

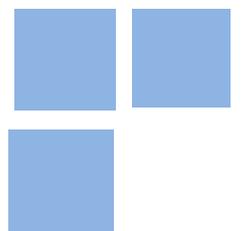
Programa de Ensino Integral e desempenho Escolar em São Paulo

YURI PASSUELO

VENEZIANO C. ARAÚJO

ELOIZA R. F. DE ALMEIDA

SOLANGE GONÇALVES



Full-time Education Program and School Performance in São Paulo

Yuri Passuelo (yuri.mpassuelo@gmail.com)

Veneziano C. Araújo (veneziano.araujo@unifesp.br)

Eloiza R. F. de Almeida (eloiza.almeida@usp.br)

Solange Gonçalves (solange.goncalves@usp.br)

Abstract:

The full-time school programs have been one of the main public policies used by the different Brazilian administrative spheres in an attempt to improve educational outcome. This work evaluates the effect of the full-time school program PEI in the public schools of the state of São Paulo, for the Portuguese and mathematics exams of the Prova Brasil exam between the years 2013 and 2017, for students of the last year of elementary and junior high school, through Average Treatment Effect on Treated and Propensity Score Matching. The main results indicate that the program obtain, on average, higher levels of proficiency both in Portuguese and in Mathematics both in the 5th and 9th years of elementary school. In addition, greater proficiency levels are observed in Mathematics than in Portuguese.

Keywords: Full-time school, public policy, Prova Brasil exam, propensity score matching

JEL Codes: I28, H52, C21

Programa de Ensino Integral e desempenho Escolar em São Paulo

Resumo:

Os programas de escola em tempo integral tem sido uma das principais políticas públicas visando melhorar o desempenho educacional dos alunos do Ensino Regular. Este artigo avalia o efeito do programa de escola integral (PEI) nas escolas públicas do Estado de São Paulo, para os exames de Língua Portuguesa e Matemática da Prova Brasil entre os anos de 2015 e 2017, para os alunos dos últimos anos do Ensino Fundamental ciclo I e ciclo II, por meio do método empírico de *Average Treatment Effect on Treated e Propensity Score Matching*. Os principais resultados apontam que os alunos participantes do programa obtêm em média maiores níveis de proficiência tanto em Língua Portuguesa quanto em Matemática tanto nos 5^{os} quanto nos 9^{os} anos do Ensino Fundamental. Além disso, são observados maiores níveis de proficiência em Matemática do que em Língua Portuguesa.

Palavras-Chave: Escola em tempo Integral, Política Pública, Prova Brasil, Propensity Score Matching.

Programa de Ensino Integral e desempenho Escolar em São Paulo

Yuri Passuelo¹
Veneziano C. Araújo²
Eloiza R. F Almeida³
Solange Gonçalves⁴

RESUMO

Os programas de escola em tempo integral tem sido uma das principais políticas públicas visando melhorar o desempenho educacional dos alunos do Ensino Regular. Este artigo avalia o efeito do programa de escola integral (PEI) nas escolas públicas do Estado de São Paulo, para os exames de Língua Portuguesa e Matemática da Prova Brasil entre os anos de 2015 e 2017, para os alunos dos últimos anos do Ensino Fundamental ciclo I e ciclo II, por meio do método empírico de *Average Treatment Effect on Treated e Propensity Score Matching*. Os principais resultados apontam que os alunos participantes do programa obtêm em média maiores níveis de proficiência tanto em Língua Portuguesa quanto em Matemática tanto nos 5^{os} quanto nos 9^{os} anos do Ensino Fundamental. Além disso, são observados maiores níveis de proficiência em Matemática do que em Língua Portuguesa.

Palavras-chave: Escola em tempo Integral. Política Pública. Prova Brasil. Propensity Score Matching.

ABSTRACT

The full-time school programs have been one of the main public policies used by the different Brazilian administrative spheres in an attempt to improve educational outcome. This work evaluates the effect of the full-time school program PEI in the public schools of the state of São Paulo, for the Portuguese and mathematics exams of the Prova Brazil exam between the years 2013 and 2017, for students of the last year of elementary and junior high school, through *Average Treatment Effect on Treated and Propensity Score Matching*. The main results indicate that the program obtain, on average, higher levels of proficiency both in Portuguese and in Mathematics both in the 5th and 9th years of elementary school. In addition, greater proficiency levels are observed in Mathematics than in Portuguese.

Keywords: Full Time School. Public Policy. Prova Brasil exam. Propensity Score Matching

JEL: I28, H52, C21.

¹ Graduado em Ciências Econômicas pela Universidade Federal de São Paulo (UNIFESP). E-mail: yuri.mpassuelo@gmail.com

² Professor do Departamento de Economia da Universidade Federal de São Paulo (UNIFESP). E-mail: veneziano.araujo@unifesp.br

³ Doutoranda em Teoria Econômica na Universidade de São Paulo (FEA-USP). E-mail: eloiza.almeida@usp.br

⁴ Professora do Departamento de Economia da Universidade de São Paulo (FEA-USP). E-mail: solange.goncalves@usp.br

1. INTRODUÇÃO

Mesmo depois da superação do desafio de universalização do acesso à educação nos anos 1980 e 1990, a educação brasileira ainda figura entre as últimas posições em exames de avaliação internacional como o PISA. Para os exames realizados em 2018, o país ficou na 72º posição no exame de Matemática e em 67º em Ciências, entre 79 países. Já em leitura, o país conseguiu alcançar a 59º posição, dentre 78 países participantes.

Diante desse contexto de baixo desempenho em avaliações internacionais, uma das principais políticas educacionais adotadas é a das escolas de tempo integral, que consiste na extensão do período de permanência dos estudantes nas escolas. Existem diversas variações nos planos pedagógicos desse tipo de intervenção. Como é o caso do programa de escola em tempo integral de Pernambuco avaliado por Rosa et al. (2020) e o Programa Mais Educação (PME) do Governo Federal, avaliado por Almeida et al. (2015).

O Estado de São Paulo apresenta dois diferentes programas de escola de tempo integral. O Programa de Escola Integral (PEI), foco do presente trabalho, é o programa mais recente, implementado a partir de 2012 como um novo programa de escola em tempo integral, que visa expandir o tempo de permanência e instrução do estudante na escola promovendo atividades como aulas de reforço, acompanhamento estudantil e orientação do estudante na construção de projetos e atividades relacionadas a formação do aluno como indivíduo. Já o programa Escola em Tempo Integral (ETI), que foi implementado em meados de 2006 (logo, anterior ao PEI), com o objetivo de promover atividades esportivas e culturais como complemento às atividades em sala de aula. Este programa sofreu uma série de mudanças durante a sua trajetória e ainda está vigente e sua existência será considerada na análise.

