humanas.

Para a realização de tal pesquisa toma-se como ponto de partida a ideia de trajetórias escolares, ou seja, o acompanhamento dos estudantes em situação de pobreza por meio da utilização de dados referentes aos estudantes vinculados ao Programa Bolsa Família (PBF) e posteriormente o seu cotejamento com dados das escolas brasileiras. O PBF é um programa federal de transferência de renda voltado para pessoas em situação de pobreza, em funcionamento desde 2004, possui um caráter intersetorial e estabelece critérios relacionados a condicionalidades para permanência no programa. Na área da educação, a frequência à escola é a condição para o recebimento do benefício pelas famílias.

O presente artigo busca discutir as questões metodológicas que envolvem o desenho de uma pesquisa quantitativa que coteje o olhar sobre o perfil dos estudantes e seus percursos escolares, relacionados à permanência no PBF e ao cumprimento da condicionalidade na educação. A discussão neste momento centra-se em como lidar com o manejo de uma base de dados administrativa e em como usar tais dados para uma reflexão sobre os impactos de uma política pública de distribuição de renda dentro do campo educacional. Isto também traz a necessidade de um debate sobre o conceito de pobreza, seu uso no Brasil e o modo como está relacionada ao acesso à educação, anotando, ainda, como as políticas públicas articulam ações para a garantia de tais direitos.

A proposta aqui apresentada é de construção de uma base de pes-

que contava com 3 eixos de ação: formação (por meio de cursos de Especialização e Aperfeiçoamento para profissionais envolvidos com o Programa Bolsa Família), pesquisa (sobre a temática) e por fim, difusão dos conhecimentos. A parceria entre essas duas instituições consolida-se através de um coletivo pesquisando a temática, mesmo findada a iniciativa governamental.

quisa com dados do Sistema Presença, sistema que coleta informações sobre a frequência de estudantes beneficiados pelo PBF em escolas públicas ou privadas. O banco, organizado sob a perspectiva de análise longitudinal, tem como coorte selecionada para a pesquisa os estudantes nascidos em 2002 e registrados no sistema em 2015. A data se justifica pelo início da pesquisa e o recorte dos estudantes se dá porque aos 13 anos estariam na transição do Ensino Fundamental para o Médio e que é marcado por evasões e reprovações. Para a construção da análise longitudinal, conforme será detalhado adiante, utilizar-se-ão os dados referentes aos anos de 2015, 2016 e 2017.²

A escolha da coorte é dada por compreender a importância de pensar a escola para a juventude pobre, verificar seu percurso e quantos desses estudantes têm seu direito à educação protegido, incluindo a conclusão do Ensino Fundamental e o acesso ao Ensino Médio. A escolha do Ceará e do Paraná se dá por serem os estados dos pesquisadores vinculados.

Discussões sobre metodologia de pesquisa são pouco exploradas na área de Educação. Porém, entende-se que são necessárias, especialmente quando se trabalha com um grande volume de dados. A discussão de construção e organização dos dados é extremamente importante, não apenas por permitir que se repliquem os estudos, mas porque colaboram com entendimento dos resultados, além de possibilitar outros olhares sobre a utilização dos dados e novas formas de desenho metodológico.

O presente artigo, para atender os objetivos aqui propostos, inicialmente, discute as ideias de pobreza e desigualdade que ancoram a

²Cabe ressaltar que tal banco será cotejado, posteriormente, com informações sobre as condições de escola. A discussão metodológica do cotejamento desses dados com as informações do Censo Escolar não será debatida no presente texto devido à sua complexidade.

discussão e contextualizam o PBF, haja vista que esse usa uma determinada medida de pobreza para definir quem pode ou não ser beneficiado. Posteriormente, o texto discute o trabalho metodológico com as bases de dados do PBF, descrevendo os passos para sua construção. E por fim apresenta algumas estratégias metodológicas de análise dos dados em perspectiva transversal e longitudinal.

Pobreza, desigualdade e o Programa Bolsa Família: questões teóricas e conceituais

A afirmação que o nível socioeconômico das famílias tem impacto sobre o “sucesso” escolar dos estudantes é antiga e tem como expoente os estudos norte-americanos que resultaram na publicação do Relatório Coleman³ na década de 1960. Nele constatou-se a influência dos aspectos econômicos, sociais e culturais das famílias sobre a aprendizagem, ou seja, o *background* familiar impacta significativamente no percurso do estudante e no resultado dos testes de avaliação em larga escala. Ainda que se descreva o impacto dos professores e da estrutura das escolas, a ideia central do relatório referia-se ao baixo impacto da ação escolar. Desde a publicação do Relatório Coleman e seus desdobramentos, outros estudos (MADAUS, AIRASAIN, KELLAGHNA, 2008; RUTTER, *et. al.*, 2008) foram feitos com intenção de dialogar com os achados e contrapor as análises, apontando suas limitações.

Na França, aparecem as teses sociológicas de Pierre Bourdieu e Jean-Claude Passeron (2014) sobre a “reprodução”. A principal delas

³James S. Coleman foi o coordenador da pesquisa encomendada pelo governo norte americano que tinha como objetivo “descrever a distribuição diferencial das oportunidades educacionais nos Estados Unidos, em meados da década de 60” (BROOKE; SOARES, 2008, p. 14).

— a de que a instituição escolar (francesa) legitimaria a dominação e a reprodução de desigualdades sociais — foi um divisor de águas para a Sociologia da Educação e segue sendo importante referência de interlocução para as pesquisas educacionais.

Essas novas formas de investigação deram origem aos estudos, denominados de “efeito-escola”. Tratam-se de estudos que, mesmo levando em conta a reprodução, ou apesar do efeito *background*, consideram a escola como fator significativo no aprendizado e nos resultados estudantis. Tais pesquisas apresentam novos questionamentos de caráter teórico e metodológico, buscando formas mais sofisticadas de tratamento dos dados obtidos pelos estudos.⁴

As pesquisas, geralmente de grupos vinculados ao campo da Sociologia da Educação e ou das políticas educacionais, têm apresentado dificuldades em debater simultaneamente o “efeito-escola” e as desigualdades sociais de origem. Em alguns casos, o nível socioeconômico não é objeto de debate das pesquisas e tampouco das políticas públicas, sendo dado como algo “certo”, a ser apenas constatado.

Nesse sentido, parece ímpar colocar o debate do nível socioeconômico em um contexto mais amplo que implica a discussão sobre pobreza e desigualdade. Ao relacionar educação e pobreza, estudos afirmam (DUARTE, 2012; DUARTE; YANNOULAS, 2011; BONFIM, 2018) a necessidade de estabelecer contornos mais conceituais sobre a temática. Embora constata-se um crescente interesse, os debates

⁴Há uma vasta literatura sobre o efeito do nível socioeconômico dos estudantes nos seus resultados (ALVES; SOARES, 2007; ANDRADE; LAROS, 2007; FRANCO; BONAMINO, 2005; SOARES; ANDRADE, 2006; SOARES; COLLARES, 2016). Contudo, é possível identificar diversas características escolares que podem mitigar as influências das características sociais, denotando que o fenômeno das desigualdades escolares tem como núcleo as desigualdades sociais. É importante também ressaltar o vínculo inverso, com a produção de desigualdades sociais a partir das desigualdades educacionais e a produção de desigualdades próprias dos sistemas de ensino.

são desenvolvidos predominantemente por pesquisadores da área da Saúde, Economia e Serviço Social e estão mais direcionados aos impactos dos programas de transferência de renda e inclusão social. A educação é tratada apenas como uma variável positiva relacionada com o nível de renda. Assim, verifica-se a necessidade da consolidação da temática como um campo de reflexão, com pesquisas de continuidade e aprofundamento teórico.

A noção de pobreza absoluta está relacionada para alguns autores com a privação de acesso aos meios para a subsistência, isto é, “as condições básicas que devem ser cumpridas para sustentar uma existência física saudável” (GIDDENS, 2012, p. 344). Uma pessoa em situação de pobreza (no sentido absoluto) é aquela que não possui o necessário para prover a sua subsistência e de seus dependentes. Mas, sociologicamente falando, essa noção de pobreza absoluta varia tanto no tempo como no espaço. Ela varia inclusive nas diferentes conceituações dos diversos autores. A questão a se saber é se é possível definir com precisão quais seriam as condições básicas necessárias para a subsistência de humanos em qualquer contexto temporal e espacial. Isto é, um conceito de pobreza absoluta que pudesse ser universalmente aplicável, sociologicamente falando, parece ser difícil de estabelecer, haja vista que a mensuração, por mais objetiva que seja, terá sempre que definir marcadores que são, em última instância, subjetivos.

Dito isso, empiricamente, as instituições sociais operam recortes que são capazes de dizer onde começa e onde acaba a pobreza. Do ponto de vista das políticas públicas, tais definições são essenciais e orientam onde os diferentes atores sociais investem sua energia, tempo e recursos. No entanto, não se pode deixar de sublinhar que uma definição de pobreza, qualquer que seja, está fundada em valores subjetivos, sejam eles morais, éticos, políticos, religiosos ou outros.

