

EDI FLORES REYNA

**NOVAS EVIDÊNCIAS SOBRE O EFEITO DA PRÉ-ESCOLA NOS TESTES DE
PROFICIÊNCIA DO SAEB**

Dissertação apresentada à Universidade
Federal de Viçosa, como parte das exigências
do Programa de Pós-Graduação em Economia
Aplicada para a obtenção do título de *Magister
Scientiae*

VIÇOSA
MINAS GERAIS - BRASIL
2019

**Ficha catalográfica preparada pela Biblioteca Central da Universidade
Federal de Viçosa - Câmpus Viçosa**

T

F634n Flores Reyna, Edi, 1993-
2019 Novas evidências sobre o efeito da pré-escola nos testes de
proficiência do Saeb / Edi Flores Reyna. – Viçosa, MG, 2019.
vi, 55 f. : il. (algumas color.) ; 29 cm.

Inclui anexos.

Inclui apêndices.

Orientador: Maria Micheliana da Costa Silva.

Dissertação (mestrado) - Universidade Federal de Viçosa.

Referências bibliográficas: f. 40-44.

1. Educação pré-escolar - Aspectos econômicos. 2. Capital humano. 3. Rendimento escolar. 4. Análise de regressão.

I. Universidade Federal de Viçosa. Departamento de Economia Rural. Programa de Pós-Graduação em Economia Aplicada.

II. Título.

CDD 22. ed. 372.21

EDI FLORES REYNA

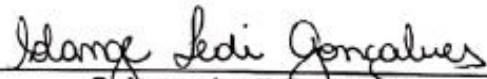
**NOVAS EVIDÊNCIAS SOBRE O EFEITO DA PRÉ-ESCOLA NOS TESTES DE
PROFICIÊNCIA DO SAEB**

Dissertação apresentada à Universidade Federal de Viçosa, como parte das exigências do Programa de Pós-Graduação em Economia Aplicada, para obtenção do título de *Magister Scientiae*.

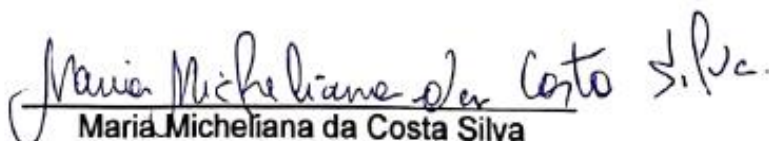
APROVADA: 19 de fevereiro de 2019.



Bladimir Carrillo Bermudez



Solange Ledi Gonçalves



Maria Micheliana da Costa Silva
(Orientadora)

À Litzy Celeste e Hillary Elizabeth, por serem
fonte da minha inspiração.

AGRADECIMENTO

Agradeço primeiramente a Deus por me permitir conhecer pessoas especiais e maravilhosas durante este período da minha vida e pela dádiva de ser parte de uma família que desde cedo me ensinou o verdadeiro significado do amor.

Ao meus pais, Wilson Flores Pezo e Liria Olivia Reyna Gonzales, exponho aqui minha gratidão e admiração. Vocês são e sempre serão a minha fonte de inspiração e orgulho. Agradeço aos meus irmãos e primas, Beldi, Corazón de Jesús, Ingri Ivoni, Wilson, Camilo Adrian, Litzy Celeste, Hillary Elizabeth e José Marcos, que fizeram da saudade e distância inspirações para perseverar a cada minuto da minha estadia fora de casa, a fim de alcançar cada objetivo traçado.

À minha orientadora professora Maria Micheliana da Costa Silva. A ela meu sincero obrigado por ter confiado em mim e pela paciência durante a elaboração deste trabalho. A professora Viviani Silva Lírio agradeço por cada sugestão no decorrer desta caminhada.

Aos meus amigos, Aline, Andrezza, Augusto, Bruna, Débora, Daniel, Cleyton, Gerciana, Jardesson, João Vitor, Peter e Rubicleis, agradeço sempre por ter conhecido cada um deles, por me permitirem ser parte da vida deles compartilhando experiências e conselhos, mais ainda, por terem sido parte desta grande aventura!

À UFV, ao Departamento de Economia Rural e ao Programa de Pós-Graduação em Economia Aplicada-PPGEA, deixo registrado meus agradecimentos pela oportunidade de continuar minha formação acadêmica e pela infraestrutura proporcionada. Aos professores do PPGEA, agradeço pelo apoio durante os meses de mestrado.

Aos professores que aceitaram participar da avaliação do projeto ou da banca de defesa da dissertação, agradeço por todas as contribuições.

Agradeço também ao CNPq pelo apoio financeiro imprescindível que recebi durante meu mestrado.

SUMÁRIO

RESUMO	v
ABSTRACT	vi
1 INTRODUÇÃO.....	1
1.1 Considerações Iniciais.....	1
1.2 Problema e sua Importância	5
1.3 Hipótese.....	8
1.4 Objetivos	8
1.4.1 Objetivo Geral.....	8
1.4.2 Objetivos Específicos.....	8
2 REFERENCIAL TEÓRICO	9
3 METODOLOGIA.....	12
3.1 Modelo Analítico: Estratégia de Identificação	13
3.2 Fontes de Dados	17
4 RESULTADOS E DISCUSSÕES.....	19
4.1 Pré-escola e notas do SAEB.....	19
4.2 Construção dos grupos de tratamento e controle	24
4.3 Resultados das estimativas da relação pré-escola e desempenho acadêmico.	28
5 CONCLUSÕES.....	37
6 REFERÊNCIAS.....	40
APÊNDICES.....	45
APÊNDICE A	45
APÊNDICE B	46
APÊNDICE C	47
APÊNDICE D	48
APÊNDICE E.....	51
ANEXOS.....	52
ANEXO I.....	52
ANEXO II	53

RESUMO

REYNA, Edi Flores, M.Sc., Universidade Federal de Viçosa, fevereiro de 2018. **Novas evidências sobre o efeito da pré-escola nos testes de proficiência do SAEB.** Orientadora: Maria Micheliana da Costa Silva. Coorientadora: Viviani Silva Lório.

Diversos estudos mostram os efeitos do cuidado e da educação durante a primeira infância no acúmulo de capital humano e nos resultados futuros dos indivíduos, como por exemplo, grau de instrução, estoque de saúde, produtividade e salários. Nesse contexto, o presente trabalho visou verificar o efeito de cursar a pré-escola sobre os testes de proficiência do Sistema de Avaliação da Educação Básica (SAEB), aplicado aos alunos do quinto ano do ensino fundamental no Brasil. Fornece-se novas evidências sobre o possível efeito da educação infantil na média e ao longo da distribuição das notas dos testes de proficiência, por meio de uma nova estratégia de identificação, baseada em uma estimação de duas etapas, considerando *inputs* contemporâneos e passados do processo educativo do indivíduo. Para tanto, foram utilizados os bancos de dados de Censo Demográfico de 2010 e o SAEB de 2015. A combinação dos dois bancos de dados foi exequível depois de se definir coortes por características invariantes no tempo, como sexo, raça, UF de residência, ano de nascimento e participação na pré-escola. Na primeira etapa, identificou-se grupos de tratamento e controle por meio do *propensity score matching*. Em seguida, obteve-se as estimativas de mínimos quadrados ordinários, considerando os pesos obtidos no pareamento. Além disso, estimou-se também uma regressão quantílica não condicional para verificar se o possível efeito da pré-escola varia ao longo da distribuição das notas. Os resultados encontrados corroboram com a hipótese de que cursar a pré-escola possui relação positiva com o desempenho acadêmico nas provas de proficiência em português e matemática dos alunos do 5º ano do ensino fundamental. Com a regressão quantílica, observou-se que há uma maior relação entre frequentar a pré-escola e as notas nos quantis mais altos dos escores do teste do SAEB, cujos resultados podem refletir habilidades cognitivas mais desenvolvidas. Mesmo não sendo possível definir o efeito exato da pré-escola sobre os testes, a estratégia sugerida permite, no mínimo, garantir a existência de uma relação positiva entre as variáveis. O coeficiente médio encontrado para pré-escola representa 9.1% do desvio padrão no teste de português e 10,1% no teste de matemática. Logo, pode-se afirmar que as crianças que frequentam a pré-escola tendem a possuir maior capacidade de adquirir os conhecimentos e habilidades que devem ser desenvolvidas nos anos iniciais do ensino fundamental.

ABSTRACT

REYNA, Edi Flores, M.Sc., Universidade Federal de Viçosa, February, 2018. **New evidences of the pre-school effects in Primary Education Assessment proficiency tests.** Adviser: Maria Micheliana da Costa Silva. Co-adviser: Viviani Silva Lírio.

Several studies show the effects of care and education in the early childhood in human capital accumulation and the individual future outcomes such as educational level, health stock and wage. In this context, this work seeks to verify the effects of pre-school over the Primary Education Assessment (Sistema de Avaliação da Educação Básica - SAEB) proficiency tests, applied to elementary school students in 5th grade in Brazil. It provides new evidences of the possible effects of child education in and along the average test grades distribution through a new identification strategy, based in a two steps estimation, considering contemporary and past inputs of the individual educational process. For this, it was used the 2010 Demographic Census and 2015 SAEB data. The combination of two different data was possible after defining cohorts by time invariants characteristics like sex, race, federation unity (UF), birth year and early childhood participation. In the first stage, was identified control and treated groups by propensity score matching. After, was obtained the OLS estimators considering the parity weight. Furthermore, was estimated an unconditional quantile regression to verify if the possible effects of child education varies in the grades distribution. The results found corroborate the hypothesis that pre-school is positively related in Portuguese and mathematical performance. With the quantile regressions, was observed a higher relationship between pre-school and the grades in higher quantiles of SAEB scores, which can reflect more developed cognitive abilities. Even though it was not possible to define the exact effect of pre-school over the tests, the suggested strategy allows at least guarantee the existence of a positive relationship between the variables. The average coefficient found in pre-school represents 9.1% of the standard deviation in Portuguese test and 10.1% in mathematics. So, it can be affirmed that children who attend pre-school tend to be more capable of acquiring the knowledge and abilities that must be developed in the initial years of elementary school.

1 INTRODUÇÃO

1.1 Considerações Iniciais

Desde a década de 1960, foram realizadas diversas pesquisas com o intuito de determinar os impactos, tanto micro como macroeconômico, do acúmulo do capital humano nas sociedades. Esta coletânea de estudos consolidou o papel da educação nas discussões sobre políticas públicas, relativo à redução das desigualdades sociais, e sobre estratégias de crescimento econômico dos países. Por esse motivo, conhecer detalhadamente o processo de acúmulo de capital humano e seus diversos efeitos tornou-se um desafio constante no mundo acadêmico. Estudos mostram que este capital gera impactos diretos e indiretos nos indivíduos, os quais são classificados como retornos privados e sociais. Estes fazem referência aos efeitos no ambiente socioeconômico do indivíduo, já aqueles referem-se aos benefícios obtidos apenas no nível individual como, por exemplo, melhores salários. Além disso, o conceito de capital humano também é incluído nas pesquisas de caráter macroeconômico como sendo um fator relevante para explicar as diferenças nos níveis de rendas e nas taxas de crescimento do PIB (Produto Interno Bruto) dos países (BARRO, 1990; BECKER, 1962; DE HOLANDA BARBOSA FILHO; DE ABREU PESSÔA, 2010; HANUSHEK, 2013; MINCER, 1970, 1974; ROMER, 1989; SCHULTZ, 2010, 1961).

O termo capital humano define-se como o acúmulo de habilidades cognitivas e não cognitivas adquiridas pelo indivíduo durante todo o ciclo da sua vida e, que após serem adquiridas, são indissociáveis dele. Soma-se a isso, a qualidade de saúde do indivíduo; existe uma relação recíproca entre o estoque de saúde e o processo de aquisição e de produtividade das habilidades. Portanto, o nível educacional e a qualidade de saúde do indivíduo fornecem uma ideia da qualidade e quantidade de capital humano que ele é capaz de acumular (BECKER, 1964; CARNEIRO; HECKMAN, 2003; CURRIE, 2009; RUSSO; DIAS, 2016; SCHULTZ, 1961)

Durante muito tempo a média de anos de estudos dos indivíduos foi utilizada como medida do estoque de capital humano. Porém, para Hanushek (2013) e Hanushek e Woessmann (2008, 2011), este tipo de medida apresenta um conjunto de dificuldades. De acordo com os autores, esta perspectiva reconhece que os anos estudos são a única fonte de capital humano, ignorando as dimensões importantes, entre elas a saúde. Assim sendo, esta forma de mensurar o capital humano pode fornecer uma ideia pouco precisa sobre impacto nos resultados futuros

dos indivíduos. Russo e Dias (2016), por exemplo, mostraram que pessoas saudáveis tendem a ser mais produtivas no trabalho afetando diretamente seus retornos privados (salários).

Atualmente, existe uma abordagem muito em voga na literatura internacional que define como proxy do capital humano a proficiência dos alunos em testes como o *Programme for international student Assessment* (Programa Internacional de Avaliação de Estudantes-PISA). Esta abordagem foi introduzida por Lee e Lee (1995) e adotada nos trabalhos de Hanushek e Kimko (2000), Hanushek e Woessman (2008, 2011a, 2011b, 2012), Hanushek (2013), Balart, Oosterveen e Webbink (2018) e outros. O PISA é uma iniciativa da OCDE (Organização para a Cooperação e Desenvolvimento Econômico) que visa fornecer indicadores da qualidade educacional dos países através da aplicação de testes de matemática, ciência e leitura em uma amostra de estudantes na faixa etária dos 15 anos.