Nesse contexto, o objetivo deste artigo é avaliar o programa de escola em tempo integral PEI para as séries do 5º e 9º ano do Ensino Fundamental, no período entre os anos de 2015 a 2017. O intuito é investigar se o programa obteve efeito sobre o desempenho acadêmico dos alunos participantes em comparação com alunos de mesmas características da rede pública estadual, porém que estudaram em escolas que não participaram de nenhum outro programa. As comparações são feitas apenas para o ano de 2017 no caso dos 5ºs anos, enquanto dos 9ºs anos são feitas para os anos de 2015 e 2017.

Este artigo está organizado em cinco seções além desta introdução. Na seção 2, é apresentada a revisão da literatura sobre a análise de impactos de programas com aumento

de tempo de instrução no desempenho escolar. A seção 3 expõe brevemente o funcionamento e estrutura do PEI. Na seção 4, são detalhadas as fontes de dados utilizadas no trabalho, a seleção de variáveis para a especificação econométrica e as estratégias de identificação aplicadas no estudo. Já a seção 5, apresenta os resultados principais do trabalho, além dos exercícios de robustez. Por fim, a seção 6 contempla as considerações e discussões finais do artigo.

2. TEMPO DE INSTRUÇÃO E DESEMPENHO ESCOLAR

Tanto a literatura nacional quanto a internacional buscam avaliar o quanto o tempo de instrução impacta o aprendizado do aluno, artigos como Lavy (2015) e Rivkin e Schuman (2015) buscam entender como variações no tempo de instrução entre países tem impactos sobre o desempenho nos exames do PISA de 2006 e 2009 respectivamente, os resultados apontam em ambos que existe uma associação positiva, porém Rivkin e Schuman (2015) destacam o papel condicional que variáveis como ambiente escolar e qualidade da aula ministrada tem sobre o efeito do tempo de instrução. Já Cattaneo et al. (2017) e Battistin e Meroni (2016) aprofundam análises dentro de países como Suíça e Itália e chegam a resultados divergentes, enquanto Cattaneo et al. (2017) encontra uma associação positiva, porém menor do que os efeitos apontados por Lavy (2015) e Rivkin e Schuman (2015) para o caso da Suíça, Battistin e Meroni (2016) não encontra associação positiva a direta entre desempenho escolar e tempo de instrução para o programa analisado na Itália, sendo essa apenas presente no conjunto de estudantes com pior desempenho. Diante essa variedade de resultados que mostram diferentes efeitos e impactos o trabalho Aronson et al. (1999) investiga as diferentes manifestações que o tempo de permanência do estudante na escola é alocado, e classifica essas manifestações em quatro grupos sendo elas tempo alocado, tempo de instrução, tempo não instrucional e tempo de aprendizagem acadêmica, sendo o primeiro o tempo em que o estudante permanece na escola não necessariamente em aula, o segundo e o terceiro o tempo em aula e o tempo fora de aula e por último o tempo de aprendizado

Na literatura nacional são diversos os artigos que buscam entender o impacto do tempo de instrução e avaliar programas de escola em tempo integral, como é o caso deste artigo. Menezes Filho (2007) investigou os principais determinantes do desempenho escolar para o SAEB de 2005 e Vasconcellos e Hermeto (2015) analisam o impacto do programa de

escola integral de Belo Horizonte sob o desempenho no exame municipal por meio de uma regressão DID entre os anos de 2007 e 2011 e chegam a resultados que associam positivamente o impacto aos resultados do exame. Rosa et al. (2020) analisa o impacto do programa de escola em tempo integral de Pernambuco sobre o desempenho no exame de avaliação estadual entre 2010 e 2016 e chega também a resultados positivos nos exames de Língua Portuguesa e Matemática.

Almeida et al. (2016) e Oliveira e Terra (2018) analisam o programa mais educação (PME) sob óticas e períodos diferentes, enquanto o primeiro trabalho faz uma avaliação geral no período de 2008 a 2011 sob indicadores como resultados da Prova Brasil e taxas de abandono, o segundo trabalho busca avaliar se a mudança a partir de 2012 na priorização de escolas com maior presença de alunos beneficiário do bolsa família trouxe resultados positivos nas escolas selecionadas sob indicadores como resultados da Prova Brasil e taxas de aprovação reprovação e abandono entre os anos de 2010 a 2013, ambos os trabalhos trouxeram resultados semelhantes aos de onde o programa não teve impactos diretos sobre os indicadores analisados.

3. O PROGRAMA DE ESCOLA INTEGRAL (PEI)

O PEI foi implementado nas escolas estaduais de São Paulo no ano de 2012 como um novo modelo de programa de ensino integral, o foco inicial do programa foi atender as escolas do ensino médio, e progressivamente foi implementado no ensino fundamental.

O desenho do programa engloba a carga curricular comum aos alunos da rede estadual e acrescenta um conjunto de atividades complementares como aulas de Língua inglesa, aulas práticas de Informática e Ciências (Física, Química e Biologia), horário dedicado à orientação de estudo, e atividades listadas como construção de um projeto de vida e a de participação clubes juvenis.

Além da mudança na matriz curricular existem mudanças quando necessário, na estrutura física da escola e no corpo pedagógico. Os professores que lecionam nas unidades atendidas pelos programas são colocados sobre um regime de trabalho diferente do resto da rede Estadual, o regime de dedicação plena e integral (RDPI), tem como principal diferença um salário 75% maior em relação a faixa salarial base do professor, e dedicação exclusiva por um período de 40 horas semanais na unidade escolar. O processo de

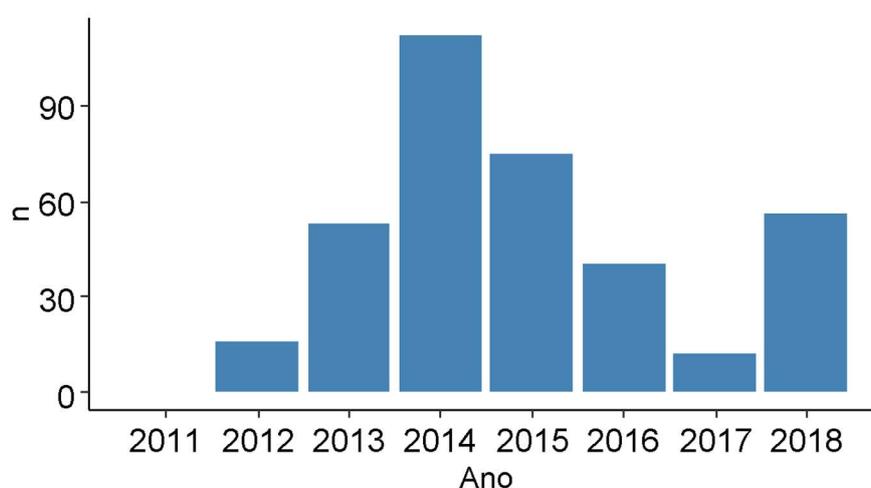
contratação também se dá de forma diferente, exigindo que o candidato seja professor da rede estadual há pelo menos três anos e tenha um curso de licenciatura relacionado à matriz curricular. Quanto à gestão escolar, também há requisito de experiência de, no mínimo, três anos como diretor, e, caso também atue como docente, são necessários oito anos de experiência como docente, além de possuir curso superior em Pedagogia ou pós-graduação na área da educação.