Busca-se, assim, para esta empreitada de pesquisa, sublinhar algumas tentativas institucionais de objetivar a pobreza, sobretudo, aquelas que deram origem aos marcos legais do Programa Bolsa Família. Como exemplo deste esforço de definição da pobreza ressaltamos aqueles empreendidos pelos Relatórios do Desenvolvimento Humano (RDHs).⁵ De acordo com a edição de 2019, a referência internacional para se determinar que uma pessoa está na linha da extrema pobreza, e, portanto, não possui o mínimo necessário para prover a subsistência, é a mesma adotada pelo Banco Mundial (BM).

Desde 2015, o valor arbitrado pelo BM para aferir as situações de pobreza é de U\$ 1,90 por pessoa, por dia.⁶ Entretanto, tal valor foi considerado demasiadamente baixo e linhas mais altas foram estabelecidas pela instituição: U\$ 3,20 *per capita*, de rendimento diário, para países de renda média baixa e U\$ 5,50 em países de rendimento alto.

Se compararmos os critérios de corte de renda arbitrados pelo PBF e pelo Banco Mundial, verificamos quão discrepantes são os valores utilizados como referência. No âmbito do PBF, o recorte de renda para definir uma situação de pobreza é de até R\$ 178,00 mensais — valor que delimita quem receberá ou não o benefício. Desta forma, o critério de renda estabelecido pelo BM é cerca de 3 vezes superior

⁵Documentos editados anualmente pelo Programa das Nações Unidas para o Desenvolvimento (PNUD) com relatórios de dados, diagnósticos (sobre a vida social, econômica e política), prescrições e propostas diversas para amenizar os riscos inerentes à pobreza e a desigualdades intensas. Os RDHs, segundo Maria José de Rezende (2015), além de condensarem múltiplas vozes, desde grupos de interesses econômicos e políticos dominantes a militantes de movimentos sociais populares, agrupam “inúmeras informações, muitas das quais colhidas (...) de debates, discussões, pesquisas e investigações produzidas ao longo de anos pelas Ciências Sociais e Humanas sobre pobreza, desigualdades, participação política, democracia, acesso à educação e à saúde.” (REZENDE, 2015, p. 34).

⁶Com o critério de apenas U\$1,90 por dia, o próprio Banco Mundial (WORLD BANK, 2018) indica que 736 milhões de pessoas vivem abaixo da linha da extrema pobreza no mundo.

ao determinado pelo PBF. Esse simples exemplo é suficiente para demonstrar como não existe uma leitura unívoca do que seria ser pobre. Diferentes instituições têm leituras diferentes. E os diferentes atores que são regidos por essas definições se batem para modificar ou para manter as linhas de corte.

A pobreza de rendimento ou a insuficiência de renda, embora tenha grande peso na mensuração escalar da pobreza, é apenas uma das suas múltiplas expressões. Tendo em tela que a pobreza está relacionada para alguns autores ao acesso a meios para prover a subsistência, isso passa pela questão da nutrição, da saúde, da habitação e da segurança. Para esses as dimensões imediatamente vinculadas à sobrevivência cotidiana dos indivíduos não podem estar simplesmente circunscritas ao valor de U\$ 1,90 por dia. “Os mais desfavorecidos padecem de privações sobrepostas, de normas sociais discriminatórias e da ausência de capacitação política.” (PNUD/RDH, 2019, p. 63).

Assim, se constata que a definição de pobreza é um fenômeno social que suscita diferentes interpretações. Basta que se pense que trabalhos considerados canônicos do campo de combate à pobreza, especialmente inspirados por Amartya Sen,⁷ vinculam-na às noções de capacidade e liberdade no sentido de produzir indicadores que possam lidar simultaneamente com suas múltiplas dimensões.

Quando apresenta a noção de capacidade, Sen (1993; 2008), está preocupado com as possibilidades concretas dos humanos executarem e realizarem seus planos com liberdade. Em outras palavras, terem meios para construir o futuro que almejam ou ainda, em um nível

⁷ Amartya Kumar Sen é um economista indiano, atualmente pesquisador da Universidade de Harvard, consultor do PNUD/ONU para questões relativas ao desenvolvimento humano. Foi ganhador do prêmio Nobel de Economia, em 1998, justamente pelos resultados de suas pesquisas sobre problemas fundamentais em bem-estar social.

mais profundo, terem perspectivas que possibilitem sonhar com um futuro.

Os dados trabalhados ao longo de décadas pelo PNUD demonstram que países que mantêm em níveis altos a desigualdade social, tendem a possuir mais pessoas consideradas (por critérios objetivos subjetivamente construídos) como estando em situação de pobreza e tendem a não diminuir estas cifras de pobreza (PNUD/RDH, 2019, p. 84).⁸ Em outras palavras, o crescimento econômico não significa necessariamente a redução das desigualdades sociais e, portanto, pode não contribuir para a redução da pobreza.

No Brasil, alegando ação estratégica para o cumprimento do desígnio constitucional de “erradicar a pobreza e a marginalização e reduzir as desigualdades sociais e regionais” (BRASIL, 1988, Art. 3º, II), em 2004, o governo instaurou com o PBF uma estratégia política pública de transferência de renda direta condicionada que resultou em notável redução dos índices que o poder público define como pobreza no país.

O Programa Bolsa Família é programa social de combate à fome, instituído com o objetivo de mitigar os efeitos imediatos da pobreza, potencializando acesso às políticas públicas essenciais nas áreas da educação e da saúde. Esse Programa tem sido objeto de estudo de diversos pesquisadores das Ciências Humanas (ZIMMERMANN, 2006; LEÃO REGO; PINZANI, 2013; PIRES, 2013; MARTINS; RÜCKERT, 2019).

O PBF nasce como uma política intersetorial, consolidando-se como uma política que articula ações de assistência social, educação e saúde, sob a responsabilidade do extinto Ministério do Desenvolvi-

⁸O relatório de 2019 apresentou estudo indicando que a redução do índice Gini teria impacto muito mais significativo na redução da pobreza do que o crescimento do PIB per capita (PNUD/RDH, 2019, p. 84).

mento Social;⁹ e apresenta critérios de elegibilidade, com uma linha de corte de renda que determina quem será ou não contemplado pelo Programa.

Em 2015 (ano demarcatório do início de nossa coorte), esse recorte de renda estabelecia que uma família se enquadrava na situação de extrema pobreza se contasse com uma renda *per capita* entre zero e R\$ 77,00 mensais e era considerada em condição de pobreza, se possuísse uma renda *per capita* entre R\$ 77,01 e R\$ 154,00 mensais.

Para garantir a manutenção do recurso mensal às famílias beneficiárias, o desenho do Programa prevê o cumprimento de condicionalidades que envolvem as políticas públicas da educação, da saúde e da assistência social. No âmbito educacional, área de interesse da presente pesquisa, o PBF determina que as crianças e adolescentes, além de devidamente matriculados nas instituições de ensino, devem cumprir a frequência escolar mínima exigida pelo Programa: 85% para a faixa etária de 6 aos 15 anos e de 75% para os adolescentes entre 16 e 17 anos de idade.

A determinação de contrapartidas às famílias beneficiárias para garantir o acesso a uma renda mínima apresenta muitas controvérsias e opiniões divergentes no que remete à garantia de direitos (ZIMMERMANN, 2006; BICHR, 2010; PIRES, 2013), o que sinaliza a ausência de um conceito universal não só do que é pobreza, mas também das políticas para dela sair.

Quando há descumprimento de condicionalidades, as famílias enfrentam os efeitos gradativos, também chamados de sanções, que vão desde uma advertência — que não gera qualquer efeito sobre o recurso monetário — até a situação de bloqueio ou suspensão do benefício. O

⁹Atualmente o PBF é de responsabilidade do Ministério da Cidadania, especificamente da Secretaria Especial do Desenvolvimento Social.

poder público é responsável pelo acompanhamento da frequência escolar dos estudantes pobres, condicionada ao recebimento do valor monetário. Esta dinâmica possibilita ao poder público identificar os casos de não cumprimento, e traçar ações para o acompanhamento dessas famílias.

Percurso metodológico para a construção de um banco de dados

No campo das Ciências Humanas, as discussões conceituais são bastante presentes, e versam sobre embates teóricos assumidos à luz de determinada forma de enxergar o mundo, a realidade e o objeto a serem investigados. Discussões metodológicas, por outro lado, raramente ocupam espaço central nas discussões e, as descrições mais amplas dos passos e definições, muitas vezes, são apresentadas de forma rasa e sem maiores detalhamentos.