No Brasil, o sistema que provê as informações sobre a educação básica do país é o Sistema de Avaliação da Educação Básica (SAEB). Este contém um conjunto de avaliações sobre o desempenho dos alunos do 5º e 9º ano do ensino fundamental e do 3º ano do ensino médio nas disciplinas de matemática e português. Desde a sua criação, em 1990¹, o SAEB avalia as escolas públicas a cada 2 anos, e desde 1997 uma amostra de escolas privadas fazem parte desta avaliação. Como o PISA, as provas de desempenho do SAEB são divididas em quantis de notas, em que cada um deles reflete os diversos níveis de habilidades que o estudante possui (INEP, 2018a, 2018b, 2018c).

A nova abordagem pretende medir o estoque de capital humano mediante as notas em testes de proficiência. As notas refletem o conjunto de habilidades cognitivas e não cognitivas adquiridas pelos indivíduos ao longo da vida acadêmica, o estoque de saúde e a qualidade educacional das escolas. Dessa forma, tem-se uma noção mais real sobre o estoque de capital humano dos indivíduos (BALART; OOSTERVEEN; WEBBINK, 2018; BORGHANS; SCHILS, 2012; HANUSHEK, 2013; HANUSHEK; WOESSMANN, 2008; LEE; LEE, 1995).

Como descrito por Balart, Oosterveen e Webbink (2018) os dois tipos de habilidades supracitadas apresentam significância estatística ao explicar as notas dos alunos nos testes do PISA. Isto deve-se ao fato de que a performance do aluno, durante a primeira parte do teste, é definida principalmente pelas capacidades cognitivas, ao passo que o declínio da mesma está estritamente atrelado ao nível de motivação e esforço dos alunos, habilidades não cognitivas

¹ Desde o ano de 2001 o foco de avaliação do SAEB são as áreas de matemática e português.

(BALART; OOSTERVEEN; WEBBINK, 2018; BORGHANS; MEIJERS; TER WEEL, 2008; BORGHANS; SCHILS, 2012; HECKMAN; PINTO; SAVELYEV, 2013).

O capital humano medido na dimensão das habilidades é um ótimo preditor de resultados futuros para os indivíduos (BORGHANS; SCHILS, 2012). A dotação de habilidades cognitivas e não cognitivas dos indivíduos afeta o grau de escolaridade que eles conseguem alcançar, os seus estoques de saúde e seus comportamentos no entorno social na vida adulta. Para o caso específico de habilidades não cognitivas (motivação, esforço, tenacidade, paciência, aversão ao risco, autoestima e autocontrole), elas potencializam o aprendizado de novas habilidades, melhoram a inserção do agente econômico no mercado de trabalho e o seu bem-estar. Estes efeitos se estendem para a sociedade através da redução do crime, do número de fumantes, da taxa de gravidez na adolescência, do abandono escolar e do aumento da média da expectativa de vida (ALMLUND et al., 2011; CARNEIRO; CUNHA; HECKMAN, 2004; CARNEIRO; HECKMAN, 2003; HANUSHEK, 2013; HECKMAN, 2006, 2007; HECKMAN; PINTO; SAVELYEV, 2013)

Nesse sentido, compreender o processo de formação das habilidades pode ajudar a oferecer um conjunto de estratégias mais eficientes para os formuladores de políticas públicas. A neurociência, a psicologia comportamental, a biologia molecular, a epigenética e a economia avançaram muito em relação ao estudo deste processo (CENTER ON THE DEVELOPING CHILD., 2016). Essas ciências concordam que a tecnologia de formação de habilidades é um processo acumulativo que tem seu início logo após o nascimento e se estende durante toda a vida do indivíduo. (CASTRO; ROLLESTON, 2018; CUNHA; HECKMAN, 2007; HECKMAN, 2006, 2007, 2008; HECKMAN; PINTO; SAVELYEV, 2013).

A formação de habilidades é um processo sequencial de estágios que podem incluir períodos críticos e sensíveis para o desenvolvimento cerebral. Define-se como período sensível aquele período em que é mais efetivo aprender determinadas habilidades. Por outro lado, se uma habilidade é produzida apenas em um único período, este será definido como período crítico. Estes períodos podem começar e terminar em diferentes momentos da vida, por exemplo: a construção do circuito neural da visão, escuta e tato começam e terminam antes dos 5 anos. A sequência de estágios revela que as habilidades desenvolvidas pelos indivíduos possuem complementariedade dinâmica e estática. Esta se refere ao impacto das habilidades inatas sobre a criação de habilidades futuras, já aquela refere-se ao efeito das habilidades

adquiridas no passado sobre as habilidades do presente (HECKMAN, 2006, 2007; CENTER ON THE DEVELOPING CHILD., 2016).

Embora a formação de habilidades seja um processo que acontece durante todo o ciclo de vida das pessoas, as bases da arquitetura do cérebro são estabelecidas durante a primeira infância. Além disso, a capacidade de adaptação do cérebro apresenta uma tendência decrescente em relação ao tempo. Portanto, a primeira infância é um período de mudanças rápidas e mais abertas a influências tanto positivas como negativas por meio dos relacionamentos e das experiências (CENTER ON THE DEVELOPING CHILD., 2016).

As primeiras experiências vivenciadas pelas crianças afetam a natureza e a qualidade das bases da arquitetura cerebral. Situações de estresse como ambientes familiares desfavoráveis, desnutrição entre outros podem minar o desenvolvimento cognitivo e não cognitivo das crianças gerando efeitos negativos duradouros. A consequência destas situações pode ser um ciclo de pobreza intergeracional; pais que tiveram um ambiente hostil durante a infância possuem um baixo estoque de habilidades cognitivas e não cognitivas o que resulta em menores níveis de renda na vida adulta, isto por sua vez pode influenciar o desenvolvimento cerebral do seus filhos (CURRIE, 2009; HECKMAN; PINTO; SAVELYEV, 2013; PAXSON; SCHADY, 2007).

De acordo com Heckman, Pinto e Savelyev (2013), o QI (Quociente de inteligência) do indivíduo se estabelece aproximadamente aos 10 anos, sugerindo que os 10 primeiros anos de vida é um período sensível para formação de habilidades. A base da arquitetura cerebral relativa às competências sociais, por sua vez, é estabelecida durante os 5 primeiros anos. Estas se relacionam com a capacidade das crianças de se adaptar aos desafios da vida na escola, permitindo assim maior facilidade para aprender através das experiências. Aliás, a aquisição de altos níveis de habilidades não cognitivas como a de prestar atenção, seguir regras, resolver problemas, motivação e controlar impulsos também são desenvolvidas na primeira infância. Portanto, o ambiente de convívio e cuidados na primeira infância são relevantes para o desenvolvimento da criança, gerando assim, a ativação de potenciais genéticos positivos ou negativos relativos às habilidades (CENTER ON THE DEVELOPING CHILD, 2016).

Em virtude da importância da primeira infância na formação de habilidade, Heckman (2007, 2008) afirma que os investimentos nesta etapa são os que geram maiores retornos econômicos, medidos nos rendimentos futuros, tornando-os mais eficientes. Becker (1964), por sua vez, concluiu que investir na infância proporciona maior tempo para usufruir dos benefícios

de dito investimento. Do mesmo modo, Benzil, Hansen e Kristensen (2008), alegam que capacitar adultos com baixo estoque de habilidades é menos efetivo do que investir em crianças. Isto porque o aprendizado do indivíduo durante a primeira infância é mais fácil quando comparado com as outras etapas educacionais, dado a plasticidade cerebral das crianças e o processo “*learning begets learning*” (CARNEIRO; CUNHA; HECKMAN, 2004).

Destarte, a educação na primeira infância possui relevância para desenvolver as habilidades necessárias que permitam um maior acúmulo de capital humano. Assim sendo, a relação a ser explorada nesta pesquisa é a existente entre a educação infantil e os escores dos testes de proficiência. Dessa maneira, define-se a educação infantil como objeto de estudo, especificamente a pré-escola².

1.2 Problema e sua Importância

A pré-escola pode ser pensada como um investimento complementar ou compensatório, realizado pelas famílias e pelo Estado. Entende-se como um investimento complementar quando existe um ambiente favorável para as crianças dentro do lar. Esta primeira experiência longe dos pais seria considerada como uma situação de estresse positiva, que gera um desenvolvimento cognitivo e não cognitivo saudável. Por outro lado, no caso de investimento compensatório, a pré-escola tem o papel fundamental de criar uma relação estável entre a criança e seu responsável, fornecendo assim suporte e segurança para construção das capacidades chave que permitam à criança se adaptar melhor às mudanças e fazer frente às adversidades. (CURRIE; VOGL, 2013; CENTER ON THE DEVELOPING CHILD., 2016)

Qualquer que seja o caso, intervenções na infância como a pré-escola podem ajudar a evitar situações como as relatadas nos testes do PISA. Hanushek (2013) mostra que 69,1% dos estudantes brasileiros participantes do PISA matemática de 2009 se encontravam no nível 1. De acordo com o autor, com este nível de conhecimento os estudantes teriam muita dificuldade de participar da força de trabalho que inclua novas tecnologias. O limitado estoque de capital incapacitaria o processo de adaptação dos indivíduos às constantes mudanças tecnológicas no mercado de trabalho.

² No Brasil, a pré-escola contempla crianças entre os 4 e 5 anos de idade.

Nesse contexto, o presente problema de pesquisa se resume em encontrar qual é o efeito de frequentar a pré-escola sobre o capital humano acumulado na infância, medido por meio de testes de proficiência do Sistema de Avaliação da Educação Básica (SAEB).

Evidências na literatura internacional mostram os efeitos da educação infantil sobre o acúmulo de capital humano. Os programas desenvolvidos nos Estados Unidos, *Perry Preschool Program*, *The Carolina Abecedarian* e *Head Start*, mostraram efeitos positivos sobre as notas de proficiências, redução da taxa de encarceramento, redução da delinquência e crime e sobre a probabilidade de terminar o ensino médio e frequentar a faculdade. As estratégias de identificação dos efeitos dos programas *Perry Preschool* e *The Carolina Abecedarian* levaram em conta a randomização dos grupos de tratamento e controle, já para o *Head Start* foi utilizado dados longitudinais com efeito fixo de irmão para controlar a endogeneidade da família (CURRIE, 2001; GARCES; THOMAS; CURRIE, 2002; HECKMAN; PINTO; SAVELYEV, 2013). Na Dinamarca, Bauchmüller, Gortz e Rasmussen (2014), fazendo uso do MQO, defendem a existência apenas de uma correlação entre qualidade da pré-escola e as performances dos alunos no final da escola primária³.

Na América Latina existem pesquisas que vinculam a educação infantil e o acúmulo de habilidades. Na Bolívia, por exemplo, mostrou-se os efeitos positivos sobre os resultados cognitivos e psicossociais para crianças com mais de 7 meses de exposição ao programa PIDI (*Programa Integral de Desarrollo Infantil*) (BEHRMAN; CHENG; TODD, 2004). Do mesmo modo Berlinski, Galiani e Gertler (2009) ao explorar um experimento natural, a expansão das infraestruturas pré-escolares na Argentina, encontraram que as habilidades comportamentais adquiridas pelos estudantes na pré-escola têm efeito positivo sobre os testes. Em suma, os autores encontraram que um ano a mais de pré-escola aumenta em 8% a média nos testes de proficiência para os alunos do terceiro ano do ensino fundamental⁴. No Chile, Cortázar (2015) baseou seu estudo no método de *propensity score matching* com uso de variáveis observáveis contemporâneas aos testes e também encontrou resultados similares para a relação entre programas de educação na primeira infância⁵ e testes de proficiências, porém ela alega que os

³ No final da escola primária, no sistema educativo dinamarquês, os indivíduos têm entre 15 e 16 anos de idade (BAUCHMÜLLER; GORTZ; RASMUSSEN, 2014).

⁴ O sistema de educação básica argentino obrigatório possui duas opções de estrutura: A primeira estrutura um está composta por 2 anos de educação infantil, 6 anos de ensino fundamental e 6 anos de ensino médio; A segunda está composta por 2 anos de educação infantil, 7 anos de ensino fundamental e 5 anos de ensino médio.

⁵ Na pesquisa feita por Cortázar (2015) foram considerados programas antes da pré-escola uma vez que o acesso a pré-escola é universal.

benefícios não são iguais entre as crianças, pois varia conforme sexo e status socioeconômico das famílias.

Mesmo com diversos estudos sobre a primeira infância na literatura, ainda existe espaço para discussão sobre a importância da pré-escola. No Brasil, não há dados longitudinais de políticas de educação infantil e nem experimentos de abrangência nacional que permitam adotar as estratégias utilizadas nos artigos internacionais, sendo necessária a implementação de novas estratégias de identificação. Além disso, cada país tem características e condições culturais, sociais, políticas e econômicas próprias que não permitem definir uma abordagem global e única sobre o fornecimento de educação infantil no mundo (OLMSTED; WEIKART, 1989). Ademais, é reconhecido que dentro do próprio Brasil existe uma grande heterogeneidade socioeconômica.

A análise do efeito da educação infantil na distribuição das notas dos testes é uma lacuna ainda não explorada pela literatura. Ao analisar o comportamento desta distribuição, pode-se fornecer o efeito da pré-escola sobre diversos níveis e tipos de habilidades desenvolvidas pelos alunos durante os primeiros 5 anos do ensino fundamental. No quantil mais baixo da distribuição das notas, por exemplo, pode-se dizer que as crianças não possuem as habilidades elementares para sua idade, precisando, portanto, de atenção especial. Por outro lado, nos quantis mais elevados, as crianças apresentam habilidades mais aprimoradas, tais como facilidade de fazer inferências a partir de diversos tipos de textos ou imagens e de resolver problemas de caráter monetário (soma e subtração de valores monetários e conversão de uma quantia em moedas ou notas de valores menores) (INEP 2018b, 2018c).