De acordo com a SEESP (2014), a equipe escolar como um todo - quadro de docentes e equipe supervisora do projeto pedagógico e implementação do programa - é submetida a avaliações de desempenho. As competências avaliadas englobam itens como protagonismo, didática e domínio do conhecimento, formação contínua do profissional e autodesenvolvimento, gestão, relacionamento e corresponsabilidade, criatividade e solução de problemas e, por último, difusão e multiplicação de boas práticas de ensino.

A escola escolhida para receber o programa pode inicialmente manter sua estrutura docente e pedagógica, desde que a equipe se enquadre sob o regime de dedicação do programa e seja colocada sob um período de adaptação. Ao fim desse período, a gestão será avaliada sob as métricas dispostas pelo programa.

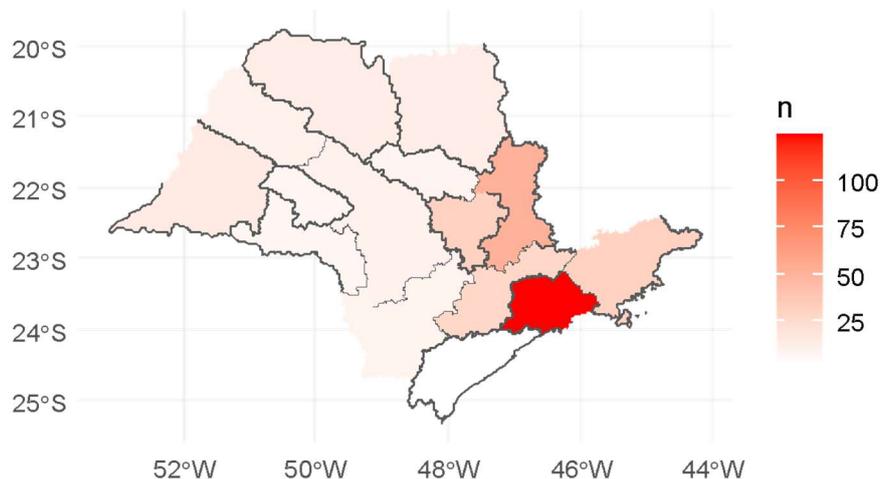
O programa também prevê que as unidades escolares que recebem a implementação do programa, tenham um sistema de priorização de alunos já matriculados na unidade escolar com o intuito de evitar a substituição de alunos de outras unidades da rede estadual ou do ensino privado.

Figura 1 – Ingresso de escolas no PEI por ano do programa



Fonte: Elaboração própria a partir dos dados da SEESP (2020)

Figura 2 – Distribuição geográfica do número de escolas inscritas no PEI por Mesorregião



Fonte: Elaboração própria a partir dos dados da SEESP (2020).

Por meio da Figura 1, é possível observar que o início do programa ocorreu em 2012 com a entrada de 16 escolas. Uma aceleração no número de escolas inscritas ocorre nos dois anos subsequentes, seguida de uma redução entre o quarto e quinto ano do programa (2015-2017) e um novo aumento em 2018. A distribuição geográfica das escolas participantes é apresentada na Figura 2, em que quanto mais escura a cor da mesorregião, maior a incidência da implementação do programa. É possível verificar que a incidência do PEI está concentrada na mesorregião de São Paulo Capital e nas demais unidades próximas à Capital, como Campinas, Piracicaba, Região Macro Paulista e Vale do Paraíba.

Tabela 1 – ATT especificação principal

Variável	Ensino Regular	PEI	Teste t	
<i>Características da escola</i>				
Número de Escolas	4.888	364		
Turmas	102.557	3.782		
Alunos	3.147.632	117.723		
Alunos/Escola	(643,95)	(323,41)	-40,6066	***
	374,1832	110,7026		
Alunos/Turma	30,69	31,13	1,3267	
	(6,6015)	(6,0655)		
Turmas/Escola	20,9814	10,3901	-48,5918	***
	(10,4656)	(3,0227)		
<i>Infraestrutura da Escola</i>				
Salas	14,9033	14,2487	-2,4831	**
	(5,4288)	(4,8064)		
Presença de Biblioteca/sala de leitura	0,8815	0,9849	13,1184	***
	(0,3232)	(0,1218)		
Presença de Cozinha	0,9847	0,9944	2,2706	**
	(0,1227)	(0,07431)		
Presença de Refeitório	0,5745	0,8356	9,9052	***
	(0,4994)	(0,4841)		
Presença de Laboratório de Ciências	0,1391	0,4637	12,2013	***
	(0,3461)	(0,4987)		
Presença de Laboratório de Informática	0,9167	0,9614	4,1230	***
	(0,2764)	(0,1926)		
Presença de Quadra Esportes	0,9374	0,991	8,8740	***
	(0,2422)	(0,0944)		
Presença de Computador	0,9898	0,9929	0,6678	
	(0,1006)	(0,0842)		
Presença de Internet	0,9879	0,9897	0,3260	
	(0,1095)	(0,101)		

Fonte: Elaboração própria a partir dos dados do Censo Escolar (2018).

Notas: A tabela inclui apenas as escolas estaduais em funcionamento e de Ensino Regular. Os níveis de significância são * $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$.

Fora as diferenças na estrutura do corpo docente e pedagógico mencionadas, as escolas que atendem ao programa também apresentam diferenças relacionadas à estrutura física e características da escola quando comparada ao restante da rede estadual, a Tabela 1 mostra a comparação entre a estrutura das escolas participantes do programa e o as escolas de ensino regular, por dela é possível perceber que as escolas participantes do programa possuem um menor número médio de turmas por escola apesar do tamanho dessas turmas

permanecer praticamente o mesmo, além disso possuem maior presença de bibliotecas ou salas de leitura, refeitório, laboratórios de ciências e informática e quadras de esporte.