No caso das pesquisas quantitativas, que não são tão frequentes nos estudos educacionais no Brasil (GATTI, 2004; PEREIRA; ORTIGÃO, 2016), a descrição mais ampla da metodologia é válida para compreensão e entendimento dos resultados. O Brasil conta com diversos levantamentos e bases de dados, sejam eles administrativos ou não. Os Censos Populacionais, por exemplo, são realizados desde o Império e se somam às Pesquisas Nacionais de Amostra Domiciliar e diversas pesquisas temáticas. No campo educacional, o Instituto Nacional de Estudos e Pesquisas Anísio Teixeira (*INEP*) realiza levantamentos anuais sobre as escolas, turmas, matrículas, docentes e gestores/diretores¹⁰ brasileiros, além de avaliações que produzem dados que podem auxiliar a pesquisa educacional. Apesar disso, ainda

¹⁰ A coleta referente aos gestores iniciou-se apenas em 2019.

somos considerados um país com modesta tradição estatística (IBGE, s/d) no que se refere ao uso desses dados.

A pesquisa quantitativa, segundo Gatti (2004), é importante para compreender melhor as ações do Estado, além de permitir uma análise mais ampla do fenômeno a ser observado. Segundo a autora: “A análise da política educacional não pode prescindir do confronto do que é declarado nos pressupostos e metas que as orientam, com os dados a respeito dos resultados alcançados [...]”. (GATTI, 2004, p. 19).

Há muitas possibilidades de tratamento dos dados quantitativos. Em geral, as pesquisas apresentam recortes transversais ou séries históricas. Os estudos transversais são aqueles que tratam as informações obtidas em determinados momentos precisos, como o caso dos levantamentos já enumerados, durante os recenseamentos populacionais ou educacionais. Os estudos que analisam séries históricas apresentam o movimento das informações coletadas em estudos transversais em uma sequência temporal.

Outra possibilidade de tratamento dos dados é apresentada pelos estudos longitudinais, que definem um recorte como ponto de partida para a análise, ou seja, estabelecem um critério para definir uma coorte inicial e organizam as informações coletadas ao longo de um período temporal de forma a acompanhar a movimentação dos dados referentes à coorte. Neste caso, os indivíduos do estudo não mudam ao longo do tempo, permanecem os mesmos para que seja possível analisar as mudanças desse grupo.

Pesquisas longitudinais podem ter caráter prospectivo ou retrospectivo. No primeiro caso, os estudos coletam os dados na medida em que o período temporal analisado avança e o planejamento de pesquisa prevê a coleta dos dados em situações futuras. Já os estudos retrospectivos utilizam dados coletados anteriormente ou informações

sobre eventos passados (GAYA; BRUEL, 2019).

Esta pesquisa se identifica como um estudo quantitativo, que utiliza uma composição de estratégias de análise transversal em série histórica e longitudinal retrospectiva, pois acompanha as informações sobre um conjunto de estudantes a partir de uma coorte inicial utilizando dados coletados no passado. Apresentaremos uma possibilidade de tratamento dos dados nas duas perspectivas a fim de discutir a articulação e complementaridade entre as análises descritivas de uma mesma variável, o que será detalhado à frente.

Os dados utilizados caracterizam-se como dados secundários e têm caráter administrativo, ou seja, não são coletados pelos pesquisadores e não são pensados/desenhados para um estudo, e sim, servem como mecanismo de acompanhamento da condicionalidade no âmbito da educação. A fonte de tais dados é o Sistema Presença, conhecido também como “Sistema de Acompanhamento da Frequência Escolar do Programa Bolsa Família (Sistema Presença) que foi desenvolvido em uma plataforma Web, a fim de permitir o acesso ao Sistema em todas as regiões do País” (MEC/SECADI, 2018, p. 3).

Esse sistema permite o acompanhamento da frequência e consequentemente da condicionalidade do programa no que se refere ao âmbito educacional. Tal acompanhamento é de responsabilidade do Ministério da Educação em parceria com os estados e municípios brasileiros. No âmbito dos estados e municípios, as secretarias estaduais e municipais, juntamente com as escolas, são essenciais para o funcionamento do sistema. Com esse sistema

[...] é possível saber o motivo da baixa frequência de qualquer aluno da base de acompanhamento (aproximadamente 17 milhões de crianças e adolescentes). A apropriação dessas informações é de extrema valia para a formulação de políticas públicas de enfrentamento de situações de abandono e evasão escolar e acionamento da rede de

proteção à infância [...]. (CURRALERO, 2010, p. 163).

O sistema de coleta das informações de frequência é feito de forma nominal, sendo necessário informar o cumprimento ou não da condicionalidade. A coleta é feita bimestralmente iniciando-se em fevereiro e finalizando em novembro. O Sistema conta com diversos relatórios, sendo que o utilizado para a presente pesquisa denomina-se de CON-DEDU (Condicionalidade Educação), o qual reúne informações sobre as condicionalidades da Educação. Tal banco contém as informações mostradas no Quadro 1.

Ainda que a coleta seja feita nominalmente, nos relatórios é possível apenas ter acesso ao Número de Identificação Social (NIS) e este é o mesmo para o aluno durante toda sua permanência no Programa. Os motivos e submotivos de baixa frequência permitem identificar as causas das ausências a escola e vão desde gravidez, problemas de saúde, situação de rua e greve até ausência de oferta escolar. A identificação dos motivos e submotivos de ausência na escola são importantes, primeiro porque colaboram no entendimento da realidade e das dificuldades enfrentadas pelos estudantes para manter-se na escola, bem como retratam questões externas ao cotidiano escolar (greve, enchente).

Como a coleta é feita bimestralmente foi necessário inicialmente cotejar os bancos de dados de 2015 a 2017 (período a ser analisado). Primeiramente, agregou-se todos os períodos de coleta de 2015, cabendo destacar que o número de estudantes em cada mês de coleta é diferente, ou seja, não necessariamente os estudantes permanecem todos os meses (períodos de coleta) cadastrados e/ou vinculados ao sistema, como se pode perceber na Tabela 1.

Esse total de estudantes acompanhados pelo PBF, corresponde a aproximadamente 48,5% dos estudantes brasileiros entre 6 a 17 anos (34.988.440). Cabe destacar que o número variável em cada período

Quadro 1 – Variáveis que compõem o banco CONDEDU do Sistema Presença

Variável	Descrição
Número da linha	Identifica a quantidade de beneficiários no ano e no mês respectivo
Tipo de registro	Caracteriza a informação como referente frequência do aluno
NIS do aluno	Refere-se ao código do estudante beneficiário do programa Bolsa Família, enquanto este estudante for beneficiário ele será identificado por esse código.
Data de nascimento do aluno	Informa o ano, mês e dia que o aluno nasceu
INEP Escola	Código da escola que o aluno beneficiário pertence
Série	Indica qual ano o aluno está cursando;
Situação do aluno	Indica se o aluno é ativo, se não conta com informação da escola, se o aluno não estuda porque o município não tem oferta educacional referente à série/ano que deveria estar cursando;
Frequência do primeiro mês	Identifica a frequência em um dos meses coletados
Motivo baixa frequência primeiro mês	Discrimina o motivo da baixa frequência
Submotivo baixa frequência primeiro mês	Especifica ainda mais o motivo da baixa frequência
Frequência do segundo mês	Identifica a frequência em um dos meses coletados
Motivo baixa frequência segundo mês	Discrimina o motivo da baixa frequência
Submotivo baixa frequência segundo mês	Especifica ainda mais o motivo da baixa frequência

Fonte: MEC, s/d. Organizado pelos autores.

se refere aos estudantes que vão sendo desligados do Programa e outros que são incorporados. O desligamento pode ocorrer por motivos

Tabela 1 - Total de estudantes vinculados ao Programa Bolsa Família, Brasil, 2015

Mês	Total beneficiários
Fev/mar	17.259.683
Abr/maio	17.272.090
Jun/jul	16.992.327
Ago/set	17.046.955
Out/nov	16.757.353

Fonte: MEC/Sistema Presença (2015), dados trabalhados pelos autores.

de cancelamento do benefício, ou seja, pelo não cumprimento da condicionalidade¹¹; ou por outros motivos, tais como, aumento da renda, término do Ensino Médio etc. Também há casos de estudantes que ficam suspensos do banco por um período, retornando em coletas posteriores.

Alguns estudantes, em alguns períodos de coleta, não contam com informações de escola, série ou mesmo de frequência. Tais fatores relacionam-se com frequência à transferência de escolas ou municípios que tais estudantes fazem durante seu percurso escolar. Além de problemas de atualização da situação pela família junto ao Cadastro Único e do próprio caráter das informações coletadas pelo Sistema.