Os trabalhos desenvolvidos no Brasil indicam os efeitos da educação infantil sobre diversos resultados. Curi e Menezes-Filhos (2009), por exemplo, encontraram uma relação positiva entre a educação pré-primária e o nível educacional, a proficiência escolar e os rendimentos futuros dos indivíduos. Os resultados foram estimados por MQO controlando por variáveis observáveis e contemporâneas. Felício e Vasconcelos (2007), por sua vez, encontraram efeitos positivos da educação infantil sobre os testes de proficiência em matemática. Frequentar a pré-escola aumenta a média nos testes de proficiência em 6,3% para os alunos da quarta série. Os autores trabalharam sob a hipótese de que as características dos indivíduos e das famílias foram invariáveis no tempo, portanto, usaram dados obtidos no ano da aplicação dos testes para realizar o pareamento.

Em ambos os casos, os autores fizeram hipóteses relativamente fortes para controlar o processo decisório da família de inserir a criança na educação infantil. Considerar apenas inputs contemporâneos significa dizer que somente eles são suficientes para determinar a produção cognitiva, pois conseguem capturar perfeitamente os efeitos dos inputs passados, tornando-os invariantes no tempo. Porém, a decisão de matricular a criança na educação infantil difere do momento da aplicação dos testes, portanto, os inputs podem variar e ter impactos diferentes a depender do período. A renda familiar, por exemplo, apresenta tendência crescente através do ciclo da vida (PARISH; WILLIS, 1993). Ao mesmo tempo, Todd e Wolpin (2003, 2004) mostram que não há nenhum suporte para utilizar modelos restritos contemporâneos devido às fortes suposições que devem ser feitas.

Dentro deste contexto, o presente trabalho contribui para a literatura nacional e internacional de duas formas: fornecendo novas evidências sobre o possível efeito da educação infantil ao longo da distribuição das notas dos testes de proficiência. Outrossim, propõe-se uma nova estratégia de identificação que permita que o acesso à pré-escola tente se aproximar de um experimento ideal, por se basear em uma estimação de duas etapas considerando *inputs* contemporâneos e passados do processo educativo do indivíduo. Isto possibilita controlar a tomada de decisão das famílias de matricular as crianças na pré-escola no período correspondente e, posteriormente, controlar o desempenho nos testes no momento da sua aplicação.

1.3 Hipótese

A pré-escola tem impactos positivos no desempenho acadêmico em português e matemática dos alunos do 5º ano do ensino fundamental. Esses efeitos são maiores nos quantis mais altos dos escores do teste do SAEB.

1.4 Objetivos

1.4.1 Objetivo Geral

Identificar o efeito da pré-escola sobre o teste de proficiências de português e matemática dos alunos do 5º ano do ensino fundamental em 2015.

1.4.2 Objetivos Específicos

- a. Verificar os fatores que influenciaram a decisão em frequentar a pré-escola, cinco anos antes da realização do teste.
- b. Analisar o efeito da pré-escola sobre as notas nos testes de proficiência do SAEB.

- c. Identificar o impacto da pré-escola nos diferentes quantis de nota.

2 REFERENCIAL TEÓRICO

Nesta seção apresenta-se a teoria que dará suporte à presente investigação. Adota-se aqui o marco teórico utilizado por Heckman (2008), Cunha e Heckman (2007), Aizer e Cunha (2012), Castro e Rolleston (2018) e Adhvaryu e Nyshadhan (2016).

A capacidade de produção de habilidades é um processo que inclui vários estágios concatenados entre si, em que cada um deles representa uma etapa da vida de um indivíduo, aqui cognominado, indivíduo i . Os estágios decorridos são responsáveis por gerar os inputs necessários para os estágios subsequentes. Logo, assume-se que o nível de capital humano do estágio $t+1$ está em função de variáveis contemporâneas e passadas.

Formalmente define-se a função de produção de capital humano da seguinte maneira:

$$A_{it+1} = f_t(\beta, A_{it}, I_{itk}, \sigma) \quad k \in \{S_c, S_n\} \quad (1)$$

Na equação (1) a variável A_{it+1} denota o vetor de habilidade cognitivas (S_c) e não cognitivas (S_n) desenvolvidas pelo indivíduo i no tempo t . O vetor β representa inputs relativos a background familiar; A_{it} mostra as habilidades adquiridas no período anterior. O vetor I_t inclui os investimentos feitos para aprimorar as habilidades. Neste trabalho a variável I_t representará os investimentos realizados pelos pais para melhorar as habilidades das crianças. σ é um escalar que representa as habilidades inatas dos indivíduos.

As habilidades inatas são características não observáveis difíceis de mensurar que geram endogeneidade para os modelos econométricos. Porém, conforme Castro e Rolleston (2018), esse elemento não introduz um viés significativo no acúmulo de capital humano. Mesmo assim, esta característica sugere cautela na análise da relação de interesse desta pesquisa. As variáveis pré-natais, por exemplo, podem influenciar a predisposição genética dos indivíduos.

Assume-se também que esta função é estritamente crescente em I_t . Define-se $f(.)$ como sendo côncava em I_t e duas vezes continuamente diferenciável em todos seus elementos. A partir disso pode-se ter que maior investimento nos indivíduos produz mais habilidades, dado que $\frac{\partial f_t}{\partial I_t} > 0$.

A complementariedade dinâmica das habilidades é definida pela derivada cruzada entre habilidades no período t e o investimento no período $t+1$, $\frac{\partial^2 f_t}{\partial A_t \partial I_t} > 0$. Esta complementariedade

explica como os investimento em estgios superiores so mais efetivos em indivduos que possuem grandes estoques de habilidades. Logo, os estudantes com altos nveis de capital humano na infncia demonstram maior eficincia na aquisio de novas habilidades.

Com relao  auto-produtividade, entende-se que vetor de A_t permite a criao de maiores nveis de habilidades no perodo A_{t+1} com grande facilidade. Portanto,  possvel escrever que $\frac{\partial f_t}{\partial A_t} > 0$. Esta, conjuntamente como a complementariedade dinmica, mostra como o investimento na primeira infncia gera maiores retornos e so mais eficientes do que em perodos posteriores. Estas duas caractersticas chave geram um efeito multiplicador que pode ser chamando de “*skill begets skill*”.

Nessa perspectiva, esta pesquisa levar em considerao um processo de acumulao de capital de dois estgios, seguindo a modelo apresentado por Heckman (2008) e Cunha e Heckman (2007). A equao que denota o processo  designada como:

$$A_{i2} = f_2(\beta, A_{i1}, I_1, I_2, \sigma) \quad (2)$$

A equao (2) pode ser escrita assumindo que a relao entre os investimentos apresenta a forma funcional de uma CES (*Constant Elasticity of Substitution*), permitindo assim relacionar os investimentos em dois perodos distintos, passado e presente.

$$A_{i2} = f_2\left(\beta, A_{i1}, \sigma, [\gamma(I_1)^\phi + (1 - \gamma)(I_2)^\phi]^{\frac{1}{\phi}}\right); \phi \leq 1 \text{ e } 0 \leq \gamma \leq 1 \quad (3)$$

O parmetro ϕ denota o grau de substituio entres os investimentos feitos no perodo 1 e 2. O parmetro γ captura os efeitos da auto-produtividade e complementariedade das habilidades (“*skill begets skill*”) permitindo observar os efeitos diretos e indiretos dos investimentos. Dado a complementariedade entre os investimentos, discutida amplamente por Heckman (2008) e Cunha e Heckman (2007), assume-se que:

$$\lim_{\phi \rightarrow -\infty} A_{i2}(\cdot) \quad (4)$$

$$A_{i2} = \min\{I_1; I_2\} \quad (5)$$

Desse modo, para valores pequenos de ϕ observa-se que intervenes tardias no podem compensar a falta de investimento em capital humano de estgios iniciais da vida.

Define-se o preo do investimento no perodo 1 igual a R\$ 1, no perodo 2 igual a R\$ $\frac{1}{1+r}$, r sendo a taxa de juros da economia e $\frac{1}{1+r}$ um fator de desconto. A quantidade de capital humano

produzida por uma unidade de I_1 e R\$ $(1+r)$ de I_2 são dadas por γ e $(1+r)(1-\gamma)$ respectivamente (CUNHA; HECKMAN, 2007; CUNHA; HECKMAN; SCHENNACH, 2010). A partir disso, e por meio de um simples processo de maximização se obtém que a razão ótima de investimento entre os dois períodos é dada por:

$$\frac{I_1}{I_2} = \left[\frac{\gamma}{(1-\gamma)(1+r)} \right]^{\frac{1}{1-\phi}} \quad (6)$$

O nível de investimento da família nas crianças é definido também por um processo de maximização de utilidade do domicílio, como mostrado no trabalho de Adhvaryu e Nyshadham (2016). Nesse sentido, a equação (7) é uma adaptação do processo de maximização de utilidade domiciliar utilizada pelos autores mencionados.

$$\begin{aligned} U &= u(A_{it}) + C \\ \text{s. a. } \rho I_t + C &\leq M \end{aligned} \quad (7)$$

Onde $u(\cdot)$ é uma função crescente, côncava e duas vezes diferenciável em cada um dos seus elementos. C representa o consumo e será considerado como um bem numérico; a renda do domicílio é dado por M .

Pelas condições de primeira ordem da maximização em relação a I_t e C se obtém:

$$\frac{\partial u(\cdot)}{\partial A_{it}} \left(\frac{\partial A_{it}}{\partial I_t} \right) = \lambda \rho \quad (8)$$

$$1 = \lambda \quad (9)$$

Para observar a complementariedade dos investimentos com relação as habilidades adquiridas no período anterior ou até mesmo com as habilidades inatas se recorre ao teorema da função implícita. Logo obtém-se que:

$$\frac{\partial A_{it}}{\partial I_t} \left[\frac{\partial^2 u}{\partial^2 A_{it}} \left(\frac{\partial A_{it}}{\partial I_t} \frac{\partial I_t}{\partial A_{it-1}} + \frac{\partial A_{it}}{\partial A_{it-1}} \right) \right] + \frac{\partial u}{\partial A_{it}} \left[\frac{\partial^2 A_{it}}{\partial^2 I_t} \frac{\partial I_t}{\partial A_{it-1}} + \frac{\partial^2 A_{it}}{\partial I_t \partial A_{it-1}} \right] = 0 \quad (10)$$

Resolvendo para $\frac{\partial^2 A_{it}}{\partial I_t \partial A_{it-1}}$, encontra-se a equação (11)

$$\frac{\partial^2 A_{it}}{\partial I_t \partial A_{it-1}} = - \frac{\left[\frac{\partial u}{\partial A_{it}} \frac{\partial^2 A_{it}}{\partial^2 I_t} \frac{\partial I_t}{\partial A_{it-1}} + \frac{\partial A_{it}}{\partial I_t} \left[\frac{\partial^2 u}{\partial^2 A_{it}} \left(\frac{\partial A_{it}}{\partial I_t} \frac{\partial I_t}{\partial A_{it-1}} + \frac{\partial A_{it}}{\partial A_{it-1}} \right) \right] \right]}{\frac{\partial u}{\partial A_{it}}} \quad (11)$$

O sinal desta derivada implícita pode oferecer ideia de como os investimentos reagem às dotações iniciais ou estoque de capital de períodos anteriores. O sinal negativo significa que o investimento de tempo t seria substituto aos estoques de capital do tempo $t-1$, mostrando que existiria um processo de compensação. Já o sinal positivo mostraria o processo de complementariedade entre essas duas variáveis. Para este trabalho assume-se a existência do processo de complementariedade dinâmica entre as habilidades, pois leva-se em consideração as discussões feitas por Cunha e Heckman (2007) e Heckman (2008).

3 METODOLOGIA

A fim de identificar a relação entre a participação na pré-escola e as notas do SAEB, buscou-se primeiramente um experimento natural que torne essa decisão de exógena ou aleatória. Encontrar este tipo de relação possui relevância no quesito de políticas públicas, pois revela a efetividade dos investimentos públicos em programas sociais. O experimento ideal eliminaria o viés de seleção através da criação de dois grupos aleatórios, com características similares, em média. Consequentemente, esta aleatoriedade permitiria definir a relação de causa e efeito entre a variável de interesse e as variáveis dependentes (notas de proficiência em português e matemática) uma vez que seria possível gerar um grupo contrafactual.

Um experimento natural que geraria a variação exógena sobre a decisão de ingressar na pré-escola é a lei N° 12.796, de 4 de abril de 2013. Esta lei estabelece a obrigatoriedade da educação básica desde os 4 anos de idade até os 17 anos, organizada em: pré-escola, ensino fundamental e ensino médio. Todavia, a lei apresentou um prazo de adequação para os municípios e estados até 2016, impossibilitando assim utilizar os resultados desta variação exógena para a presente análise (MEC, 2018; BRASIL, 2018).

Por outro lado, foi cogitado realizar a avaliação da educação infantil pelo lado da oferta através do programa Proinfância do FNDE⁶, seguindo a estratégia aplicada na Argentina por Berlinski, Galiani e Gertler (2009). Contudo, ser beneficiado pelo programa Proinfância não é definido aleatoriamente, dado que é o próprio município quem manifesta interesse em ser beneficiado pelo programa por meio da elaboração de um Plano de Ações Articulada. Portanto, existem características não observáveis de cada município que fazem com que eles sejam

⁶ Fundo Nacional de Desenvolvimento da Educação.

beneficiados, como por exemplo, a motivação e nível de compromisso dos governantes para com a população. (FNDE, 2018).

Desse modo, ao não ser encontrado um evento exógeno ou experimento natural para a decisão de participar da pré-escola, o trabalho apresenta uma nova estratégia de identificação que tenta imitar o contexto de randomização para essa decisão. Esta será apresentada na subseção subsequente.