4. BASE DE DADOS E ESTRATÉGIA EMPÍRICA

Para este artigo são três as fontes de dados utilizadas para a análise. A primeira delas se trata da Prova Brasil (PB) no INEP, os dados utilizados dessa fonte se resumem as proficiências de Língua Portuguesa e Matemática, calculadas na escala SAEB⁵, e as informações do questionário socioeconômico, fontes de informações a nível aluno que descrevem características da residência, *background* escolar e *background* familiar. Além da Prova Brasil, o Censo Escolar também do INEP é utilizado como fonte de informações a nível escolar, listando uma série de características relacionadas à estrutura da escola, essa informação é pareada com a PB pelo código identificador da escola. A última fonte utilizada é a relação de escolas participantes do programa, e seu respectivo ano de entrada, disponibilizada pela SEESP, sendo essa informação da mesma forma que o Censo Escolar, pareada pelo código identificador da escola.

Para a estratégia empírica de avaliação do efeito das escolas de tempo integral é utilizada a metodologia do *Average Treatment Effect on Treated* (ATT) ou efeito médio de tratamento nos tratados. Seguindo a abordagem de resultados potenciais proposta por Rubin (1974), o ATT será a diferença média no desempenho da PB (y) de um aluno i quando esse recebeu o tratamento ($D=1$) comparado ao cenário em que não recebeu o tratamento ($D=0$). Como tratado, considera-se que o aluno está em uma escola com um dos programas de Ensino Integral. Podemos representar esse efeito como:

$$ATT = E(y_i|D = 1) - E(y_i|D = 0)$$

Para a estimação do ATT três metodologias são adotadas, o *Propensity Score Matching* (PSM), a distância de *Mahalanobis* e o balanceamento por entropia, ambas têm por objetivo selecionar um grupo de estudantes do grupo controle com características mais semelhantes aos estudantes do grupo sob tratamento, porém existem diferenças técnicas entre elas.

A aplicação do PSM é dada em duas etapas. Primeiramente calcula-se o *propensity score* a partir das características observáveis dos indivíduos, utilizando-se de metodologias de

⁵ Escala SAEB com média zero e desvio padrão 1

classificação binária e, na sequência, realiza-se o pareamento dos indivíduos (*matching*), a partir do escore de propensão, utilizando modelos de classificação que permitam identificar como controles os indivíduos mais parecidos quanto possível para cada tratado. Dessa forma, é possível obter o ATT comparando o desempenho dos indivíduos tratados e controles.

De acordo com Rosenbaum e Rubin (1983), o escore de propensão seria a probabilidade condicional de se participar de um tratamento específico dado um conjunto de variáveis observáveis (X), e dessa forma os resultados potenciais são independentes da atribuição do tratamento sendo apenas baseada em características observáveis. Essa premissa é conhecida como independência condicional e é dada por:

$$Y_0 \perp D | X \quad (2)$$

Uma segunda condição garante que para cada indivíduo tratado existe um par no grupo de controle cujo resultado reproduz o que seria o resultado do indivíduo tratado na ausência do tratamento. Ou seja, a hipótese de suporte comum ou de sobreposição, garante que as características dos tratados sejam representadas no grupo controle e vice-versa. Formalmente, esta hipótese é dada por:

$$0 < P(D = 1 | X) < 1. \quad (3)$$

As características observáveis contidas no vetor X são inúmeras e compará-las individualmente torna o pareamento inviável (problema de dimensionalidade). A utilização do score de propensão proposto por Rosenbaum e Rubin (1983) visa eliminar este problema tornando a comparação unidimensional, utilizando uma função de X que resume toda a informação contida neste vetor, no chamado *propensity score*:

$$p(X) = P(D = 1 | X) \quad (4)$$

Considerando que o resultado potencial é independente da atribuição do tratamento, mas condicionada às características observáveis, tal premissa se aplica ao *propensity score*, podendo ser então reescrita para:

$$Y_0, Y_1 \perp D | X \quad (5)$$

Dessa forma, sob as hipóteses de identificação, dado o *propensity score*, é possível criar um grupo de controle com características observáveis semelhantes às do grupo de tratamento e eliminar o viés da estimativa do ATT. Para o cálculo do *propensity score* a

metodologia utilizada é por meio do modelo *Logit*, e como metodologia para a etapa de *matching* é utilizado um KNN com $k = 1$ com reposição.

A distância de *Mahalanobis*, diferentemente do PSM, não reduz as dimensões por meio da aplicação de um modelo multivariado, mas sim calcula as distâncias das observações sob tratamento para as observações do grupo controle e por último realiza o *Matching* escolhendo as observações controle mais próximas, de maneira formal a distância entre dois pontos no R^n é calculada pela fórmula abaixo:

$$d(\vec{x}, \vec{y}) = \sqrt{(\vec{x} - \vec{y})^T S^{-1} (\vec{x} - \vec{y})} \quad (6)$$

Em que x e y são os respectivos pontos relacionados às observações tratamento e controle e S^{-1} é a matriz de covariância do conjunto de variáveis.

O balanceamento por entropia se baseia de forma semelhante ao *Propensed Score Weighting* (PSW), onde o objetivo é estimar $E(D = 0)$ utilizando todo conjunto das observações pertencentes ao grupo controle, de forma que Y_i é ponderado por um conjunto de pesos d_i :

$$E(D = 1) = \frac{\sum_{i|D=0} Y_i d_i}{d_i} \quad (7)$$

Porém diferentemente do PSW, no balanceamento por entropia os pesos d_i não são calculados por meio de um *Propensed Score*, mas sim estimados por meio de um processo de otimização. A otimização consiste em minimizar a função $h(w_i) = w_i \log\left(\frac{w_i}{q_i}\right)$, em função de w_i , aonde q_i é um conjunto de pesos bases para cada observação definidos como $1/n_0$. As restrições para a otimização são dadas por:

$$\sum_{i|D=0} w_i = 1 \quad (8)$$

$$w_i \geq 0 \quad (9)$$

$$\sum_{i|D=0} w_i c_{ri}(X_i) = m_r \quad (10)$$

Ou seja, o conjunto de pesos w_i deve ser maior ou igual a 0 e a soma dos pesos deve ser igual a 1. Além disso outra restrição imposta a otimização é que os pesos w_i diminuam a distância entre os r momentos de cada variável do grupo controle para o grupo tratamento, na qual $r \in 1, \dots, R$.