Tendo tais situações a perspectiva de análise longitudinal, é melhor para acompanhar a trajetória em relação a uma análise transversal, que toma como referência apenas um momento determinado e

¹¹Em relação ao não cumprimento da condicionalidade, consideram-se 4 efeitos: advertência, bloqueio, suspensão e cancelamento. “Para a progressão de um efeito para o seguinte, considera-se o intervalo de seis meses. Por exemplo, caso uma família tenha sido advertida, em março de 2014, e venha a incorrer em um novo descumprimento, em período inferior ou igual a seis meses (ou seja, até setembro de 2014), o efeito progride para bloqueio. Mas, se o novo descumprimento ocorrer em prazo superior a seis meses, o efeito será a advertência, isto é, reinicia-se a aplicação gradativa dos efeitos”. (MDS, 2015, s/p).

considera a totalidade de estudantes, sem captar as mudanças de um mesmo indivíduo ao longo de um período. A coorte da pesquisa foi definida a partir de três critérios que consideraram o cruzamento entre um conjunto de variáveis, quais sejam: 1) ser estudante vinculado ao PBF, com pelo menos um registro de frequência no Sistema Presença no ano letivo de 2015; 2) ser estudante nascido em 2002; 3) ser estudante matriculado em escolas dos estados do Ceará ou do Paraná em pelo menos um dos registros de frequência do ano de 2015.

Para selecionar tal coorte foi necessário criar uma variável de ano de nascimento por período de coleta e selecionar os estudantes nascidos nesta data em qualquer um dos meses estabelecidos. Os filtros aplicados de ano de nascimento e dos dois estados consolidou um banco de dados com 161.261 estudantes beneficiários. Esses estudantes, foram definidos como a coorte a ser analisada. As coletas de 2016 e 2017 foram agregadas e seu cotejamento foi feito tomando como base o número de estudantes citados, ou seja, agregaram-se informações dos mesmos estudantes nos outros 10 períodos de coleta bimestral (5 do ano de 2016 e 5 do ano de 2017), quando estavam disponíveis.

Após tal cotejamento, iniciou-se um trabalho de organização do banco, sendo que um dos primeiros elementos considerados se refere ao ano de nascimento. Analisando essa variável percebeu-se que havia em torno de 0,5% dos estudantes que tinham informações diferentes para as datas de nascimento. Objetivando compreender melhor a situação dos estudantes em relação ao ano de nascimento, realizou-se uma comparação entre a quantidade de meses de acompanhamento e o ano de nascimento informado, possibilitando a construção de um percentual (*Tabela 2*) de informações possivelmente incorretas referentes ao ano de nascimento (comparada com o número de acompanhamen-

Tabela 2 - Percentual de informações de ano de nascimento diferente de 2002, Ceará e Paraná, 2015 a 2017.

Informação com variação	Frequência	%
Mais de 93%	115	0,07
Entre 92% e 51%	329	0,20
Entre 1 a 50%	499	0,29
Nenhuma	160.318	99,42
Total	161.261	100,00

Fonte: MEC/Sistema Presença (2015, 2016, 2017), dados trabalhados pelos autores.

tos). Por exemplo, um estudante com 15 meses de acompanhamento e 8 informações incorretas de série obtêm um percentual de 53% de informações incorretas.

Tendo em vista que um banco de dados, por vezes, pode apresentar inconsistências e erros de digitação e objetivando “perder” o menor número de casos (estudantes) a serem analisados, estabeleceu-se que não seriam considerados nas análises os estudantes que obtiverem, na proporção entre ano de nascimento informado com variação de valores e os meses de acompanhamento, um percentual superior a 50% (ou seja quando mais de 50% das informações de meses estiverem provavelmente incorretas). Foram assim excluídos da análise 444 estudantes e a coorte analisada passou a ser composta por 160.817 estudantes nascidos em 2002 e que estiveram vinculados ao PBF no Paraná e Ceará durante o ano de 2015.

No que se refere à coorte analisada, em torno de 33,2% dos estudantes estudaram no Paraná e 66,8% no Ceará. A grande diferença entre o percentual de estudantes do Programa Bolsa Família nos dois estados não é percebida apenas na coorte analisada: ao comparar o percentual de vinculados ao Programa em nível nacional, têm-se que 7% deles são do Ceará e 3% do Paraná, ou seja, o número de estudan-

tes cujas famílias são beneficiárias em cada estado é bastante diferente. Ressalta-se que, em 2015, o número de matrículas do Paraná e Ceará era bem próximo, sendo de 2.568.570 e 2.229.711.¹² O percentual de beneficiários do PBF em relação ao total de estudantes do Estado também é diverso, sendo de aproximadamente 8% no Paraná e 47% no Ceará¹³ para a coorte que estamos estudando. Essas diferenças são explicadas pela situação de pobreza de cada estado: segundo dados do IBGE (2019) o percentual de pessoas vivendo em pobreza ou extrema pobreza no CE circundava em 43% e no PR de 14%.

Considerando as ausências de informação, as saídas e retornos dos estudantes ao banco, optou-se por inicialmente versar a análise das trajetórias dos estudantes que permaneceram vinculados ao PBF durante todo período analisado (15 bimestres). Esses estudantes representam 52% (83.545) de toda a coorte.

Findada a organização mais ampla do banco, apresenta-se o tratamento metodológico que será dado às variáveis que apresentam ausência de informação. Para tal exemplificaremos os procedimentos a partir da variável “série” em duas perspectivas, articulando as análises transversais da série histórica e análises longitudinais para o mesmo conjunto de dados, ambas com caráter descritivo.

¹²Em 2019 o número de matrículas no PR era de 2.572.007 e no Ceará de 2.161.816 (LABORATÓRIO DE DADOS EDUCACIONAIS, 2020).

¹³Os dados calculados consideram as informações referentes ao sistema presença de abril/maio (2015) que coincidem com o período de coleta dos dados do Censo Escolar. Ao considerar os estudantes de 6 a 17 anos que são o público alvo da condicionalidade educacional do PBF o percentual de vinculados sobe para 26% no PR e 69% no CE.

Como olhar alunos sem presença e sem série no sistema? Trabalhando com *missings* e caracterizando tipos de não respostas

Sempre se pode lidar com as ausências de respostas no sistema da forma mais corriqueira nas Ciências Humanas: ignorando as não respostas (os *missing* do sistema) e levando-se em conta apenas os casos válidos. No caso de trabalho com populações como a dos estudantes beneficiários do Bolsa Família, ignorar os casos com ausência de informação pode corromper as análises, haja vista que diferentes estudos mostram como as não respostas têm forte correlação com os mais pobres e os menos educados, como já mostraram Bourdieu (2017) ou Collomb (1977).

Tomando como referência esses estudos que apontam que a ausência de informação, ou seja as “não respostas” estão diretamente relacionadas a situações de fragilidade social, supõe-se que, no banco de dados utilizado, uma boa parcela dos estudantes que constam como “sem informação de série” sejam exatamente aqueles egressos das famílias mais vulneráveis, os mais pobres entre as famílias pobres e menos educadas entre aquelas de pouca educação formal.

As possibilidades de situações para que ocorra a “não resposta” são as mais diversas: os responsáveis pela inclusão das informações podem deixar em branco; os alunos podem estar em processo de transferência de escola sem formalizar tal ação junto ao serviço social; podem estar em processo de reclassificação em outra série dentro da própria escola ou transferido para Educação de Jovens e Adultos (EJA) em outro estabelecimento; ou ainda, diversas situações pelas quais passam os mais pobres em seu percurso escolar sem que necessariamente desejem por elas passar.

Para conhecer a trajetória desses alunos acredita-se necessário, o quanto possível for, não perder dados. Sublinha-se que esta pesquisa tem interesse em melhor conhecer a trajetória de alunos oriundos das classes empobrecidas e, portanto, o esforço de encontrar informações sobre o maior número possível dos membros da coorte visa evitar passar ao largo dos que mais necessitam do auxílio econômico e dos serviços de educação escolar. Assim, a metodologia que se adota nesta pesquisa procura reduzir estas ausências e será explicada na sequência do texto.

Primeiro, serão apresentados alguns dos procedimentos para trabalhar com o conteúdo ausente (os *missings*). Qual o sentido desta ausência? Os momentos da série histórica em que as não respostas aparecem precisam ser investigados (se no início do ano letivo, em final de semestre, ou ao final do ano escolar). Existem variações nas taxas de não resposta no correr do ano ou entre os diferentes anos analisados?

Em segundo lugar, tenta-se buscar, entre os dados ausentes na série histórica, aqueles que podem ser encontrados, fazendo cruzamentos deles com outras informações conhecidas. Serão apresentados adiante alguns exemplos de como a metodologia com a qual a pesquisa se propõe a trabalhar funciona (mesmo que neste exemplo não se apresente ainda uma análise dos dados gerados, haja vista que o objetivo é ilustrar como a metodologia adotada busca recuperar a informação ausente).

Existem dois tipos de informações no banco: as que são obrigatórias e as que não o são. Os diferentes itens que devem ser preenchidos pelo coordenador municipal não se comportam da mesma maneira quando não recebem uma informação: existem itens que são obrigatórios para que o preenchimento da base continue e itens que não são

obrigatórios e que permitem que o operador prossiga sem tê-los respondido.