3.1 Modelo Analítico: Estratégia de Identificação

Em cada etapa da estratégia de identificação, utilizou-se um método distinto -*Propensity Score Matching* e Mínimos Quadrados Ordinários-, os quais possuem pressupostos diferentes que devem ser atendidos para conseguir efeitos causais. Nessa lógica, as estimativas realizadas tentaram atender os pressupostos; entretanto, ainda existe a probabilidade de que eles não sejam satisfeitos, devido a limitações dos dados existentes. Não obstante, isso não impossibilita a utilização desta abordagem, pois a combinação dos métodos irá fornecer maior robustez aos resultados encontrados.

Uma vez que não é possível acompanhar o mesmo indivíduo no momento em que houve a decisão de entrar na pré-escola e na realização do teste de proficiência, cinco anos depois, recorre-se a construção de coortes, com os dados provenientes do Censo Demográfico 2010 e do SAEB de 2015. Para agregar estes bancos de dados, foram estabelecidas coortes considerando as seguintes características fixas dos indivíduos: sexo, raça, ano de nascimento, UF e participação na pré-escola. A fim de minimizar o problema de atrito na amostra, foram excluídos dos coortes aqueles indivíduos que reprovaram pelo menos uma vez ou abandonaram a escola anteriormente. Os indivíduos que declararam ter entrado na escola após o primeiro ano do ensino fundamental também foram removidos da amostra. Assume-se que aqueles indivíduos que ingressaram depois na escola apresentaram uma defasagem de conhecimento em relação aos seus pares, portanto, pode-se supor que o rendimento dos mesmos será menor por causa do período de adaptação escolar.

Com os dados domiciliares do Censo 2010, identificou-se os grupos de tratamento e controle usando as bases do *Propensity Score Matching*. Em programas nos quais o tratamento não é designado de forma aleatória, a simples comparação entre indivíduos do grupo controle e tratado gera estimativas enviesadas. Em ambos os grupos, existem fatores tanto observados como não observados altamente correlacionados com a decisão de participar de um dado tratamento. Estes fatores de confusão, como os define Becker e Ichino (2002), são minimizados

quando se identifica grupos de indivíduos que são tão semelhantes quanto possível, com base num vetor de características observáveis (X).

O método de *Propensity Score Matching* gera uma única variável, denominada *propensity score* (escore de propensão), que resume um conjunto de características observadas em cada indivíduos antes do tratamento (BECKER; ICHINO, 2002). A partir dessa variável, identifica-se os grupos de controle e tratado que são comparáveis entre si. A efetividade do método está sujeita ao cumprimento de duas suposições: a CIA (Suposição de Independência Condicional) e o equilíbrio entre as variáveis (*Balancing Hypothesis*). A CIA assume que dado um conjunto de características observáveis dos indivíduos os resultados potenciais são independentes do *status* de participação do programa. Formalmente Becker e Ichino (2002) o define como:

$$\begin{aligned} D &\perp X|p(X) \\ Y_1, Y_0 &\perp D|p(X) \end{aligned} \quad (12)$$

Por sua vez, a hipótese de equilíbrio afirma que se as observações possuem o mesmo *propensity score*, então elas devem ter a mesma distribuição de características observáveis (X), e não observáveis independentemente de ser beneficiado o não pelo tratamento. Em outras palavras, o status de tratamento deve ser ortogonal as características dos indivíduos dado o *propensity score* ($D \perp X|p(x)$), o que significaria que as unidades de tratamento e controle são em média observacionalmente idênticas (BECKER; ICHINO, 2002).

Assim sendo, o efeito médio do tratamento (ATE) medido por este método é calculado por:

$$ATE = E \left[\frac{(D_i - p(X_i))Y_i}{(1 - p(X_i))P(D_i)} \right] \quad (13)$$

Onde D representa uma *dummy* que mostra se o indivíduo fez ou não parte do tratamento. Y é o vetor de resultados do tratamento; e $p(X_i)$ é chamado por Wooldridge (2002) como função de *propensity score*.

Como foi mencionado anteriormente, a relação entre as variáveis de interesse desta pesquisa não será calculada por este método, uma vez que foi utilizado somente para a criação dos grupos de controle e tratado. Nesse ponto de vista, atenta-se aqui na função de *propensity score* e no teste da hipótese de equilíbrio entre as variáveis. Afim de aprimorar a qualidade dos

pareamentos, impõe-se a condição de suporte comum, como é sugerido por Becker e Ichino (2002).

A função $p(X_i)$ será estimada através do método paramétrico logit, pois considera-se, que a decisão da criança frequentar a pré-escola é o resultado de um processo de maximização de utilidade da família, como visto na seção anterior. Portanto, a variável latente do problema de pesquisa é dada pela diferença entre a utilidade que a família consegue ao matricular ou não a criança na pré-escola. Além disso, assume-se que os componentes aleatórios que definem o nível de utilidade de cada opção possuem uma distribuição independente de valor extremo de tipo 1. Logo a diferença entre esses componentes gera uma distribuição logística (CAMERON; TRIVEDI, 2005).

A função de *propensity score* a ser estimada é dada pela equação (14).

$$Inf_{2010i} = X'_{2010}\beta + \epsilon_i \quad (14)$$

Na equação (14) estão presentes os seguintes elementos: Inf_{2010i} é uma *dummy* que obtém valor 1 se a criança fez pré-escola e 0 caso contrário; X representa o vetor de características observadas associadas com a participação da criança na pré-escola, o qual inclui características da família, da criança, disponibilidade de escola e localização; ϵ_i é o termo de erro estocástico. Como mencionado, considera-se as variáveis decisórias contemporâneas à entrada na pré-escola, algo negligenciado pela literatura nacional⁷.

Para definir os indivíduos que fazem parte dos grupos tratados e controles, recorre-se aos métodos de pareamentos mais utilizados na literatura: vizinhos mais próximos e Kernel matching. De acordo com Becker e Ichino (2002), cada um desses métodos possuem diferentes níveis de *trade-offs* entre qualidade e quantidade de observações pareadas, porém a utilização conjunta deles pode garantir robustez à estratégia de identificação. O pareamento por vizinho mais próximo escolhe para cada indivíduo do grupo de tratamento um conjunto de k elementos que apresentem a menor distância entre os *propensity scores* (CAMERON; TRIVEDI, 2005). Já com o pareamento por Kernel, todas as unidades de tratamentos são pareadas com a média ponderadas das observações do grupo controle, com um peso que é inversamente proporcional à distância entre os *propensity scores* dos tratados e controles (BECKER; ICHINO, 2002).

⁷ As estimativas dos trabalhos de Curi e Menezes-Filho (2009) e Felício e Vasconcellos (2007) foram obtidas considerando apenas variáveis coletadas no momento da realização dos teste de proficiência. Desconsiderando, dessa forma, variáveis relativas à entrada na pré-escola.

Após a definição dos grupos, as coortes que fazem parte de cada grupo e que estejam presentes em 2015, permitem o cálculo do possível efeito da pré-escola na equação de interesse desta pesquisa.

$$A_{i2015} = X_{if2015}\beta + X_{is2015}\alpha + X_{it2015}\gamma + \omega Educ_i + \mu_i \quad (15)$$

A equação (15) esta composta por: A_{i2015} nota média da coorte i no teste do SAEB em matemática ou português; $X_f; X_s; X_t$ são os vetores de características das famílias e criança, das escolas e dos professores do coorte i respectivamente; μ_i é o erro idiossincrático ; e $Educ_i$ é a variável de interesse deste trabalho com característica binária, onde valores iguais a 1 indica que o coorte i frequentou a pré-escola, 0 caso contrário.

O parâmetro ω da equação (15) mostrará o efeito multiplicador das habilidades apresentado na seção anterior (γ). Este parâmetro indicará o valor conjunto do efeito direto do investimento na educação dos indivíduos (pré-escola) e indiretamente através da criação de um maior estoque de capital humano da criança que permitisse uma maior eficiência no aprendizado na escola. Por esse motivo, espera-se um sinal positivo e significativo para esta estimativa.

Para observar o efeito da pré-escola sobre a distribuição das notas do SAEB será utilizada a abordagem de efeitos do tratamento quantílico não condicional (*quantile treatment effect on the treated, QTT*). Esta abordagem, desenvolvida por Firpo (2007) e utilizada também por Andrews, Li e Lovenheim (2012), define que o efeito do tratamento no quantil é dada pela diferença entre os quantis da distribuição marginal dos resultados potenciais⁸. Logo, o QTT pode ser escrito como:

$$QTT = \Delta_{\tau|T=1} = q_{1,\tau|T=1} - q_{0,\tau|T=1} \quad (16)$$

Na equação (16) tem-se que: $q_{1,\tau|T=1}$ representa a distribuição marginal do quantil τ para o grupo de tratamento ($T=1$) e $q_{0,\tau|T=1}$ é a distribuição marginal para o grupo contrafactual. A estimação desta técnica é dada em duas etapas e sob duas suposições. Na primeira etapa é estimado o *propensity score*, permitindo encontrar o grupo contrafactual, e na segunda se realiza a diferença entres os quantis dos grupos (FIRPO, 2007). Para isso supõe-se que a seleção dos

⁸ Nota do aluno dado que frequentou ou não a pré-escola.

grupos se baseia nas características observáveis (seleção em observáveis) e que existe *rank permanence* - o tratamento não altera a posição dos indivíduos na distribuição das notas. A divisão dos quantis de notas será feita de acordo com a escala de proficiências apresentada pelo SAEB (INEP, 2018b, 2018c). Portanto a equação (16) pode ser reescrita da seguinte maneira:

$$QTT = \{\inf_q P[f(A_i; X^* = x_1^*; T = 1 \leq q) = \tau] - \left\{ \inf_q P[f(A_i; X^* = x_1^*; T = 0 \leq q) \geq \tau] \right\} \quad (17)$$

Onde X^* contém todos os vetores de controles identificados na equação (15).

3.2 Fontes de Dados

Os dados utilizados nesta pesquisa são provenientes de duas fontes. Para a primeira parte da estratégia, recorre-se ao Censo Demográfico de 2010 disponibilizado pelo Instituto Brasileiro de Geografia e Estatística (IBGE). A partir dele, foram considerados domicílios em que exista pelo menos uma criança com idade entre 4 e 6 anos. O motivo de utilizar estes dados deve-se ao fato de tentar controlar, em observáveis, a decisão da família de inserir a criança na pré-escola no momento em que o evento aconteceu.

Na segunda parte do trabalho, fez-se uso do banco de dados do SAEB de 2015 disponibilizado pelo Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira (INEP), em que se considera os alunos que fizeram o 5º ano em 2015, que correspondem às crianças ingressantes na educação infantil em 2010, sendo possível associar os dois bancos de dados. As variáveis que serão utilizadas no trabalho são apresentadas no Quadro 1.

A não existência da variável renda no banco de dados do SAEB gera uma limitação no trabalho, devido à renda familiar estar diretamente ligada ao estoque de capital humano dos filhos (PAXSON; SCHADY, 2007). Todavia, afim de superar essa limitação, recorre-se a uma variável proxy, construída com base na metodologia da Associação Brasileira de Empresas de Pesquisa (ABEP). A forma de cálculo é apresentada nas tabelas do ANEXO I. Este critério foi utilizado em pesquisa que vincula o trabalho infantil e o desempenho escolar (BEZERRA, 2006; BEZERRA; KASSOUF, 2006; BEZERRA; KASSOUF; ARENDS-KUENNING, 2009; ONO, 2015).

Quadro 1: Variáveis utilizadas nas estimações

Dimensões	Variáveis	Banco de dados.
Características das crianças	Sexo (feminino=1) Cor ou raça (branco=1) Idade Lê jornais (sim=1) Lê livros (sim=1) Lê revistas (sim=1) Tempo em trabalho doméstico Tempo acessando internet Costuma ir na biblioteca Trabalha fora de casa Faz dever de casa	Censo Demográfico 2010 e/ou SAEB 2015
Características das famílias	Localização (urbano=1) Idade da mãe Escolaridade da mãe Salário da mãe Escolaridade do pai Horas de trabalho do pai e da mãe Renda familiar per capita Região Município Computador (sim=1)	Censo Demográfico 2010 e/ou SAEB 2015
Características das escolas e professores	Lab. de informática Lab. de ciências Tipo de escola (pública=1) Oferta de educação infantil ⁹ Qualidade dos computadores Acesso à internet Biblioteca (sim=1) Sala de leitura Quadra de esporte Livros podem ser emprestados Sexo dos professores Idade dos professores Raça dos professores Escolaridade Licenciatura em matemática Licenciatura em letras Possui pós-graduação Salário dos professores Experiência como professor Professor lê jornais Professor lê livros Qual disciplina leciona	SAEB 2015
Variável de interesse	Pré-escola (sim=1)	Censo Demográfico 2010 e SAEB 2015
Variáveis dependentes	Notas nos testes de português (prof_p) e matemática (prof_mat)	SAEB 2015

⁹ É calculado dividindo o número de escolas pelo número de criança de 4 a 6 anos presentes município em 2010. O número de escolas por município foi obtido do Censo Escolar 2010

4 RESULTADOS E DISCUSSÕES

Seguindo a estratégia de identificação supracitada, esta seção apresenta os principais resultados da pesquisa. Inicia-se com uma análise das variáveis de interesse na subseção 4.1. A segunda subseção mostra os resultados relativos à criação dos grupos de tratamento e controle. Por fim, os resultados da relação de interesse são apontados na subseção 4.3.

4.1 Pré-escola e notas do SAEB

O número de pessoas que frequentam a pré-escola vem apresentando tendência crescente no decorrer dos anos. Curi e Menezes-Filho (2009) evidenciam esse comportamento ao mostrar que das pessoas nascidas na década de 1930 menos do 5% tinham frequentado a educação infantil, enquanto que das nascidas em 1974 essa porcentagem era superior ao 40%. Já Felício e Vasconcellos (2007) mostraram que do total de crianças em idade de frequentar a pré-escola, 71% o faziam em 2004, 11 pontos percentuais (p.p) a mais quando comparado com 1999. Em dados mais recentes, de acordo com a Figura 1, observa-se que o número de matrículas na educação infantil passou de 6,5 milhões em 2007 para 8,5 milhões em 2017. Especificamente na pré-escola as matrículas passaram de 4,9 milhões para 5,1 milhões no mesmo período.