A principal motivação para a aplicação de metodologias como a distância de *Mahalanobis* e o balanceamento por entropia além do próprio PSM é a fragilidade apontada por King

& Nielsen (2019), em contextos de amostras já balanceadas o PSM pode acabar produzindo amostras desbalanceadas e consequentemente em resultados subótimos em comparação com metodologias de Blocagem Completa, como é o caso da distância de *Mahalanobis*, já a metodologia de balanceamento por entropia é sugerida por Hainmuller (2017) como uma alternativa para produzir amostras balanceadas no contexto de estudos observacionais.

A seleção de variáveis utilizadas para o pareamento entre o grupo controle e o grupo de tratamento segue a lógica de trabalhos como Rocha et al. (2018), Costa (2017), Dutra (2016) e Xu et al. (2017), selecionando variáveis relacionadas a características do domicílio, escolaridade dos pais, histórico escolar (reprovação e abandono) e características de infraestrutura das escolas. Devido a existência de uma variedade de variáveis relacionadas a características do domicílio (oito ao todo) foi aplicado a Análise dos Componentes Principais (ACP)⁶⁷ para esse conjunto com o intuito de reduzir o número de dimensões e, portanto, potenciais problemas de multicolinearidade, para a especificação principal apenas o primeiro componente desse resultado será utilizado.

Para além da aplicação dessas três metodologias, quatro exercícios de robustez também são aplicados com o intuito de averiguar a estabilidade dos resultados segundo diferentes especificações, em todos os exercícios foram utilizadas variações do PSM, o primeiro deles se trata da utilização de quatro componentes de características do domicílio ao invés de apenas um, no segundo exercício são retirados o conjunto de variáveis a nível escola, no terceiro é utilizada uma variação na especificação de *matching* utilizando um KNN com $k = 1$ sem reposição na amostra, e por último é utilizado um KNN com $k = 3$ com reposição.

5. RESULTADOS

A Tabela A1 do apêndice mostra os resultados dos modelos *logit* utilizados no cálculo do *propensity score*, e as figuras A1 e A2 também do apêndice mostram a distribuição do *propensity score* antes e depois do pareamento, inicialmente, os estudantes do grupo de

⁶ Para mais detalhes ver em Hongyu et al. (2015).

⁷ Os resultados do ACP estão disponíveis ao público por meio de solicitação.

tratamento e de controle apresentam diferenças nas distribuições do *propensity score*, mas a aplicação *Matching*, resulta em grupos com balanceamento semelhante.

Tabela 1 – ATT especificação principal

Disciplina	Modelo	5º Ano		9º Ano	
		2017	2015	2017	2017
Língua Portuguesa	PSM	0,2035 ***	0,1195 ***	0,1567 ***	***
	Mahalanobis	0,1712 ***	0,1298 ***	0,1298 ***	***
	Entropia	0,1996 ***	0,1621 ***	0,2111 ***	***
Matemática	PSM	0,2641 ***	0,1479 ***	0,2253 ***	***
	Mahalanobis	0,2157 ***	0,1565 ***	0,2537 ***	***
	Entropia	0,2512 ***	0,1968 ***	0,2705 ***	***

Fonte: Dados da Prova Brasil, Censo Escolar e da SEESP e elaboração própria. Notas: Os níveis de significância são * $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$. O número de observações controle pareados para cada respectivo exercício foi de 1.317, 1.326 e 105.548 para o caso dos 5ºs anos em 2017. Para o caso dos 9ºs anos em 2015 foram 4.094, 4.131 e 297.656 observações. E para os 9ºs anos em 2017 5.331, 5.390, 152.617 observações para os três exercícios.

A Tabela 1 mostra os resultados da aplicação do ATT para as disciplinas de Língua Portuguesa e Matemática em ambas as séries utilizando as três metodologias. Os resultados indicam forte associação positiva e significativa do programa em todos os casos apresentados. Para o 5º ano o efeito estimado em Língua Portuguesa é de 0,17 até 0,20 pontos de score, em matemática o efeito é maior na faixa dos 0,21 a 0,26 pontos.

Para o 9º ano em 2015 o efeito mensurado em Língua Portuguesa ficou na faixa de 0,12 a 0,16 pontos, em Matemática o mesmo também foi um pouco maior na faixa de 0,15 a 0,19. Para 2017 esse efeito foi maior nas duas disciplinas, sendo em Língua portuguesa entre 0,16 e 0,21 e em Matemática entre 0,22 e 0,27.

Tabela 2 – ATT proficiências de Língua Portuguesa para os exercícios de Robustez

Disciplina	Modelo	5º Ano		9º Ano	
		2017		2015	2017
Especificação					
Língua Portuguesa	Exercício 1	0,2154 ***		0,1443 ***	0,1904 ***
	Exercício 2	0,0953 ***		0,0583 **	0,2291 ***
Matemática	Exercício 1	0,2843 ***		0,1569 ***	0,2427 ***
	Exercício 2	0,0941 ***		0,0896 ***	0,2643 ***
Mudanças de Parâmetros					
Língua Portuguesa	Exercício 3	0,1938 ***		0,1304 ***	0,1899 ***
	Exercício 4	0,1873 ***		0,1277 ***	0,1891 ***
Matemática	Exercício 3	0,1899 ***		0,1651 ***	0,2434 ***
	Exercício 4	0,2471 ***		0,1734 ***	0,2441 ***

Fonte: Dados da Prova Brasil, Censo Escolar e da SEESP e elaboração própria.

A Tabela 2 apresenta os resultados do ATT para os quatro exercícios de robustez realizados. Os resultados indicam na mesma forma que o exercício principal, uma relação positiva da implementação do programa com as proficiências de Língua Portuguesa e Matemática da PB para ambos as séries e anos de análise, porém com algumas diferenças de intensidade, para o 5º ano o resultado em 2017 teve uma faixa de variação do ATT de 0,09 a 0,28 pontos de *score* para Língua Portuguesa e 0,18 a 0,24 para Matemática, no caso dos 9ºs anos em 2015, esse intervalo ficou na faixa 0,05 a 0,15 para Língua Portuguesa e 0,12 a 0,17 para Matemática, por último em 2017 a faixa ficou entre 0,19 e 0,26 para Língua Portuguesa e 0,18 a 0,24 para Matemática.

Os resultados dos testes apresentam robustez as diferentes especificações sempre resultando em associação positiva do programa com melhores resultados, apenas um exercício de robustez provocou maiores variações, sendo o exercício de retirada das variáveis a nível escola, para este exercício os efeitos estimados foram menores para as notas de Língua Portuguesa em ambas séries para o ano de 2015.

Um resumo dos resultados principais e dos exercícios de robustez podem ser visualizados nas Figuras A3 e A4 do apêndice que apresentam o ATT para cada exercício e os respectivos intervalos de confiança.