É possível exemplificar essa diferença em duas variáveis do Banco. Dados quanto à frequência do aluno (cumpriu/ não cumpriu a condicionalidade) são imprescindíveis para a administração do benefício do Bolsa Família, e esses são essenciais. Sem essa resposta o sistema trava e o operador não consegue seguir com seu trabalho. Já quanto à informação relativa à série do aluno, se o operador deixar o espaço sem resposta, o sistema segue adiante, sem nenhum entrave. Em conversa com um coordenador municipal sobre a questão, ele nos informa:

Muitas vezes o operador do sistema está muito atarefado e não tem tempo de preencher todas as informações. Como preencher a série [do aluno] não é obrigatório, o operador deixa em branco. O Sistema [Presença] também tem problemas... por exemplo, no mês “tal” o operador preenche a série do menino, vamos dizer que foi o sétimo ano; no mês seguinte o Sistema deveria repetir a informação, mas aí ela vem em branco ou errada, aí o operador corrige. E o problema se repete e se repete. Aí, como a informação não trava o sistema, há ocasiões onde o operador, por estar muito atarefado, deixa em branco a situação daquele menino (OPERADOR MUNICIPAL).

Como a pesquisa pretende acompanhar a trajetória dos estudantes vinculados ao PBF ao longo dos três anos analisados, foram selecionadas inicialmente duas variáveis diretamente relacionadas à vida escolar: a série em que os estudantes estão matriculados e a situação do aluno em relação ao vínculo escolar. Ao analisar as duas variáveis, percebeu-se que há grande percentual de ausência de informação sobre a série, mesmo para estudantes que possuem registro de vínculo escolar ativo. Assim, foi necessário que o grupo de pesquisa se voltasse para aspectos metodológicos do tratamento das informações ausentes no sistema.

O caso da ausência de informação sobre a série é uma parte importante dos dados que aqui trabalhamos e representa em média 31,73% das respostas coletadas ao longo dos quinze bimestres (com um desvio padrão de 15,02). Número bastante expressivo! Esse fato é pelo menos parcialmente, explicado pela possibilidade de o operador deixar essa informação em branco e prosseguir alimentando o sistema.

Como pode-se observar na Tabela 3, em média 91,7% dos casos onde essa informação não foi preenchida dizem respeito a alunos de quem se conhece tanto a escola que frequentam, como se sabe que estão ativos (seja cumprindo ou não a condicionalidade da educação). Provavelmente essa ausência reflete uma das situações em que o responsável pela consolidação da coleta não quis ou não pode preencher a informação por não ter tempo, por achar que ela não tem importância ou então por desconhecer-la. Na outra ponta encontram-se exatamente os alunos que estão matriculados em escola sem INEP e, certamente, a informação neste caso é mais difícil. Sabe-se também que essa situação pouco influencia a ausência de informação, pois ela representa apenas 0,3% das ausências.

Entre os dois extremos se posicionam as situações dos alunos que não encontraram oferta ou atendimento educacional (2,4% dos casos) e, a dos de alunos que não foram encontrados na escola e tampouco no município (5,4% dos alunos sem informação de série). Em tese, poder-se-ia saber em qual série o aluno sem oferta de educação estaria, no entanto, o operador do sistema provavelmente achou mais importante informar a ausência de oferta e deixou em branco a referência à série.

Tabela 3 – Distribuição percentual média das não respostas sobre a série do aluno em função de seu vínculo escolar, Ceará e Paraná, 2015 a 2017

Ausência de informação da série do aluno X sua situação de vínculo escolar	%	Desvio padrão
Aluno ativo: matriculado em uma escola com INEP	91,7	3,3
Aluno não se encontra na escola onde estava em nenhuma outra escola em seu município	5,4	2,1
Aluno com inexistência de oferta/atendimento educacional	2,4	1,4
Alunos que estão matriculados em uma escola que não possui INEP	0,3	0,1

Fonte: MEC/Sistema Presença (2015, 2016, 2017), dados trabalhados pelos autores.

Trabalhando com dados ausentes

A seguir, mostramos um exemplo de como trabalhamos com dados ausentes dentro do Sistema Presença. Na apresentação deste exemplo, o objetivo é, sobretudo, mostrar as possibilidades de análise de “não respostas”. Para tal, serão tratadas duas situações descritas acima (*Tabela 3*), haja vista serem as que mais colaboram para a composição da categoria “sem série informada”. A saber: a subcategoria 1) “*alunos matriculados em escola com INEP*” e a subcategoria 2) “*aluno não se encontra na escola onde estava e em nenhuma outra escola de seu município*”. A primeira tende a diminuir ao longo do tempo, enquanto a segunda sofre uma ligeira tendência de crescimento, como pode ser visto pelas retas negras (ou linhas de tendência) que percorrem o desenho feito pela taxa de não resposta ao longo do tempo (linha azul) no *Gráfico 1* a seguir.

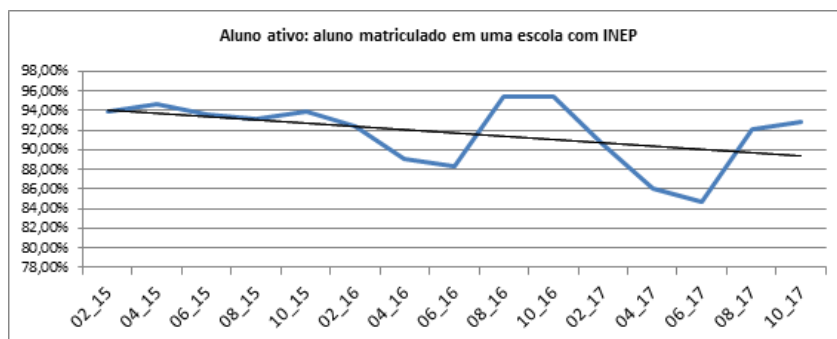


Gráfico 1: Percentual de aluno ativo entre os que compõem a subcategoria sem série indicada, Ceará e Paraná, 2015 a 2017.

Fonte: MEC/Sistema Presença (2015, 2016, 2017), dados trabalhados pelos autores.

As não respostas sobre a série do aluno nos dois Estados são compostas, em média, por 91,3% (com um desvio padrão de 3,3 pontos) de estudantes ativos (de quem se conhece a escola, se sabe se estão ou não cumprindo a condicionalidade). A coleta de dados informa, por ausência, que parte dos alunos, nos dois estados, não tem série informada a não ser o não preenchimento pelo operador e os problemas do sistema, explicitado anteriormente. Para este grupo, não existe razão aparente que explique a ausência do dado. Nota-se também que a frequência dessas não respostas, nessa categoria, pouco varia em 2015: ela sobe um pouco em 04/15 e volta a cair abaixo do valor inicial em 06/15, tornando a subir para o valor de início da série em 10/15.

Em 2016, apesar da linha de tendência continuar apontando queda, os valores começam a ter maior amplitude: entre o início do ano letivo e o mês de junho os operadores do Sistema conseguem corrigir parte das ausências de informação, para voltar a perdê-las nos dois bimestres seguintes. Em 2017, as variações têm amplitude ainda maior, com o período que vai de 02/17 até 06/17, representando as menores

contribuições dos alunos ativos para essa categoria. Daí até o final do ano, as taxas voltam a crescer, alcançando um valor próximo e um pouco maior do que aquele do início da série histórica.

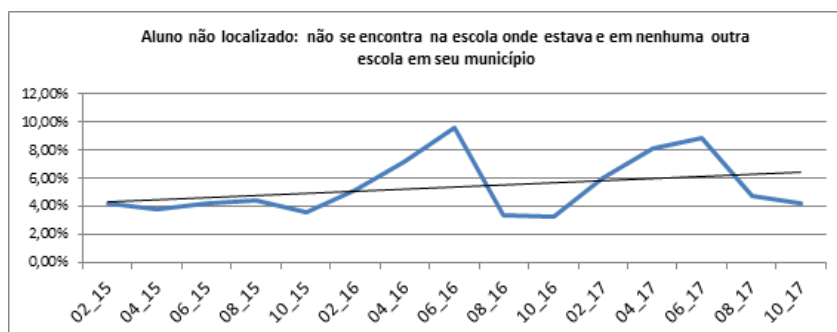


Gráfico 2: Percentual de aluno não localizado entre os que compõem a subcategoria sem série indicada, Ceará e Paraná, 2015 a 2017.

Fonte: MEC/Sistema Presença (2015, 2016, 2017), dados trabalhados pelos autores.

O que se observa no *Gráfico 2* é que a proporção varia muito pouco no ano de 2015. O fenômeno cresce a partir da primeira coleta de 2016 e vai até o final de junho do mesmo ano, quando alcança o pico da série (06/16). A partir deste momento, a dinâmica de registro de dados se inverte e chega aos menores valores desses três anos. No recomeço das atividades escolares do ano seguinte, a porcentagem volta a subir, indo nessa direção até 08/17, novamente alcançando valores próximos do ápice, voltando a cair até o final do ano de 2017.