O aumento na frequência na pré-escola está associado a fatores sociais e à mudança na estrutura familiar, como por exemplo, a lei N° 12.796 e o aumento da participação da mulher no mercado de trabalho. Almeida, Lima e Costa (2018) discutem que nas gerações mais recentes o número de crianças que crescem em famílias com mães inseridas no mercado de trabalho é maior em comparação às gerações anteriores. Esta mudança na estrutura familiar, relatada pelos autores, pode gerar uma maior demanda por educação infantil como forma de compensar a redução do tempo que a mãe se dedica ao cuidado dos filhos.

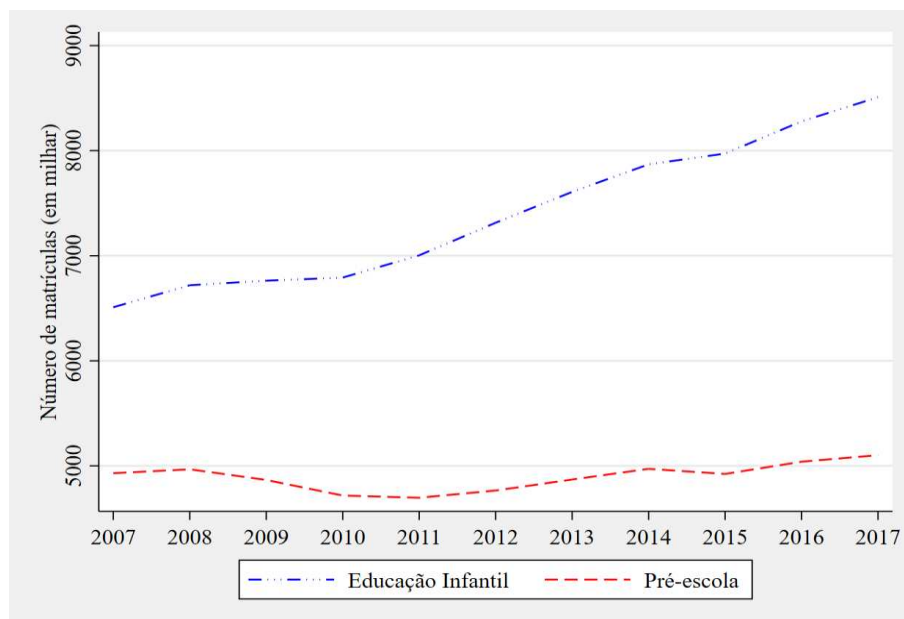


Figura 1: Número de matrículas na educação infantil no Brasil no período de 2007-2017.

Fonte: Elaborado pelo autor com base nos dados do INEP

Por outro lado, a lei N° 12.796, que torna obrigatória a frequência da criança na escola a partir dos 4 anos, pode explicar o crescimento da demanda por educação infantil após 2013. Esta lei encontra-se no contexto da Meta 1 do Plano Nacional de Educação (PNE), a qual teve como primeiro objetivo a universalização da pré-escola até 2016 (INEP, 2016). Nesse contexto, e de acordo com o relatório de educação de 2017, da Pesquisa Nacional por Amostra de Domicílio Contínua (PNADC/IBGE), a taxa de atendimento da pré-escola em 2017 era de 91,7%, tendo como destaque as regiões Sudeste e Nordeste, os quais alcançaram 93,8% e 94,8% respectivamente. Dos 8,3% das crianças de 4 e 5 anos que não frequentaram a pré-escola, 24% foi por indisponibilidade de vaga, 19% por falta de pré-escola na localidade ou porque fica muito distante e 43% foi porque os pais ou responsáveis não querem que frequente (IBGE, 2018). Este último dado permite inferir que a decisão da criança frequentar a pré-escola ainda depende altamente da decisão dos pais.

A evolução detalhada da taxa de atendimento da educação infantil é apresentada nos relatórios do monitoramento das metas do PNE. No primeiro relatório mostrou-se que no decênio de 2004-2014 a diferença entre crianças do sexo masculino e feminino que frequentaram a pré-escola permaneceu estável. Em 2004, 72,1% das meninas com idade entre 4 e 5 anos frequentavam a educação infantil, já no caso dos meninos com mesma idade essa porcentagem era 71,4%. Em 2014, essas participações aumentaram 18.1 p.p. para o caso das meninas e 17.7 p.p para os meninos, alcançando o total de 90,2% e 89,1% respectivamente (INEP, 2016).

Quando se analisa a frequência das crianças na pré-escola por área de residência, observa-se que na área urbana a porcentagem de frequência é maior que na área rural. Porém, essa diferença foi reduzida no período de 2004-2016, passando de 20,7 p.p. em 2004 para 3,2 p.p. em 2016. Esse mesmo comportamento é observado ao comparar crianças pertencentes aos grupos de 25% mais pobres e 25% mais ricos. Para o grupo dos 25% mais pobres o indicador de frequência chegou a 89,1% em 2016, evidenciando um crescimento de 26,2 p.p. em relação a 2004. Enquanto que para o grupo dos 25% mais ricos essa variação foi de 5 p.p no mesmo período, alcançando 96,4% em 2016 (INEP, 2018d).

A tendência de crescimento também é observada nos escores de proficiências, quer em português, quer em matemática, dos alunos do 5 ano do ensino fundamental. Na Figura 2 observa-se que a proficiência média brasileira em português passou de 165 pontos em 2001 para 215 em 2017, mostrando uma melhora em média de 50 pontos. Já em matemática, a proficiência média passou de 176 pontos para 224 no mesmo período.

É importante lembrar que o trabalho foca nos testes SAEB de 2015 por dois motivos principais: permite o cruzamento com os dados do Censo populacional de 2010 e é último conjunto de microdados disponibilizado pelo INEP. Logo, a descrição dos dados será referente aos resultados do SAEB 2015.

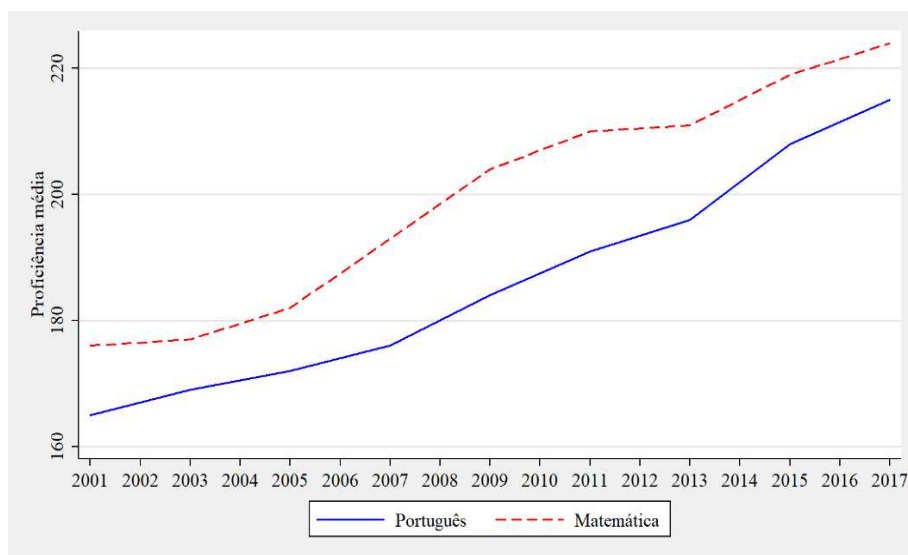


Figura 2: Proficiência média dos alunos do 5 ano do ensino fundamental em português e matemática no período de 2001 a 2017.

Fonte: Elaborado pelo autor com base nos dados do INEP

Em 2015, a proficiência média dos alunos nas disciplinas de matemática e português foi de 215 e 208 pontos, respectivamente. Referente ao teste de português, os estados que tiveram os melhores desempenho foram Santa Catarina (SC), São Paulo (SP), Paraná (PR) e Minas Gerais (MG), tal qual se observa na Figura 3. Nota-se também que as menores médias de desempenho foram para os estados de Maranhão (MA), Amapá (AP) e Alagoas (AL).

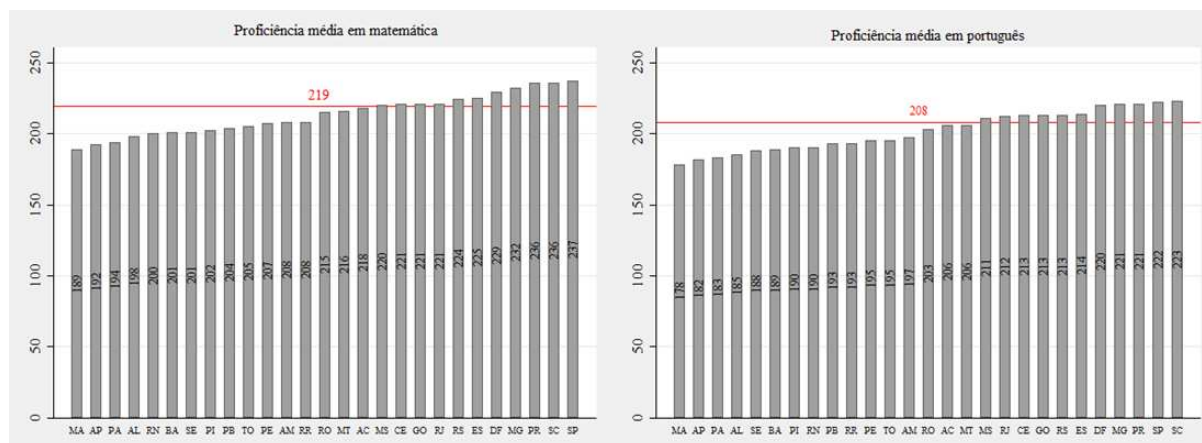


Figura 3: Proficiência média dos alunos do 5º ano do ensino fundamental nos testes do SAEB em 2015.

Fonte: Elaborado pelo autor com base nos dados do INEP

De acordo com os microdados do SAEB, ao observar de perto o desempenho dos alunos do 5º ano em português, tem-se que 46,8% tiveram um desempenho insuficiente¹⁰. Esta porcentagem encontra-se acima dos 50% nos estados do Norte e Nordeste, sendo os estados do Maranhão (Nordeste) e Amapá (Norte) aqueles que tiveram a maior porcentagem de alunos com desempenho insuficiente. Em contraste, nas regiões Sul e Sudeste apenas 36% dos alunos obtiveram um desempenho insuficiente, enquanto que mais de 10% dos alunos dessas regiões alcançaram um desempenho adequado, principalmente nos estados de Santa Catarina (Sul), São Paulo (Sudeste) e Minas Gerais (Sudeste).

No que se diz respeito à matemática e conforme a Figura 3, os estados de São Paulo, Santa Catarina, Paraná e Minas Gerais obtiveram desempenho superiores à média nacional. Estes estados fazem parte das regiões Sul e Sudeste, onde apenas 26,3% e 27,8% dos alunos do 5º ano tiveram desempenho insuficiente em matemática, respectivamente. Ao mesmo tempo, 17,58% dos alunos do Sudeste e 15,11% dos alunos do Sul lograram um desempenho adequado, mostrando que a proficiência média dessas regiões está bem acima da média nacional.

¹⁰ De acordo com as escalas do SAEB um desempenho abaixo do nível 4 é considerado como insuficiente; entre o nível 4 e 6 considera-se como básico; e para níveis maiores e iguais a 7 considera-se como desempenho adequado. Ver< http://download.inep.gov.br/educacao_basica/saeb/2018/documentos/presskit_saeb2017.pdf>

Em contrapartida, os estados de Maranhão, Amapá e Pará são os que possuem médias de desempenho aquém da média nacional em matemática. Na Figura 3, é possível notar que os três estados tiveram desempenho inferiores tanto em português e matemática. Além disso, ao observar o tipo desempenho atingindo pelos alunos, constatou-se que nas regiões, onde esses estados estão localizados, a porcentagem de alunos com desempenho insuficiente foi acima dos 50%; na região Norte foi 54% e na região Nordeste foi 55,1%. Nessas mesmas regiões menos dos 6% dos alunos alcançaram um desempenho adequado.

Do mesmo modo, as proficiências médias em português e matemática dos alunos do 5º ano do ensino fundamental variam conforme a dependência administrativa das escolas. A média de proficiência das escolas particulares é maior tanto em português como matemática quando comparadas com as escolas estaduais e municipais. Em matemática, por exemplo, a média das escolas particulares foi 244 pontos, enquanto que as escolas estaduais obtiveram 223 pontos e as municipais 215 pontos. Ao comparar essas médias com os resultados do SAEB 2013, observa-se que houve em aumento de 42 pontos para as escolas estaduais e municipais, e 17 pontos para as escolas particulares. Já na disciplina de português, as escolas privadas tiveram um acréscimo de 23 pontos, atingindo a média de 235 pontos em 2015. As escolas estaduais e municipais, por sua vez, evidenciaram uma melhora de 39 pontos em comparação com 2013, logrando uma média de 211 e 203 pontos respectivamente. (INEP, 2018e)

Identifica-se, conforme os dados apresentados, que há uma tendência comum de crescimento da taxa de atendimento da pré-escola e das notas de proficiência em português e matemática nos testes do SAEB. Por sua vez, na Figura 4, constata-se que a distribuição de notas em português e matemática dos alunos que frequentaram a pré-escola apresentam uma ligeira assimetria para a direita. Um teste¹¹ de média das proficiências entre os dois grupos revela que há uma diferença estatística nas duas disciplinas. Os alunos que frequentaram a pré-escola obtiveram 210 pontos em português e 222 pontos em matemática, enquanto que os alunos que não frequentaram esta etapa de ensino lograram 203 e 215 pontos respectivamente, esta diferença observada foi significativa a 1%. Estas informações fornecem os primeiros indícios da possível correlação entre essas variáveis.