6. CONSIDERAÇÕES FINAIS

Neste artigo foi avaliado o programa de escola em tempo integral (PEI) pertencente à rede pública estadual de educação de São Paulo, entre os anos de 2013 e 2017. Para a estimação dos efeitos de cada programa foi utilizado o método de ATT, usando os métodos PSM, distância de *Mahalanobis* e balanceamento por entropia para o balanceamento das amostras de controle e tratamento.

Os resultados da especificação principal e dos testes de robustez realizados mostram que existem evidências de uma associação positiva entre a implementação do programa e um melhor desempenho escolar dos alunos nas provas de Língua Portuguesa e Matemática da PB.

A associação entre a implementação do programa e o desempenho é mais evidente nas provas de Língua Portuguesa e Matemática, tanto nos 5^{os} quanto nos 9^{os} anos. Para os alunos dos 9^{os} anos há uma tendência de aumento do gap de desempenho observado entre o programa e as escolas que não participam de nenhum programa.

No geral, os exercícios de robustez indicam conclusões muito semelhantes à especificação principal. Dois exercícios chamaram atenção por serem mais sensíveis à mudança de especificação, mas, em geral, esses exercícios se mostraram compatíveis com os resultados da especificação principal, seja mudando o conjunto de variáveis para a construção do *Propensity Score*, utilizando variações do *Matching*, mudando os parâmetros do KNN ou mudando a metodologia para blocagem completa.

7. REFERÊNCIAS

LAVY, Victor. Do Differences in Schools' Instruction Time Explain International Achievement Gaps? Evidence from Developed and Developing Countries. *Economic Journal*, Royal Economic Society, [S. l.], ano 2015, v. 125, n. 588, p. 397-424, 17 nov. 2015.

RIVKIN, Steven. G; SCHIMAN, Jeffrey C. Instruction Time, Classroom Quality, and Academic Achievement. *The Economic Journal*, [S. l.], ano 2015, v. 125, n. 588, p. F425-F448, 17 nov. 2015.

CATTANEO, Mario A.; OGGENFUS, Chantal; WOLTER, Steffan C. The more, the better? The impact of instructional time on student performance. *Education Economics*, [S. l.], ano 2017, v. 25, n. 5, p. 433-445, 13 abr. 2017.

BATTISTIN, Erich; MERONI, Elena Claudia. Should we increase instruction time in low achieving schools? Evidence from Southern Italy. *Economics of Education Review*, [s. l.], v. 55, p. 39-56, dez. 2016.

ARONSON, Julie; ZIMMERMAN, Joy; CARLOS, Lisa. Improving Student Achievement by Extending School: Is It Just a Matter of Time?. *WestEd*, [s. l.], 30 abr. 1998.

MENEZES FILHO, Naercio Aquino. Os determinantes do desempenho escolar do Brasil. In: *O Brasil e a ciência econômica em debate*, [S.l: s.n.], v. 1. 2012.

HERMETO, Ana Maria; TEIXEIRA, Clarissa Gondim; VASCONCELLOS, Ligia Maria de. Relatório de Avaliação Econômica: Programa Escola de Tempo Integral Secretaria de Educação e Cultura do Estado do Tocantins. *Avaliação Econômica de Projetos Sociais: Itau Social*, [s. l.], 2017.

ROSA, Leonardo; BETTINGER, Eric; CARNOY, Martin; DANTAS, Pedro. The effects of public high school subsidies on student test scores: The case of a full-day high school program in Pernambuco, Brazil. *Working Paper*, [s. l.], 2020.

ALMEIDA, Rita; BRESOLIN, Antonio; BORGES, Bruna; MENDES, Karen; MENEZES, Naercio. Assessing the Impacts of Mais Educacao on Educational Outcomes: Evidence between 2007 and 2011. *Policy Research Working Paper World Bank*, Washington, DC, ano 2016, abr. 2016.

OLIVEIRA, Luis Felipe Batista de; TERRA, Rafael. Impacto do Programa Mais Educação em indicadores educacionais. *International Policy Centre for Inclusive Growth.*, [s. l.], v. 147, 2016.

GOVERNO DO ESTADO (São Paulo). Secretaria de Educação do Estado de São Paulo (SEESP). Resolução SE-52, de 2-10-2014. [S. l.], 2 out. 2014.

GOVERNO DO ESTADO (São Paulo). Secretaria de Educação do Estado de São Paulo (SEESP). Resolução SE-68, de 17-12-2014. [S. l.], 17 dez. 2014.

Rubin, D. B. (1974). Estimating causal effects of treatments in randomized and nonrandomized studies. *Journal of Educational Psychology*, 66(5), 688–701. <https://doi.org/10.1037/h0037350>.

ROSENBAUN, Paul R.; RUBIN, Donald B. The central role of the propensity score in observational studies for causal effects. *Biometrika*, [s. l.], ano 1983, v. 70, ed. 1, p. 41-55, 1983.

KING, Gary; NIELSEN, Richard. Why Propensity Scores Should Not Be Used for Matching. *Political Analysis*, [s. l.], v. 27, n. 4, p. 435 - 454, 2019.

HAINMUELLER, Jens. Entropy Balancing for Causal Effects: A Multivariate Reweighting Method to Produce Balanced Samples in Observational Studies. *Political Analysis*, [s. l.], v. 20, p. 25-46, 2017.

HONGYU, Kuang; SANDANIELO, Vera Lúcia Martins; OLIVEIRA, Gilmar Jorge. Análise de Componentes Principais: resumo teórico, aplicação e interpretação. *Engineering and Science: Scientific Journal of FAET and ICET UFMT*, [s. l.], v. 5, ed. 1, p. 83-90, 2015.

Rocha, Menezes e Komatsu (2018) - ROCHA, R. H.; MENEZES-FILHO, N.; KOMATSU, B. K. Avaliando o impacto das políticas educacionais em Sobral. *Economia Aplicada*, [S. l.], v. 22, n. 4, p. 5-30, 2018. DOI: 10.11606/1980-5330/ea110189

KOIDE, Claudine Ichitani. Avaliando a qualidade do ensino médio na rede pública em relação à das escolas privadas na cidade de São Paulo Koide, Claudine Ichitani. Orientador: Prof. Dr. Naercio Aquino Menezes Filho. 2018. 73 f. Dissertação (Programa de Mestrado Profissional em Economia) - Insper - Instituto de Ensino e Pesquisa, [S. l.], 2018.