Importa ressaltar que nesta perspectiva de análise, observando a movimentação transversal dos dados ao longo da série histórica, não é possível saber se a informação sobre a oscilação se refere aos mesmos estudantes ou se àqueles que aparecem sem informação em períodos diferentes da coleta de dados.

No caso do aluno ausente das escolas do município, de fato fica

mais complexo para o operador informar em que série a família conseguiu (se conseguiu) matricular o estudante. Segundo informou um coordenador municipal,

[...] esta situação [mudança de uma escola do município para uma escola de um município vizinho] não é incomum em regiões do município que fazem fronteira com outros municípios ou mesmo outros estados; as famílias migram em busca de melhores condições de vida ou mesmo levam os filhos para estudar em outras cidades, pois acham que a escola é melhor ou simplesmente porque está mais próxima do domicílio. (COORDENADOR MUNICIPAL).

É possível supor que muitas das informações que são incluídas ao longo dos períodos de coleta do Sistema Presença são conhecidas do operador do Sistema, mesmo que nem sempre ela seja registrada (haja vista as variações que corrigem os valores), sobretudo no que diz respeito aos alunos ativos frequentando escolas com INEP. Ao longo do tempo esses alunos voltam a ter sua série informada e há uma tendência de redução das não respostas.

Como estamos tratando com tendências que se opõem, pois quando uma aumenta a outra diminui, pelo momento, a hipótese é que o que mais faz mexer a balança entre “aluno ativo” e “aluno não encontrado no município” vem do trabalho dos operadores do sistema, que ao longo de cada ano buscam nos primeiros semestres (indo às vezes até a oitava coleta nos casos dos anos de 2015 e 2016) corrigir a informação que aparece incorreta ou vazia no Sistema. Inversamente, o final do ano escolar sempre tem aparecido como mais atribulado para os operadores. Isso faz com que nesse período ocorram com mais frequência as ausências de informações. Em todo caso, isso parece mais lógico do que imaginar que os operadores localizaram os alunos sem série informada em outros municípios.

Vale salientar que como a maior parte das ausências de informação

sobre a série do aluno diz respeito àqueles que estão matriculados em uma escola e se sabe que estão ativos no sistema, acredita-se que aqui se encontra um dos nichos onde o programa de pesquisa pode recuperar a informação. Considera-se que o estudo com caráter longitudinal permite, em certa medida, a recuperação de parte das informações perdidas nos diferentes momentos de coleta. Sem processos automáticos de imputação de valores ausentes, a comparação dos dados informados ao longo dos diferentes momentos de coleta permite estabelecer critérios para utilizar a informação sobre a situação do aluno, série, escola, município.

Além da ausência de informação nos casos das variáveis que não precisam ser obrigatoriamente preenchidas, verificou-se também a existência de variação de informações sobre o mesmo estudante entre os bimestres. A constatação de variação da informação ao longo do ano letivo, no caso dessas variáveis não pode ser tomada, a princípio, como erro de informação, uma vez que a alteração pode ser registro legítimo de movimentação própria das dinâmicas escolares.

Observando a movimentação das informações sobre a situação do aluno em uma perspectiva longitudinal verificou-se que há variação de informações ao longo do ano letivo para 10,8% dos estudantes da coorte em 2015; para 8% em 2016; e para 11,2% em 2017, como demonstra o *Gráfico 3*.

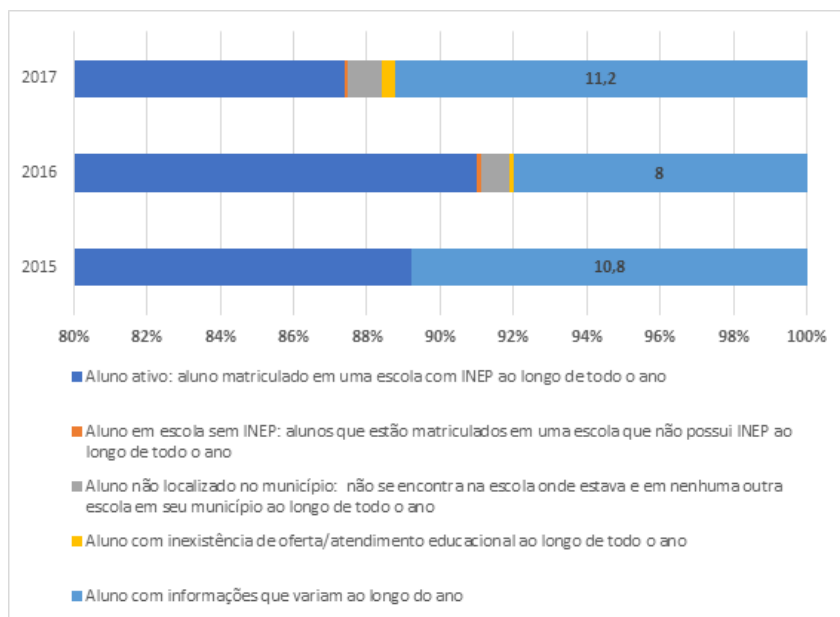


Gráfico 3 — Distribuição percentual dos registros longitudinais dos estudantes da coorte analisada em relação à situação de matrícula, Ceará e Paraná, 2015 a 2017.

Fonte: MEC/Sistema Presença (2015, 2016, 2017), dados trabalhados pelos autores.

A partir da reflexão sobre o impacto das informações ausentes e das informações que variam ao longo do período, foram definidos dois critérios para recuperar os dados relacionados aos estudantes: 1) Informações que se repetem em pelo menos em 50% dos registros ao longo do ano, seguindo o mesmo procedimento feito para recuperar as informações sobre o ano de nascimento dos estudantes; 2) Informações que aparecem pelo menos uma vez ao longo do ano letivo, considerando as informações mais recentes (preferencialmente do último bimestre de cada ano) quando se verifica a existência de disparidades entre as informações registradas ao longo dos cinco momentos de coleta em cada ano. Os dois procedimentos permitiram a recuperação de informações, com resultados distintos, que serão apresentados e

discutidos.

Os dados da *Tabela 4* indicam que a utilização dos procedimentos acima descritos permitiu a recuperação de informações sobre a série de matrícula para uma parte da coorte analisada, uma vez que é possível observar a redução do percentual de informação ausente para o conjunto dos estudantes a cada passo metodológico realizado. Observa-se também que os percentuais aumentam a cada ano da análise. Essa evolução da falta de informação não pode ser considerada aleatória e está relacionada à metodologia longitudinal de análise, pois os efeitos perversos das trajetórias irregulares possivelmente têm caráter cumulativo e se manifestam, pelo menos em parte, na ampliação desses percentuais entre 2015 e 2017.

Tabela 4 – Percentual de estudantes da coorte sem informação sobre a série de matrícula, 2015 a 2017.

Ano	Percentual de estudantes sem informação de série nos registros bimestrais ao longo do ano	médio de registros ao longo do ano	Percentual de estudantes sem informação de série pelo menos 50% dos registros bimestrais	Percentual de estudantes sem qualquer informação de série ao longo do ano
	% Médio	D. Padr.		
2015	14,82	7,39	9,7	9,7
2016	37,54	9,52	34,4	30,1
2017	42,84	9,51	41,2	33,9

Fonte: MEC/Sistema Presença (2015, 2016, 2017), dados trabalhados pelos autores.

As possibilidades de análises da situação do aluno em relação ao vínculo escolar a partir do percentual médio de estudantes sem infor-

mação sobre a série de matrícula, numa perspectiva transversal em série histórica já foram exploradas. Assim, apresentam-se as mudanças no cruzamento entre as duas variáveis depois de aplicados os procedimentos de recuperação das informações ausentes.

Observa-se que a recuperação de informações permitiu reduzir parcialmente o percentual de estudantes com informações ausentes sobre a série de matrícula bem como sobre a situação de vínculo com a escola. Assim, dos estudantes que permaneceram sem informação sobre a série, mesmo com a recuperação das informações que se repetem em pelo menos 50% dos registros bimestrais ao longo de cada ano letivo, foi possível ampliar o percentual indicado como aluno ativo para o ano de 2015. Contudo, nos anos de 2016 e 2017, o percentual se manteve abaixo da média geral de 91,7%, como indicam os dados da *Tabela 5*, pois 90,3% em 2016 e 88,4% em 2017 foram considerados como ativos em pelo menos 50% dos bimestres.

A segunda proposição para a recuperação dos dados ausentes ao longo do ano letivo refere-se à verificação de existência de registro como “aluno ativo” em algum dos bimestres dentro de cada ano letivo. Esse procedimento tornou possível encontrar informação sobre aluno ativo para todos os estudantes da coorte em 2015, pelo menos em um bimestre do ano, permitindo identificar a situação de aluno ativo em 100% dos casos analisados. Para o ano de 2016 foi possível encontrar informação sobre aluno ativo em pelo menos um bimestre para 98,8% dos casos. E em 2017 para 98,4% da coorte. Já em relação à variável “série”, após esse procedimento, 9,7% da coorte permaneceu sem informação em 2015; 30,1% em 2016; e 33,9% em 2017.