¹¹ Foi utilizado comando `ttest` do software Stata.

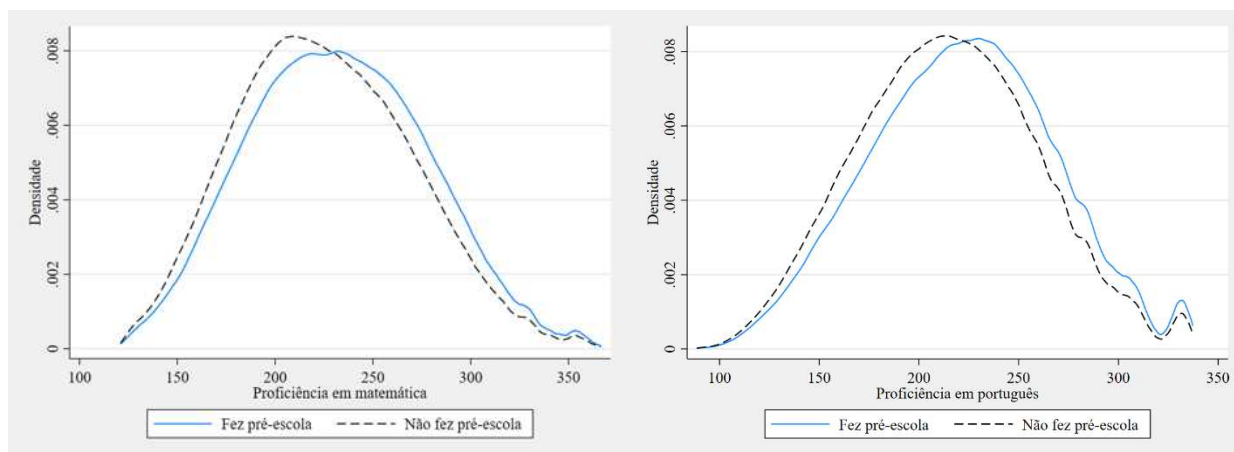


Figura 4: Distribuição de densidade das notas de português e matemática dos alunos do 5 ano do ensino fundamental em 2015.

Fonte: Elaborado pelo autor com base nos dados do INEP.

4.2 Construção dos grupos de tratamento e controle

Nesta subseção, apresenta-se os resultados do processo de definição do grupo de controle. Na Tabela 1, exibe-se os coeficientes estimados pelo modelo *logit* considerando como variável dependente a participação ou não da criança na pré-escola. Esta estimação foi realizada somente com o intuito de definir os grupos, portanto os coeficientes serão analisados com uma correlação entre variáveis. O modelo mostrou significância conjunta a 1%, de acordo com a estatística de razão de verossimilhança (LR), indicando que as variáveis escolhidas são relevantes para explicar o processo decisório da família em matricular ou não uma criança na pré-escola.

Nota-se, na Tabela 1, que o fato da criança ser do sexo feminino aumenta a probabilidade de frequentar a pré-escola. Este sinal era esperado, já que existe uma porcentagem maior de meninas matriculas nesta etapa de ensino. Por outro lado, a variável idade mostrou-se estatisticamente significativa a 1% e com sinal negativo, indicando que quanto maior a idade da criança menor é a probabilidade de estar no grupo tratado. O aumento da idade do indivíduo significa que ele está cada vez mais perto de terminar o ciclo da educação infantil e começar o ensino fundamental.

Tabela 1: Resultados do Logit para definir os grupos controle e tratado em 2010.

Variáveis	Coefficientes
Sexo	0.00979* (0.00567)
Idade	-0.52398*** (0.01095)
Cor ou raça	0.05371*** (0.00762)
Localização	0.23354*** (0.01358)
Rendimento per capita familiar	0.00004*** (0.00001)
Número de pessoas na família	-0.08362*** (0.00303)
Escolaridade da mãe	0.03404*** (0.00185)
Horas de trabalho da mãe fora de casa, por semana ¹²	0.00142*** (0.00017)
Idade da mãe	0.00339*** (0.00035)
Escolaridade do pai	0.01043*** (0.00171)
Horas de trabalho do pai fora de casa, por semana ¹²	0.00032* (0.00018)
Oferta de escolas de educação infantil	0.00269 (0.00810)
Famílias monoparentais	0.07062* (0.03743)
Norte	-0.02362 (0.04329)
Nordeste	0.38021*** (0.03504)
Sudeste	0.35821*** (0.03754)
Sul	-0.05090 (0.04097)
Constant	1.82603*** (0.07448)
LR chi2	6721.01
Prob>chi2	0.00***
PseudoR2	0.0475

Fonte: Resultados da pesquisa

Nota: Erros padrão em parênteses. * $p < .1$, ** $p < .05$, *** $p < .01$. Erros padrão clusterizado por município.¹² Número de horas trabalhadas no trabalho principal.

Com relação à localização, crianças da região Nordeste e Sudeste são as mais propensas a frequentar a escola quando comparadas com as crianças do Centro-Oeste. Essas duas regiões, que se mostraram estatisticamente significativas, apresentam as maiores taxas de atendimento da pré-escola do país. Quando se considera a diferença entre o urbano e rural¹³, foi observado que o fato das crianças residirem em uma área urbana está associado positivamente com a variável dependente.

As variáveis relativas à mãe da criança apresentaram os sinais esperados. A quantidade de horas que a mãe trabalha fora de casa está correlacionada positivamente com a probabilidade de a criança frequentar a pré-escola. A maior inserção das mulheres no mercado de trabalho se traduz na redução do tempo utilizado para cuidado dos filhos (AQUINO; PAZELLO, 2011), gerando o fenômeno relatado por Almeida, Lima e Costa (2018). Nesse sentido, a educação infantil pode ser uma alternativa de educação e cuidados dos filhos para as mães que estão inseridas no mercado de trabalho, especialmente para aquelas de famílias monoparentais. O *trade-off* entre tempo e renda, ao qual a mãe se depara no momento de decidir entrar no mercado de trabalho, torna-se mais fácil quando há um maior número de pessoas na família. Uma pessoa a mais na família aumenta a probabilidade de alguém substituir o tempo que mãe deixou de despende no cuidado dos filhos permitindo que ela se insira no mercado de trabalho (LOKSHIN; GLINSKAYA; GARCIA, 2000). Ao mesmo tempo, essa característica da família também diminui a probabilidade de a criança frequentar a escola.

Ainda na Tabela 1, observa-se que o aumento da probabilidade da criança frequentar a pré-escola está correlacionada com a escolaridade dos pais. Além disso, nota-se também que a renda familiar se relaciona de forma positiva a essa probabilidade. Mudanças na riqueza familiar aumenta a demanda por escolaridade dos filhos (GLEWWE; JACOBY, 2004). O que chama a atenção no coeficiente encontrado para a renda familiar é o seu valor pequeno, contudo, não é algo inesperado para o caso brasileiro. Como foi mostrado na subseção anterior, a diferença do número de crianças que frequentam a pré-escola entre os 25% mais ricos e os 25% mais pobres vem diminuindo rapidamente.

Para verificar a hipótese de balanceamento do método de *propensity score matching*, apresenta-se a distribuição da densidade da variável índice (*propensity score*) na Figura 5. O algoritmo utilizado para o pareamento foi: 5 vizinhos mais próximos. Resultados análogos são encontrados com o algoritmo de Kernel e são apresentados no Apêndice B e C. Na figura

¹³ Neste trabalho adota-se a diferenciação entre urbano e rural feito pelo Censo Demográfico.

observa-se uma sobreposição das densidades do *propensity score* dos grupos após o pareamento, o que evidencia balanceamento satisfatório.

Além disso, no Apêndice A mostra-se mais uma evidência sobre a qualidade do pareamento entre os grupos. Embora a média de algumas variáveis de controle não são estatisticamente iguais entre os grupos após o pareamento, a estatística R^{14} (0,64) e B^{15} (7,5%) de Rubin indicam um pareamento satisfatório. Rubin (2001) recomenda que a estatística R esteja entre 0.5 e 2, e que a B seja inferior a 25% para ter um balanceamento confiável e adequado.

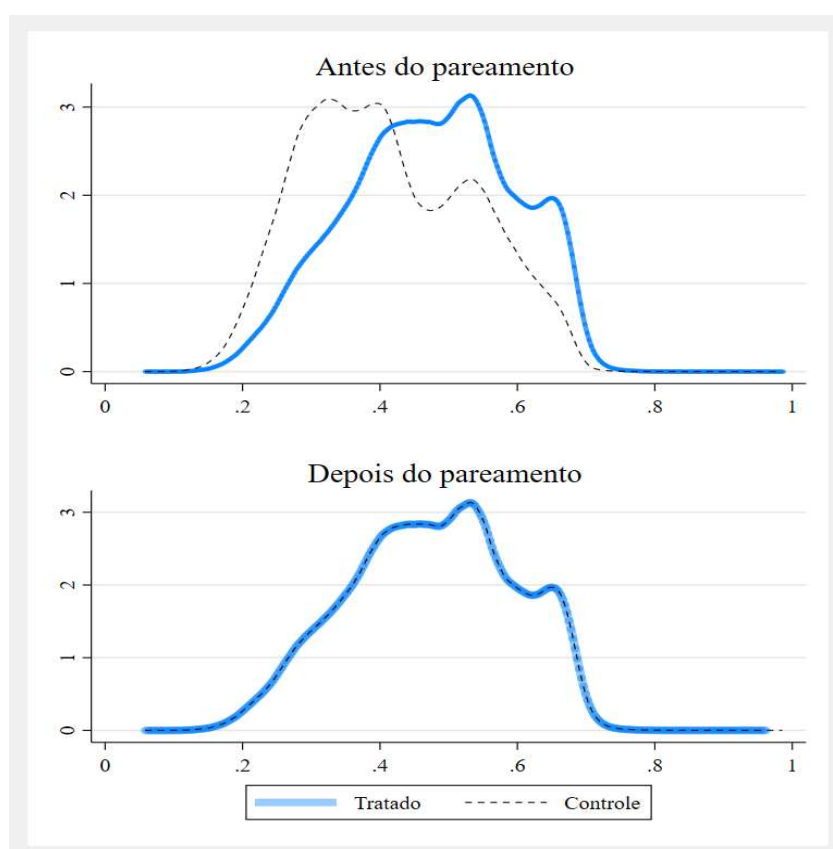


Figura 5: Distribuição do *propensity score* estimados antes e depois do pareamento.

Fonte: Resultado da pesquisa.

No processo de pareamento foram consideradas todas as observações que estavam dentro do suporte comum e com base nisso definiu-se os coortes. Nesse sentido, dado as variáveis: sexo, raça, ano de nascimento, UF de residência e frequenta ou não pré-escola, obteve-se 648 coortes na amostra. Do total de coortes definidos, 324 fazem parte do grupo controle e 324 do grupo tratado. Dadas as características da amostra, a relevância deste processo está na obtenção

¹⁴ Mostra a razão entre as variâncias do *propensity score* dos grupos após o pareamento.

¹⁵ Mostra a diferença bruta e padronizada das médias do *propensity score* entre os grupos após o pareamento.

dos pesos para as unidades de controle. Estes pesos foram utilizados nas estimações da seção subsequente.

4.3 Resultados das estimativas da relação pré-escola e desempenho acadêmico.

Nesta subseção apresenta-se os resultados principais da estratégia de identificação obtidos tanto por MQO quanto por Regressão Quantílica não Condicional (UQR), para a variável de interesse. Inicia-se mostrando a existência de uma correlação positiva entre pré-escola e o desempenho dos alunos do 5º ano de ensino fundamental nos testes de português e matemática.

Na Tabela 2, verifica-se que ter frequentado a pré-escola aumenta o desempenho médio no teste de matemática. Todavia, a partir desses resultados, não se pode deduzir uma relação de causalidade entre as variáveis uma vez que se reconhece a presença de endogeneidade nestes modelos. Um sinal deste problema é a mudança dos coeficientes de interesse entre os modelos conforme se adicionam controles. Esta variação confirma que há alguma variável não observada presente no termo de erro que pode estar influenciando o valor do coeficiente da pré-escola provocando a superestimação, por exemplo, habilidade inatas, motivação e esforço.

Tabela 2: Coeficientes estimados por MQO para as notas de matemática dos alunos do 5º ano de ensino fundamental em 2015, sem pareamento

	(1)	(2)	(3)	(4)
Pré-escola	7.11305*** (0.08181)	6.57173*** (0.15437)	6.49675*** (0.15883)	6.51137*** (0.16370)
Constante	225.454*** (0.05039)	152.80921*** (1.81695)	120.90044*** (2.11107)	113.2855*** (2.17756)
Controles (características)				
Indivíduo e da família	Não	Sim	Sim	Sim
Professores	Não	Não	Sim	Sim
Escola	Não	Não	Não	Sim
R ²	0.00588	0.09953	0.11662	0.1288
BIC	13400000	3355009	3105541	2880516
Estatística F	7560.07***	1990.46***	1359.56***	1146.26***
Prob. > F	0.00000	0.00000	0.00000	0.00000

Fonte: Resultados da pesquisa

Nota: Erros padrão em parênteses. * $p < .1$, ** $p < .05$, *** $p < .01$. Erros padrão clusterizado por municípios.

Para o teste de português, obteve-se também o sinal positivo entre a pré-escola e o desempenho do aluno. Na Tabela 3, verifica-se que, mesmo após a utilização todos os controles, os resultados do possível efeito da pré-escola entre as duas disciplinas analisadas são similares. Infere-se, portanto, que a educação na pré-escola estabelece os alicerces para a construção de

futuros conhecimentos nos primeiros anos do ensino fundamental. O estabelecimento deste alicerce pode estar diretamente relacionado com a possibilidade de desenvolver habilidades cognitivas e principalmente as não cognitivas como prestar atenção, seguir regras, controlar impulso e disciplina (CENTER ON THE DEVELOPING CHILD, 2016), as quais podem afetar igualmente a aquisição de conhecimentos nas duas áreas analisadas.