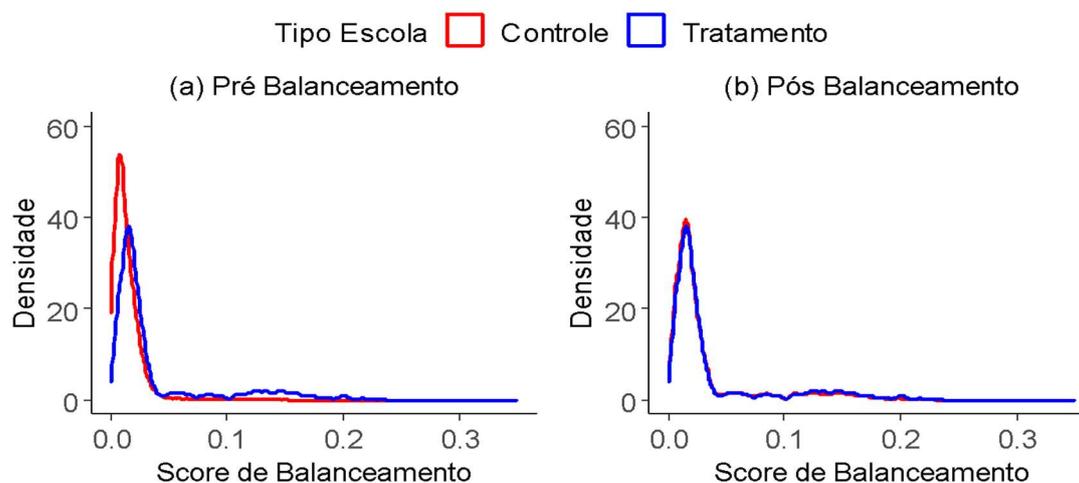
COSTA, Rafael dos Reis. Avaliação de impacto do projeto Escola Estadual de Tempo Integral em Goiás. 2017. 61 f. Dissertação (Mestrado em Administração) - Universidade Federal de Goiás, Goiânia, 2017.

DUTRA, Karoline Trindade. Efeito do PROUNI no desempenho acadêmico: uma estimação utilizando Propensity Score Matching. XX Encontro Nacional de Estudos Populacionais, Associação brasileira De Estudos Populacionais. [s. l.], p. 1-21, 2017.

LI, Yajuan; PALMA, Marco A.; YU, Zhicheng Phil. Impacts of playing after school on academic performance: a propensity score matching approach. *Education Economics*, [s. l.], v. 25, ed. 6, p. 575-589, 2017.