A ausência de informação na variável “série” mantém-se alta e muito acima da variação verificada na variável “situação do aluno”, portanto, em quaisquer dos exemplos, chegando a um terço da coorte

em 2017. Mesmo com a manutenção de parte das informações ausentes, o cruzamento entre as duas variáveis após a recuperação de todas as informações presentes no sistema, possibilitou a redução da ausência de resposta. Verifica-se, também, que mais de 97% dos estudantes que permanecem sem informação de série possuem pelo menos um registro como aluno ativo ao longo do ano, como informa a *Tabela 5*.

Tabela 5 — Distribuição percentual dos estudantes com informação ausente de série em relação à situação do aluno, depois da recuperação dos dados, Ceará e Paraná, 2015 a 2017.

Categorias de situação do aluno	Percentual de estudantes sem informação de série em pelo menos 50% dos registros bimestrais			Percentual de estudantes sem qualquer informação de série ao longo do ano		
	2015	2016	2017	2015	2016	2017
Variação de informação	0,1	1,6	0,2	0	0	0
Aluno com escola sem código INEP	0,2	0,4	0,5	0	0,1	0,3
Aluno não localizado no município	3,7	6,4	6,8	0	1,4	1,7
Aluno ativo	94,1	90,3	88,4	100,0	97,8	97,1
Inexistência de oferta	1,9	1,3	4,1	0	0,6	0,9
Total	100,0	100,0	100,0	100	100	100

Fonte: MEC/Sistema Presença (2015, 2016, 2017), dados trabalhados pelos autores.

Recuperar as informações que flutuam ao longo dos períodos de coleta do sistema presença é possível com a organização do banco de forma longitudinal, ou seja, com as informações de todos os anos analisados de cada estudante na mesma linha, assim, é possível comparar as informações registradas no sistema presença em diferentes bimestres e em diferentes anos. Importa ressaltar que quando há variação dos registros entre os bimestres não é possível identificar qual das informações corresponde à situação concreta do estudante e quais correspondem a erros de preenchimento.

Frente aos dados recuperados, é possível ressaltar a necessidade de análise sobre o conteúdo das categorias apresentadas pelas variáveis. A permanência de registro de adolescentes entre 13 e 15 anos de idade que não são considerados “ativos” porque não há oferta educacional que atenda à sua demanda por escolaridade revela uma situação de exclusão perpetuada pela ausência do poder público. Essas e outras questões de análise do conteúdo das informações presentes no banco serão tratadas pela equipe de pesquisa em outras oportunidades, uma vez que este artigo se propôs a refletir sobre aspectos metodológicos.

Considerações finais

O artigo traz um primeiro recorte de um amplo estudo sobre as trajetórias escolares de estudantes brasileiros da Educação Básica vinculados ao Programa Bolsa Família, usando dados do Sistema Presença e do Censo Escolar. Embora este texto se proponha à discussão metodológica do uso dos dados do Sistema Presença e de como lidar com dados ausentes no banco, é importante ressaltar que a pesquisa prossegue em andamento na análise dos dados trazidos no banco, com reflexões sobre os motivos de baixa frequência, o cumprimento (ou

não) das condicionalidades do PBF pelos estudantes, a reflexão sobre a trajetória escolar desses jovens e a relação entre esses alunos e as escolas em que estudam quanto a aspectos infraestruturais e educacionais; inclusive, numa perspectiva comparada entre os dois estados destacados no estudo.

O artigo procurou analisar estratégias metodológicas sobre como organizar e usar um banco de dados administrativo para realizar reflexões dentro do campo da educação, em particular, em sua relação com a pobreza, já que se tratam de estudantes vinculados ao PBF nos estados do Paraná e do Ceará.

Tendo em vista que bancos de dados administrativos orientam a execução de políticas públicas de várias naturezas e, com isso, têm grande impacto sobre seus beneficiários, estudos como o presente se mostram fundamentais. No caso analisado, o tratamento dos dados do Sistema Presença se mostra relevante para pensar não somente a maneira como políticas de distribuição de renda estão associadas ao campo educacional (e têm consequências nesse campo), mas também, exibe como a análise dos dados revela muito de como eles são preenchidos, portanto, manuseados pelos operadores de políticas públicas.

A reflexão sobre os dados contidos em um banco, o modo como os pesquisadores precisaram lidar com desafios para tornar as informações capazes de elucidar a análise sobre a trajetória dos estudantes e o modo como se comportam em relação ao fluxo escolar, quando vinculados ao PBF, ajuda no aprofundamento das discussões metodológicas quantitativas, auxiliando outros pesquisadores que precisem lidar com bancos de dados secundários de vários tipos. Esta discussão vem em bom momento, tendo em vista a falta de tradição estatística do Brasil, relatada no texto, bem como a pouca presença de análises de cunho quantitativo nos campos da Educação e da Sociologia.

A abordagem aqui enfatizada primou pela atenção aos dados e a preocupação em excluir o mínimo possível os estudantes que compõem o banco, justamente por percebê-los não como “números”, mas como jovens dentro de uma situação de pobreza e exclusão. Por isso, o esforço demonstrado no texto em criar estratégias para qualificar as informações existentes e contornar os *missings*.

A reflexão sobre ausências (*missings*) em bancos de dados traz à tona a possibilidade de tais lacunas servirem como outro elemento de exclusão aos sujeitos mais pobres ou vulneráveis. Entendemos que o artigo traz uma contribuição interessante justamente ao lidar com uma discussão sobre estratégias de recuperação de dados ausentes e em como estes podem ser reveladores de determinadas situações importantes dentro do fenômeno estudado.

Ainda que haja clareza sobre os limites de análise de dados com caráter administrativo; que o grupo de pesquisa tenha verificado a existência de ausência de informações ou variação de informações sobre o mesmo estudante, que pode gerar questionamento sobre a validade interna dos dados; a intenção deste artigo foi apresentar estratégias possíveis para o tratamento dessas informações, que possuem grande potencial de explicação da realidade analisada. Considera-se que os limites apresentados não podem gerar uma situação de negação da validade dos dados, sobretudo porque, como procurou-se indicar neste artigo, as ausências de informação podem ser bastante eloquentes sobre as dinâmicas educacionais.

Referências

ALVES, M. R. G; SOARES, J. F. As pesquisas sobre o efeito das escolas: contribuições metodológicas para a Sociologia da Educação. *Sociedade e Estado*. Brasília, v. 22, n. 2. p. 435-474, maio/ago, 2007.

ANDRADE, J. M. de; LAROS, J. A. Fatores associados ao desempenho escolar: estudo multinível com dados do SAEB/2001. *Psicologia: teoria e pesquisa*. Brasília, v. 23, n. 1, p. 33-42, jan./mar 2007.

BICHIR, R. M. O Bolsa Família na berlinda? Os desafios atuais dos programas de transferência de renda. *Novos estud.* — CEBRAP, São Paulo, n. 87, p. 115-129, Jul. 2010. Disponível em: www.scielo.br. Acesso em: 12 fev. 2020.

BONFIM, S. R. de F. Política educacional e as condições materiais e estruturais das escolas municipais de Curitiba: um olhar a partir dos estudantes vinculados ao Programa Bolsa Família. Dissertação (*Mestrado em educação*). Universidade Federal do Paraná, 2018.

BOURDIEU, P. *A distinção: crítica social do julgamento*. Porto Alegre: Zouk, 2017.

BOURDIEU, P.; PASSERON, J. C. *A reprodução: elementos para uma teoria do sistema de ensino*. 7ªed. Petrópolis: Vozes, 2014.

BROOKE, N.; SOARES, J. F. (org). Seção 1 — Comentários. In: _____. *Pesquisa em eficácia escolar: origem e trajetórias*. Belo Horizonte: Editora da UFMG, 2008, p. 14-22.

Censo é adiado para 2021; coleta presencial de pesquisas é suspensa. *Agência IBGE notícias*, Rio de Janeiro, 17/03/2020. Disponível em: agenciadenoticias.ibge.gov.br. Acesso em: 21 mar. 2020.

COLLOMB, P. Les non-réponses aux questions d'opinion sur la politique des population. *Populations*, França, n 4-5, 1977. Disponível em: www.persee.fr. Acesso em: 10 abr. 2020.

CURRALERO, C. B. et al. As condicionalidades do programa bolsa família. In: CASTRO, J. A. de; MODESTO, L. (Org.). *Bolsa Família 2003-2010: avanços e desafios*. Brasília: IPEA, 2010. p. 151- 176. p. 101-132.