Tabela 3: Coeficientes estimados por MQO para as notas de português dos alunos do 5 ano de ensino fundamental em 2015, sem pareamento

	(6)	(7)	(8)	(9)
Pré-escola	7.21409*** (0.08229)	6.43105*** (0.15416)	6.36471*** (0.15914)	6.46401*** (0.16432)
Constante	214.56420*** (0.05069)	112.08657*** (1.81222)	85.62702*** (2.10404)	79.48739*** (2.18327)
Controles (características)				
Indivíduo e da família	Não	Sim	Sim	Sim
Professores	Não	Não	Sim	Sim
Escola	Não	Não	Não	Sim
R ²	0.00597	0.11417	0.12469	0.13264
BIC	13400000	3342817	3096084	2872697
Estatística F	7685.96***	2313.97***	1524.42***	1181.86***
Prob. > F	0.00000	0.00000	0.00000	0.00000

Fonte: Resultados da pesquisa

Nota: Erros padrão em parênteses. * $p < .1$, ** $p < .05$, *** $p < .01$. Erros padrão clusterizado por municípios.

Contudo, a partir dos resultados apresentados nas tabelas anteriores, não é possível indicar que os valores encontrados sejam os efeitos da pré-escola, mesmo controlando pelas características da criança, da família, dos professores e das escolas. Estas variáveis pouco ou nada se relacionam à decisão da criança frequentar a pré-escola, portanto, é necessário controlar esta decisão com variáveis contemporâneas a ela afim de tornar os indivíduos idênticos e comparáveis entre si. Nesse sentido, aplicou-se a estratégia de identificação proposta no trabalho e os resultados são expostos na Tabela 4. Após definição dos grupos de controle e tratado, através do pareamento pelo algoritmo de 5 vizinhos mais próximos, e de ter regredido as variáveis de interesses utilizando diversos controles e os pesos, obteve-se uma relação positiva entre as variáveis de interesses, porém com uma redução de aproximadamente 2 pontos nos coeficientes da pré-escolas nos modelos da Tabela 4.

Os dois pontos observados entre as estimativas podem estar atrelados a alguma variável de confusão das famílias e das crianças. Por exemplo, a percepção da família com relação aos retornos futuros da educação e a efetividade dos investimentos na primeira infância que geram

viés para cima nas estimativas (AKRESH et al., 2012; GARCES; THOMAS; CURRIE, 2002). Além disso, ao negligenciar as habilidades inatas ou predisposição genética dos indivíduos gera-se também o mesmo viés para cima, uma vez que eles podem potencializar os ganhos da primeira infância. Nessa perspectiva, a redução encontrada com o método proposto era esperada.

Resultados positivos para esta relação também foram encontrados por Curi e Menezes-Filho (2009) e, Felício e Vasconcellos (2007). Curi e Menezes-Filho (2009) basearam sua análise em dados provenientes do SAEB de 2003¹⁶ e encontraram que a educação infantil influencia a proficiência média de matemática dos alunos da 4 série do ensino fundamental em 7,5%. Para os testes de proficiência em matemática, referentes aos anos de 2003 e 2005, Felício e Vasconcellos encontraram que o efeito da pré-escola era de 6.3% da proficiência. Observa-se então uma diferença entre esses coeficientes e os coeficientes expostos na tabela 4.

Parte significativa da diferença entre os coeficientes obtidos neste estudos e aqueles encontrados nos dois trabalhos anteriores é devido às estratégias identificação utilizadas e às mudanças sociais. Por exemplo: em fevereiro de 2006 implementou-se a lei N° 11.274, a qual estabelece o aumento de anos de estudos do ensino fundamental de 8 para 9 anos, antecipando assim a entrada das crianças no ensino fundamental (BRASIL, 2006); ou seja, o que hoje se conhece como o primeiro ano do ensino fundamental, antes da lei fazia parte da educação infantil. A partir disso, constata-se uma diferença amostral entre este trabalho e os trabalhos tanto de Curi e Menezes-Filhos (2009) e de Felício e Vasconcellos (2007).

Além disso, houve uma variação de 25 p.p. da taxa de atendimento da educação infantil entre 2000 e 2010. A taxa de atendimento da educação infantil passou de 60% em 2000 para aproximadamente 85% em 2010 (DE FELÍCIO; VASCONCELLOS, 2007; INEP, 2018d). Este aumento da participação das crianças na pré-escola conjuntamente com os efeitos da lei N° 11,274 podem explicar a diferença entre os coeficientes encontrados nos trabalhos anteriores realizados por Curi e Menezes-Filho (2009), Felício e Vasconcelos (2007) e aqueles apresentados nas estimativas 1 da Tabela 2 e 6 da Tabela 3. Estas estimativas mostram o máximo valor que os coeficientes da relação entre a frequência na pré-escola e as notas de proficiência pode adotar em 2015, quando não se controla de nenhuma forma a endogeneidade presente.

¹⁶ Em 2003, a proficiência média de matemática para os alunos da 4 série do ensino fundamental era de 177 pontos.

Tabela 4: Coeficientes estimados por MQO para as notas de matemática e português dos alunos do 5 ano de ensino fundamental em 2015 após a aplicação da estratégia de identificação

Variáveis	Português			Matemática		
	(10)	(11)	(12)	(13)	(14)	(15)
Pré-escola	5.9647*** (0.7927)	4.2432*** (0.9463)	4.1771*** (0.9213)	6.5493*** (0.734)	4.6305*** (1.2431)	4.4545*** (0.8439)
Constante	208.3459*** (2.4561)	-98.9217 (77.5330)	-64.7159 (70.6854)	218.0199*** (4.93)	-90.9294 (78.671)	-58.7542 (72.022)
Controles (características)						
Indivíduo e da família	Não	Sim	Sim	Não	Sim	Sim
Professores	Não	Sim	Sim	Não	Sim	Sim
Escola	Não	Não	Sim	Não	Não	Sim
R ²	0.0347	0.6722	0.7287	0.0333	0.6517	0.7340
BIC	5488.627	4874.918	4785.783	5551.046	4945.213	4818.903
Estatística F	20.26***	41.33***	42.92***	22.45***	40.49***	40.75***
Prob.>F	0.00001	0.00000	0.00000	0.00000	0.00000	0.00000

Fonte: Resultados da pesquisa

Nota: Erros padrão em parênteses. * $p < .1$, ** $p < .05$, *** $p < .01$. Foi utilizado o peso obtido no pareamento (5 vizinhos). Erros padrão clusterizado por UF e estimados por *bootstrap*.

No quesito da estratégia de identificação, os trabalhos brasileiros mencionados controlam a participação da criança na pré-escola apenas com variáveis contemporâneas à aplicação do teste de proficiência. Este tipo de estratégia requer uma hipótese forte de invariância das características familiares entre ambos os eventos. Contudo, Todd e Wolpin (2003, 2004) mostram que não há suporte para este tipo de modelos restritos, pois variáveis como renda, de acordo com Parish e Willis (1993), apresentam tendência crescente através do ciclo de vida dos indivíduos e das famílias. Aliás, conforme foi apresentado na seção 4.2, existem variáveis relevantes na decisão da criança frequentar a pré-escola consideradas neste trabalho, mas que foram negligenciadas pelos trabalhos predecessores, como por exemplo, número de horas de trabalho da mãe e do pai fora de casa e, número de pessoas na família.

Outra diferença nas estratégias aplicadas diz respeito à composição da amostra, neste trabalho exclui-se da amostra os indivíduos que entraram na escola depois do primeiro ano do ensino fundamental. Tanto Curi e Menezes-Filho (2009) como Felício e Vasconcellos (2007) incluíram no grupo de controle indivíduos que informaram ter começado a escola na 2ª série ou 3ª série, tratando-os com o mesmo peso daqueles que entraram na primeira série. Com os dados do SAEB, nada pode ser inferido sobre o porquê desses indivíduos terem começado a escola depois do primeiro ano. Porém, é possível inferir que os indivíduos que ingressaram nas escolas depois apresentaram uma defasagem de conhecimento em relação ao demais alunos que começaram a educação básica no primeiro ano do fundamental, o que tem como consequência menores rendimentos nos testes¹⁷. A proficiência média destes alunos em 2015 foi de 183 em português e 197 em matemática. Estes resultados foram aquém da média nacional em ambas disciplinas. Logo, as estimativas apresentadas pelos autores podem estar superestimando os efeitos da pré-escola.

Na literatura internacional, os efeitos das intervenções na primeira infância também são positivos e variam dependendo das características dos programas e do público alvo. Cada programa ou política possui peculiaridades que não permitem uma comparação direta com os resultados encontrados aqui. Por exemplo: os programas *Perry Preschool Program*, *The Carolina Abecedarian* e *Head Start* são direcionados às crianças menos favorecidas (CAMPBELL et al., 2002; GARCES; THOMAS; CURRIE, 2002). Contudo, os coeficientes mostrados na Tabela 4 se encontram próximos do intervalo definido na meta análise dos efeitos das intervenções na primeira infância de Camilli et al. (2010). Ao analisar 123 estudos, os

¹⁷ Nos dados do SAEB 2015 somente 4% dos alunos declaram ter começado a escola depois do primeiro ano do ensino fundamental e a proficiência média desse grupo foi menor que a média brasileira.

autores encontraram que os efeitos médios destas intervenções sobre os resultados cognitivos (teste) variam entre 0.12 e 0.48 do desvio padrão das notas. Os resultados encontrados na presente pesquisa mostram que os coeficientes da pré-escola sobre os testes representam 0.1 do desvio padrão¹⁸.

Nota-se, na Tabela 4, que os coeficientes da pré-escola em relação às proficiências em português e matemática são positivos e significativos a 1%. Embora não seja possível definir o valor exato desse efeito, isto confirma a existência de uma possível relação positiva entre as variáveis. Resultados similares aos apresentados na Tabela 4 são encontrados utilizando o pareamento por Kernel (ver no Apêndice E). A similaridade entre os resultados encontrados pelos métodos dá robustez às estimativas obtidas pela estratégia proposta.

Dado que cada teste de proficiência captura diferentes tipos de habilidades, a partir dos resultados idênticos entre as disciplinas avaliadas, é possível deduzir que um dos caminhos pelo qual a pré-escola afeta o desempenho dos alunos é através do desenvolvimento de habilidades não cognitivas. Berslinski, Galiani e Gertler (2009) mostram que a educação infantil tem efeitos sobre o comportamento da criança na sala de aula, em séries posteriores, especificamente nas variáveis como esforço, disciplina, prestar atenção e participação regular das aulas.

No trabalho desenvolvido para o Brasil por Curi e Menezes-Filho (2009), observa-se que os possíveis efeitos da educação infantil diminuem conforme o aluno passa da 4ª série do ensino fundamental para o 3º ano do ensino médio. Conforme a idade do indivíduo aumenta, a demanda e/ou investimento em educação deixa de ser realizada pelos pais (AKRESH et al., 2012) e passa a ser apenas uma decisão do indivíduo, logo as habilidades não cognitivas como a motivação e esforço adquirem relevância. Sendo assim, maior nível de escolaridade dos indivíduos tenderia a reduzir as diferenças não observadas entre eles. Portanto, a dissipação do coeficiente de relação entre educação infantil e desempenho escolar, de acordo com Garces, Thomas e Currie (2002), não significa necessariamente que o efeito seja temporário, senão que a educação infantil permitiu o desenvolvimento de um conjunto maior de habilidades não cognitivas.

As habilidades não cognitivas apresentam efeitos sobre os resultados futuros dos indivíduos (ALMLUND et al., 2011; CARNEIRO; CUNHA; HECKMAN, 2004; CARNEIRO; HECKMAN, 2003; GARCES; THOMAS; CURRIE, 2002; HANUSHEK, 2013; HECKMAN, 2006, 2007; HECKMAN; PINTO; SAVELYEV, 2013). Considerando isso, infere-se que a pré-

¹⁸ O desvio padrão das notas de proficiência em português e matemática é de 46 e 44 pontos respectivamente.

escola teria efeitos na vida adulta dos indivíduos de duas maneiras. A primeira seria através das habilidades não cognitivas e segunda seria por meio da melhora do desempenho acadêmico dos alunos e aumento dos anos de estudos. Curi e Menezes-Filhos (2009, 2014) apresentam uma correlação positiva entre educação infantil e a probabilidade de terminar diferentes etapas de ensino, além de uma relação positiva entre notas do SAEB e o salário futuro no Brasil.

Além de definir a relação positiva entre a pré-escola e os testes de proficiência, foi procurado também verificar o comportamento dessa relação ao longo da distribuição das notas. A partir disso, será possível entender que tipo de habilidade e conhecimento foram desenvolvidos pelos estudantes de acordo com posicionamento da sua nota alcançada na escala do SAEB. Notas altas exigem que os alunos respondam questões com maior nível de dificuldade, ou seja, requerem que o indivíduo possua maiores níveis de conhecimento e de habilidades cognitivas e não cognitivas (INEP, 2018e). Na Tabela 5, observa-se que as notas, tanto em português quando em matemática, variam dentro e entre os grupos nos diversos quantis definidos, sinalizando assim mudanças nos coeficientes da pré-escola ao longo da distribuição.

Ainda de acordo com a Tabela 5, observa-se que as notas médias em português e matemática do quantil 10 são maiores para as crianças que declararam ter frequentado a pré-escola. Esta diferença é mantida ao longo da distribuição e cresce conforme os quantis aumentam, principalmente nas notas de matemática. Porém, a simples diferença entre os valores de cada quantil não forneceria o efeito da pré-escola.