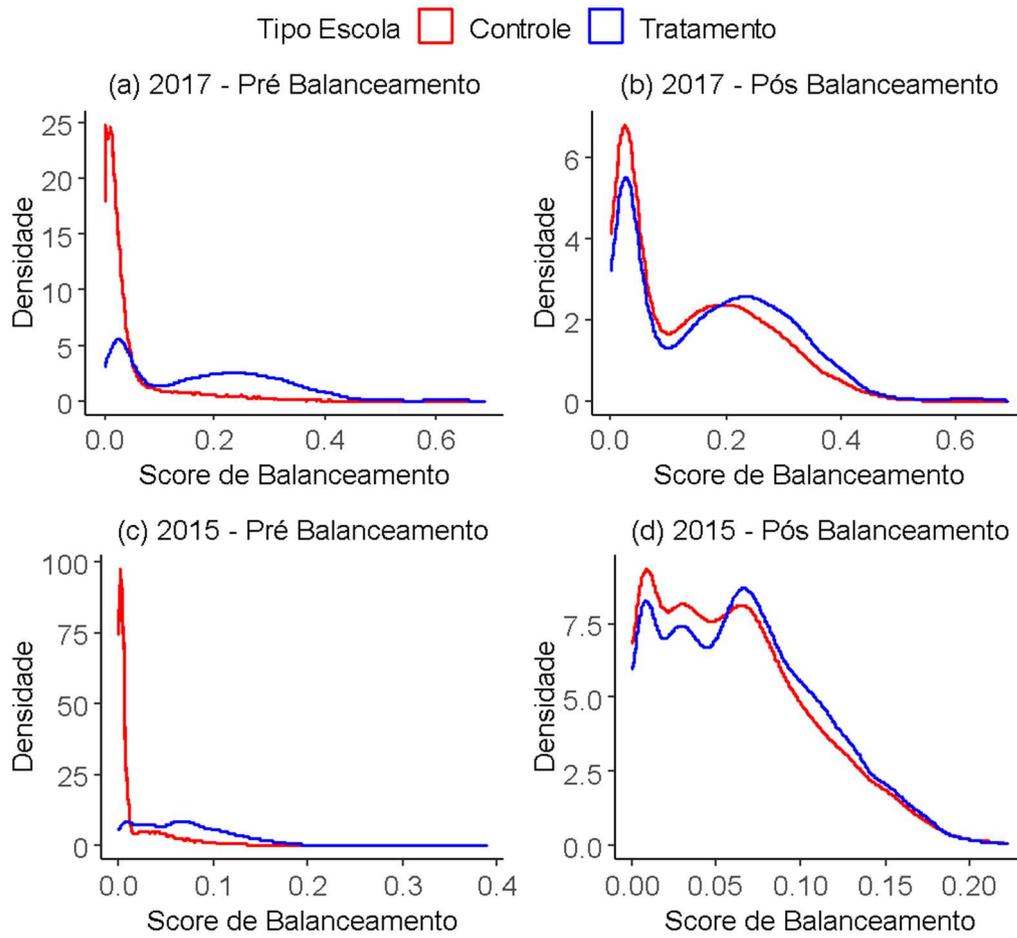
8. APÊNDICE

Figura A1 – Distribuição do *Propensity Score* Pré e Pós Pareamento 5ºs Anos



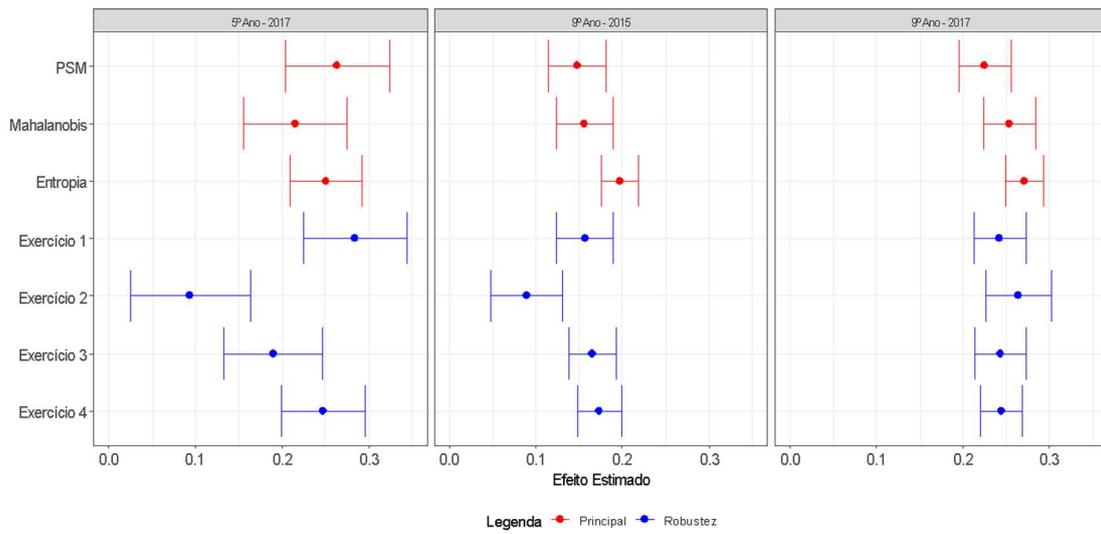
Fonte: Elaboração própria com dados da PB e do Censo Escolar e da SEESP.

Figura A2– Distribuição do *Propensity Score* Pré e Pós Pareamento 9ºs Anos



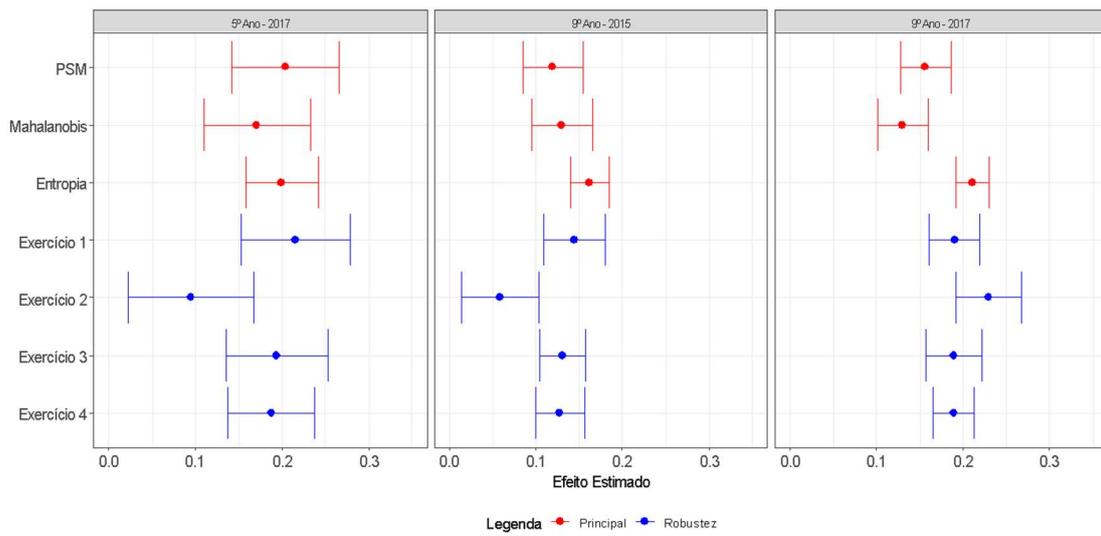
Fonte: Elaboração própria com dados da PB e do Censo Escolar e da SEESP.

Figura A3 – Resultados ATT Matemática



Fonte: Dados da Prova Brasil, Censo Escolar e da SEESP e elaboração própria.

Figura A4 – Resultados ATT Língua Portuguesa



Fonte: Dados da Prova Brasil, Censo Escolar e da SEESP e elaboração própria.

Tabela A1 – Resultados das estimativas do *Logit Propensed Score Matching*

Variáveis	5º Ano		9º Ano
	2017	2015	2017
Escola possui laboratório de informática	0.345*** (0.088)	13.39 (164.19)	-0.917*** (0.061)
Escola possui laboratório de ciências	1.990*** (0.068)	2.97*** (0.04)	2.430*** (0.027)
Escola possui quadra de esportes	0.739*** (0.118)	0.739*** (0.118)	0.130 (0.101)
Escola possui cozinha	12.177 -127.563	12.177 -127.563	12.012 -55.573
Escola possui biblioteca ou sala de leitura	0.965*** (0.085)	0.965*** (0.085)	2.280*** (0.149)
Escola possui internet	-0.234*** (0.073)	-0.234*** (0.073)	0.226*** (0.052)
Escola possui alimentação	13.449 -237.693	13.449 -237.693	0.712*** (0.204)
Número de Salas da escola	-0.104*** (0.006)	-0.104*** (0.006)	-0.155*** (0.003)
Sexo Feminino	-0.015 (0.052)	-0.015 (0.052)	0.037 (0.026)
Branco	-0.059 (0.055)	-0.059 (0.055)	-0.033 (0.026)
Idade adiantada na série	-0.112 (0.593)	-0.112 (0.593)	0.664*** (0.133)
Idade avançada na série	-0.531*** (0.164)	-0.531*** (0.164)	0.795*** (0.231)
Mora só com a mãe	-0.079 (0.081)	-0.079 (0.081)	0.012 (0.040)
Mora só com a mãe	-0.077 (0.163)	-0.077 (0.163)	-0.086 (0.072)
Mora com a mãe e o pai	-0.274*** (0.074)	-0.274*** (0.074)	-0.159*** (0.037)
Mãe tem ensino médio ou superior completo	0.257*** (0.054)	0.257*** (0.054)	0.262*** (0.026)
Mãe Alfabetizada	-0.101 (0.175)	-0.101 (0.175)	0.130 (0.091)
Mãe tem hábito de leitura	0.143 (0.096)	0.143 (0.096)	0.058 (0.042)
Realizar mais de 2 horas de trabalho doméstico	-0.132 (0.084)	-0.132 (0.084)	-0.268*** (0.038)
Estudante trabalha	0.013 (0.095)	0.013 (0.095)	-0.579*** (0.048)
Somente estudou em escola pública	-0.065 (0.063)	-0.065 (0.063)	-0.194*** (0.037)
Já reprovou	-0.255** (0.103)	-0.255** (0.103)	-0.431*** (0.039)
Já abandonou a escola	0.004 (0.113)	0.004 (0.113)	-0.013 (0.074)
PC1	-0.080*** (0.022)	-0.080*** (0.022)	-0.065*** (0.013)
Constant	-29.997 -269.760	-49.52 (241.46)	-16.942 -55.574
Observações	107,137	303,365	160,703

Fonte: Elaboração Própria.

Notas: Os níveis de significância são * $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$.