DUARTE, N. de S. Política social: um estudo sobre educação e pobreza. Tese (*Doutorado em Política Social*). Universidade de Brasília, Brasília, 2012.

DUARTE, N.; YANNOULAS, S. C. *O percurso escolar da população em situação de pobreza*. In: Anais da 34ª Reunião Anual da Associação Nacional de Pós-Graduação e Pesquisa em Educação — ANPED. Natal-RN, out. 2011. Disponível em: 34reuniao.anped.org.br. Acesso em: 21 maio 2020.

Extrema pobreza atinge 13,5 milhões de pessoas e chega ao maior nível em 7 anos. *Agência IBGE notícias*, Rio de Janeiro, 06/11/2019. Disponível em: agenciadenoticias.ibge.gov.br. Acesso em: 21 maio 2020.

FRANCO, C; BONAMINO, A. A pesquisa sobre característica de escolas eficazes no Brasil: Breve revisão dos principais achados e alguns problemas em aberto. *Revista Educação On-line*. Rio de Janeiro, n. 1, p. 1-13, 2005. Disponível em: <<http://lpp-uerj.net/olped/documentos/2081.pdf>>. Acesso em: 08 de nov. 2019.

GATTI, B. A. Estudos quantitativos em educação. *Educação e Pesquisa*, São Paulo, v. 30, n. 1, p. 11-30, jan./abr. 2004.

GAYA, T. F. M.; BRUEL, A. L; Estudos longitudinais em educação no Brasil: revisão de literatura da abordagem metodológica e utilização de dados educacionais para pesquisas em Educação. *Revista de Estudios Teóricos y Epistemológicos en Política Educativa*, Ponta Grossa, v. 4, p. 1-18, 2019. Disponível em: www.revistas2.uepg.br. Acesso em: 02 de maio 2020.

GIDDENS, A. *Sociologia*. Trad: Ronaldo Cataldo Costa; 6ª ed. Porto Alegre: Penso, 2012.

IBGE. *Memória — Censos Demográficos*. IBGE: Rio de Janeiro, s/d. Disponível em: memoria.ibge.gov.br. Acesso em: 20 abr. 2020.

LABORATÓRIO DE DADOS EDUCACIONAIS. *Indicador número de Matrículas*. Curitiba/Goiás: UFPR/UFG, 2020. Disponível em: dadoseducacionais.c3sl.ufpr.br. Acesso em: 20 abr. 2020.

LEÃO REGO, W.; PINZANI, A. *Vozes do Bolsa Família: autonomia, dinheiro e cidadania*. São Paulo: Editora Unesp, 2013.

MARTINS, B. A.; RUCKERT, F. Q. O Programa Bolsa Família e a educação: uma análise da produção científica brasileira fundamentada na base de dados SciELO (2003-2018). *Rev. Bras. Educ.*, Rio de Janeiro, v. 24, e240061, 2019. Disponível em: www.scielo.br. Acesso em: 10 fev. 2020.

MDS. *Condicionalidades*. Brasília: MDS, 27/07/2015. Disponível em: mds.gov.br. Acesso em 10 abr. 2019

MEC. *Manual de leiaute dos arquivos para troca de informações entre MEC e MDS*. Versão 4.0. Brasília: MEC, s/d.

MEC/SISTEMA PRESENÇA. *CONDEDU 2015*. Disponível em: frecuenciaescolarpbf.mec.gov.br. Acesso em: 12 fev. 2018.

MEC/SISTEMA PRESENÇA. *CONDEDU 2016*. Disponível em: frecuenciaescolarpbf.mec.gov.br. Acesso em: 12 fev. 2018.

MEC/SISTEMA PRESENÇA. *CONDEDU 2017*. Disponível em: frecuenciaescolarpbf.mec.gov.br. Acesso em: 12 fev. 2018.

MADAUS, G. F.; AIRASIAN, P. W.; KELLAGHAN, T. Insumos escolares, processos e recursos. In: BROOKE, N.; SOARES, J. F. (org). *Pesquisa em eficácia escolar: origem e trajetórias*. Belo Horizonte, editora da UFMG, 2008, p. 112-141.

PEREIRA, G.; ORTIGÃO, M. I. R. Pesquisa quantitativa em educação: algumas considerações. *Educação, Cultura e Comunicação*, Rio de Janeiro, v. 8, n. 1, p. 66-79, jan/jun. 2016.

PIRES, A. Afinal, para que servem as condicionalidades em educação do Programa Bolsa Família? *Ensaio: Aval. Pol. Públ. Educ.*, Rio de Janeiro, v. 21, n. 80, p. 513-532, jul./set. 2013.

PNUD/RDH (2019) *Relatório de Desenvolvimento Humano: Além do rendimento, além das médias, além do presente: Desigualdades no desenvolvimento humano no século XXI*. PNUD/ONU. 2019. Disponível em: www.pnud.org. Acessado em: 02/05/2020.

REZENDE, M. J. de. Uma abordagem histórico-hermenêutica dos relatórios do desenvolvimento humano (PNUD/ONU). e-l@tina. *Revista electrónica de estudios latinoamericanos*, Argentina, v. 13, n. 51, 2015. Disponível em: publicaciones.sociales.uba.ar. Acesso em: 20 maio 2020.

RUTTER, M. et al. Estudos anteriores. In: BROOKE, N.; SOARES, J. F. (org). *Pesquisa em eficácia escolar: origem e trajetórias*. Belo Horizonte, editora da UFMG, 2008, p. 142-152.

SEN, A. O desenvolvimento como expansão de capacidades. *Lua Nova Revista de Cultura e Política*, São Paulo, n. 28-29, p. 313-334, 1993.

_____. *Desenvolvimento como liberdade*. São Paulo: Cia das Letras, 2010.

SOARES, J. F.; ANDRADE, R. J. de. Nível socioeconômico, qualidade e equidade das escolas de Belo Horizonte. *Ensaio: avaliação de política pública*, Rio de Janeiro, v. 14, n. 50, p. 107-126, jan./mar. 2006.

SOARES, J. F.; COLLARES, A. C. M. Recursos familiares e desempenho cognitivo dos alunos do ensino básico brasileiro. *Revista de Ciências Sociais*, Rio de Janeiro, v. 49, n. 3, p. 615-650, 2006.

TEIXEIRA, A. A escola brasileira e a estabilidade social. *Revista brasileira de Estudos pedagógicos*, Brasília, vol. XXVIII, nº 67, 1957, p. 3-29.

UOL Economia. Página na internet disponível em: economia.uol.com.br. Acesso em: 24 maio 2020.

WORLD BANK. 2018. *Poverty and Shared Prosperity 2018: Piecing Together the Poverty Puzzle*. Washington, DC: World Bank. License: Creative Commons Attribution CC BY 3.0 IGO.

WORLD POVERTY CLOCK. Página na internet disponível em: <https://worldpoverty.io>. Acesso em: 20 maio 2020.

ZIMMERMANN, C. R. Os programas sociais sob a ótica dos direitos humanos: O caso do Bolsa Família do governo Lula no Brasil. *Sur. Rev. Int. Direitos Humanos*, São Paulo, nº. 4, Vol. 3, jun. 2006.

Resumo:

Dentro da temática das relações entre educação, desigualdades sociais e pobreza, a finalidade do artigo é apresentar o desenho e os desafios metodológicos de uma pesquisa em desenvolvimento sobre os percursos escolares de estudantes vinculados ao Programa Bolsa Família (PBF), nos estados do Ceará e do Paraná. A pesquisa centra-se em uma coorte de estudantes que nasceram em 2002 e que são acompanhados por um período de 3 anos (2015 a 2017). Trata-se de uma pesquisa de caráter quantitativo, com análise transversal e longitudinal, realizada de modo interinstitucional e interdisciplinar. Nesta etapa, ela se vale de um banco de dados criado a partir do Sistema de Acompanhamento da Frequência Escolar do Programa Bolsa Família (Sistema Presença). Ao cotejar elementos teóricos e metodológicos, introduz problemas e mobiliza estratégias possíveis para o tratamento das informações, que possuem potencial de explicação da realidade analisada.

Palavras-chave: Educação; pobreza; desigualdade.

Abstract:

Within the theme of the relationship between education, social inequalities and poverty, the purpose of the article is to present the design and methodological challenges of a research in progress on the school pathways of students linked to the Bolsa Família Program (BFP), in the states of Ceará and Paraná. The research focuses on a cohort of students who were born in 2002 and who are followed for a period of 3 years (2015 to 2017). It is a quantitative research, with transversal and longitudinal analysis, carried out in an interinstitutional and interdisciplinary way. In this step, it uses a database created from the School Attendance Monitoring System of the Bolsa Família Program (Presence System). By comparing theoretical and methodological elements, it introduces problems and mobilizes possible strategies for the treatment of information, which have the potential to explain the analyzed reality.

Keywords: Education; poverty; inequalities.

Recebido para publicação em 31/05/2020.

Aceito em 11/11/2020.