Tabela 5: Notas do SAEB 2015 dos alunos do 5 ano do ensino fundamental nos quantis 10, 25, 50, 75 e 90 da distribuição considerando como variável categoria a pré-escola.

Proficiência em português					
Pré-escola	Q10	Q25	Q50	Q75	Q90
Não	155.84	182.43	214.09	245.59	272.75
Sim	161.05	189.52	222.36	253.61	280.93
Proficiência em Matemática					
Não	168.61	192.55	223.01	256.49	285.21
Sim	173.67	199.17	231.45	264.67	292.65

Fonte: Resultados da pesquisa

Desse modo, aplica-se a estratégia proposta conjuntamente com a estimação por regressão quantílica não condicionada¹⁹ para calcular os coeficientes da pré-escola. Nas estimativas apresentadas nas Tabela 6 e 7, foram utilizados os coortes definidos anteriormente e os pesos

¹⁹ Para encontrar as estimativas foi utilizado o comando `rifreg` do stata.

encontrados pelo algoritmo de 5 vizinhos. Em todos os modelos foram inseridos os controles relativos às características da criança, da família, dos professores e das escolas, com a finalidade de obter estimativas robustas nos modelos.

Verifica-se que o coeficiente da variável de interesse aumenta conforme os quantis de nota aumentam²⁰. O coeficiente da pré-escola passa de 2.68 no quantil Q10 para 6.88 no quantil Q90 nas notas de matemática (Tabela 6). Para a disciplina de português (Tabela 7), o mesmo comportamento é observado, o coeficiente varia de 1.84 para 6.47. Porém, somente os coeficientes dos quantis Q25, Q50, Q75 e Q90 foram significativos para ambas as disciplinas.

A não significância da pré-escola para as notas dos testes no quantil inferior das notas era algo esperado. O primeiro quantil representa os níveis 0, 1 e 2 da escala do SAEB. Os alunos que obtiveram notas dentro desses níveis foram classificados com desempenho insuficiente. Isto significa que não apresentam as competências elementares em português e matemática que são esperadas dos alunos do 5º ano do ensino fundamental (nível 0) ou são capazes de identificar o tema explícito em contos e de localizar informações relativas a valores maior ou menor dentro de uma tabela ou gráfico (nível 1 e 2). Assim sendo, esperava-se que a pré-escola tivesse um efeito pequeno ou nulo neste quantil, porque a frequência da criança nesta etapa de ensino permitiria o desenvolvimento de habilidades cognitivas e não cognitivas as quais potencializariam a aprendizagem de novos conhecimentos nas etapas de ensino subsequentes.

No quantil 90, a pré-escola mostra um coeficiente positivo e significativo a 1% para ambas as disciplinas. Este coeficiente, além de ser maior quando comparados com os outros quantis, reflete a relevância da pré-escola na aquisição de conhecimentos e no desenvolvimento de habilidades capazes de influenciar no desempenho adequado nos testes de proficiência. Os alunos que obtiveram este nível de desempenho lograram notas maiores iguais ao do nível 7 na escala do SAEB. Conforme as características dos testes, notas nos níveis superiores demonstram que os alunos desenvolveram as habilidades propostas pela Base Nacional Comum Curricular, resumidas nas escalas do SAEB (INEP, 2018e; MEC, 2018)

²⁰ Foi realizado também um teste de diferenças estatísticas entre os coeficientes das regressões quantílicas.

Tabela 6: Coeficientes da regressão quantílica não condicional para a nota de matemática dos alunos do 5 ano do ensino fundamental em 2015.

	MQO	Q10	Q25	Q50	Q75	Q90
Pré-escola	4.4545*** (0.8439)	2.67780 (2.07420)	3.02310* (1.79107)	6.25993*** (1.67857)	4.72966*** (1.34725)	6.87882*** (2.39933)
Constante	-58.7542 (72.022)	-112.3791 (108.91243)	-7.61180 (74.09004)	4.14011 (76.64336)	-3.66060 (57.83408)	-36.38271 (102.33573)
Controles (características)						
Indivíduos e família	Sim	Sim	Sim	Sim	Sim	Sim
Professores	Sim	Sim	Sim	Sim	Sim	Sim
Escolas	Sim	Sim	Sim	Sim	Sim	Sim
R ²	0.7340	0.39775	0.44679	0.57980	0.50938	0.30634
Estatística F	40.75***	5.69	17.50	48.68	22.31	3.31
Prob.>F	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000

Tabela 7: Coeficientes da regressão quantílica não condicional para a nota de português dos alunos do 5 ano do ensino fundamental em 2015

	MQO	Q10	Q25	Q50	Q75	Q90
Pré-escola	4.1771*** (0.9213)	1.83453 (2.14534)	4.55524*** (1.50815)	5.16925*** (1.62239)	6.34673*** (1.68459)	6.47932*** (1.92809)
Constante	-64.7159 (70.6854)	-141.4577 (132.62722)	-87.99665 (63.97466)	41.10918 (64.48779)	68.99111 (51.01243)	51.42508 (78.43099)
Controles (características)						
Indivíduos e família	Sim	Sim	Sim	Sim	Sim	Sim
Professores	Sim	Sim	Sim	Sim	Sim	Sim
Escolas	Sim	Sim	Sim	Sim	Sim	Sim
R ²	0.7287	0.33532	0.46343	0.56492	0.44726	0.23921
Estatística F	42.92***	4.26***	15.86***	39.86***	14.00***	2.55***
Prob.>F	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000

Fonte: Resultados da pesquisa

Nota: Erros padrão em parênteses. * $p < .1$, ** $p < .05$, *** $p < .01$. Utilizou-se o peso obtido no pareamento (5 vizinhos). Erros padrão das regressões quantílicas estimados por *bootstrap*.

Em suma, a partir do método, e da robustez fornecida pela similaridade dos coeficientes entre as estimativas, afirma-se que a frequência da criança na pré-escola possui relação positiva sobre seu desempenho nos testes de proficiência de português e matemática. O coeficiente médio da pré-escola sobre o teste de português foi de 4.18 pontos, o qual representa 9.1% do desvio padrão das notas. Já para o teste de matemática, o coeficiente da pré-escola foi 4.45, ou seja, 10,1% do desvio padrão. Não obstante, não foi possível definir o valor do impacto da pré-escola, pois se reconhece a presença de endogeneidade no modelo, mesmo utilizando o método proposto. Isto deve-se ao fato de que os métodos são baseados na quase aleatoriedade da variável pré-escola, ou seja, espera-se que após ter controlado em observáveis ela não esteja correlacionada com fatores não observáveis. Esta hipótese é forte dado que variáveis relativas à percepção dos pais sobre o retorno futuro da educação dos filhos, variáveis pré-natais e predisposições genéticas podem estar correlacionadas com a decisão de colocar a criança na escola.

Por outro lado, a diferença dos coeficientes nas regressões quantílica mostra que a pré-escola facilita o processo de aprendizagem da criança na escola. A criação dos alicerces do cérebro na primeira infância prepara a crianças para a vida na escola. Características como motivação, esforço e prestar atenção na aula ajudam na aprendizagem de novos conhecimentos na escola, portanto este seria um dos caminhos pelo qual o efeito da pré-escola afeta os testes.

5 CONCLUSÕES

A literatura internacional define o acúmulo de capital humano como um processo composto de vários estágios sequenciais, nos quais os estágios iniciais são os de maior relevância. Dado que cada estágio representa uma etapa distinta dos indivíduos, os estágios iniciais desse processo se referem à primeira infância dos indivíduos. Diversos estudos mostram os efeitos do cuidado e da educação durante a primeira infância no acúmulo de capital humano e nos resultados futuros dos indivíduos. Nesse contexto, o presente trabalho visou verificar qual é o efeito que uma intervenção da primeira infância, como a pré-escola, tem sobre os testes de proficiência do Sistema de Avaliação da Educação Básica (SAEB) aplicados nos alunos do 5º ano do ensino fundamental no Brasil, como medida de conhecimento acumulado ao longo do processo de aprendizado infantil.

Para responder ao questionamento que norteou o presente trabalho, foi proposta uma estratégia de identificação de duas etapas utilizando os bancos de dados do Censo Demográfico de 2010 e o SAEB de 2015. Para permitir o uso conjunto dos dois bancos de dados, foi definido

coortes por características invariáveis no tempo como: sexo, raça, UF de residência, ano de nascimento e participação na pré-escola. Isso possibilitou verificar os fatores que influenciaram a decisão de participar da pré-escola, no momento da entrada na pré-escola, algo negligenciado pela literatura nacional, devido à ausência de dados que acompanhem o mesmo indivíduo. Nesta parte do trabalho, constatou-se que há sim limitações em considerar apenas os inputs contemporâneos aos testes para capturar os efeitos da pré-escola, pois existem variáveis relevantes na decisão da criança frequentar a pré-escola tais como: o número de horas de trabalho da mãe e do pai, o número de pessoas na família e a renda familiar que, além de não serem invariantes no tempo, elas não são incluídas nas estimações dos trabalhos nacionais.

Após isso, definiu-se os grupos de tratamento e controle, que fossem os mais idênticos possíveis por meio do método de *propensity score matching*, obtendo-se pesos usados nas estimações das regressões por mínimos quadrados ordinário. Estimou-se também uma regressão quantílica não condicional para verificar se o possível efeito da pré-escola varia ao longo da distribuição das notas, uma vez que, de acordo com a escala do SAEB, quanto maior a nota alcançada, sinaliza-se um maior acúmulo de habilidades cognitivas e não cognitivas

Os resultados da estratégia mostram a existência de um efeito positivo da pré-escola sobre os testes tanto de português como de matemática. Reconhece-se aqui que, mesmo aplicada a estratégia proposta neste trabalho não é possível afirmar que o valor encontrado do coeficiente seja de impacto da pré-escola sobre os testes de proficiência, visto que ainda podem existir variáveis de confusão relativas às características dos pais, da criança e da turma. Além disso, devido a não existência de dados sobre os indivíduos ao longo do tempo, acompanha-se o comportamento de grupos com características semelhantes invariantes no tempo. Todavia, a semelhança entre os coeficientes obtidos tanto pelas especificações usadas e pelas regressões quantílica não condicional garantem robustez da estratégia.

Os coeficientes símeis nas estimações entre as duas disciplinas avaliadas levam a deduzir que o caminho pelo qual a educação infantil afeta o desempenho dos alunos é através do desenvolvimento de habilidade não cognitivas. As habilidades como: prestar atenção, motivação para estudar e disciplina influenciam na aprendizagem dos indivíduos nos diferentes níveis de escolaridade. Portanto, as crianças que frequentam a pré-escola tendem a possuir maior capacidade de adquirir os conhecimentos e habilidades que devem ser desenvolvidas nos anos iniciais do ensino fundamental. Foi encontrado também que o coeficiente da pré-escola varia ao longo da distribuição das notas. Estes resultados confirmam o papel da pré-escola como

facilitadora no processo de aprendizagem da criança na escola, uma vez que os coeficientes são maiores no quantil mais alto da distribuição. A obtenção de notas altas exige o maior grau de esforço e conhecimento do aluno, pois as dificuldades das questões aumentam conforme aumentam os níveis na escala do SAEB. Os resultados encontrados corroboram com a hipótese planteada para esta pesquisa. A pré-escola possui efeitos positivos no desempenho acadêmico em português e matemática dos alunos do 5º ano do ensino fundamental e esses efeitos são maiores nos quantis mais altos dos escores do teste do SAEB.

A primeira infância prepara as crianças para a vida na escola e características como: motivação, esforço e prestar atenção na aula ajudam na aprendizagem de novos conhecimentos e habilidades durante a vida. É nesse ponto que reside a importância da pré-escola para o desenvolvimento do capital humano. Dessa forma, sugere-se, com base nos resultados, políticas educacionais que visem o aprimoramento da qualidade educacional desta etapa de ensino e que estas caminhem de mãos dadas com as metas do Plano Nacional de Educação da universalização da pré-escola.

As limitações enfrentadas na pesquisa dizem respeito à escassez de banco de dados. A falta de dados em painel não permite controlar variáveis fixas no tempo como as habilidades inatas e nem identificar se o indivíduo migrou de município entre o período da pré-escola e a realização do teste do SAEB. Além disso, não se encontrou variáveis relativas à ordem do filho na família, a noção dos pais sobre os retornos da educação e a idade da família. Esta última variável refere-se ao tempo de existência da família desde a sua formação. Famílias novas podem ser compostas por indivíduos mais novos e no início da vida profissional, portanto, a renda familiar seria menor e o investimento na educação dos primogênitos seria menor quando comparado com os próximos filhos.

As limitações supracitadas são oportunidades para pesquisas futuras sobre a pré-escola. Aconselha-se também explorar outro tipo de medida de capital humano que permita capturar melhor a diferenciação entre as habilidades cognitivas e não cognitivas desenvolvidas na pré-escola, além de sondar os efeitos indiretos da pré-escola no mercado de trabalho feminino e na renda familiar.

6 REFERÊNCIAS

- ADHVARYU, A.; NYSHADHAM, A. Endowments at birth and parents' investments in children. **The Economic Journal**, v. 126, n. 593, p. 781–820, 2016.
- AIZER, A.; CUNHA, F. **The production of human capital: Endowments, investments and fertility**. [s.l.: s.n.].
- AKRESH, R. et al. **Child Labor, Schooling, and Child Ability**: Working paper 5965, The World Bank. [s.l.: s.n.].
- ALMEIDA, A. C.; LIMA, J. E.; COSTA, L. V. A participação da mãe no mercado de trabalho e o diferencial de anos de estudos por gênero entre adolescentes. **Estudos Econômicos (São Paulo)**, v. 48, p. 597–622, 2018.
- ALMLUND, M. et al. Personality psychology and economics. In: **Handbook of the Economics of Education**. [s.l.] Elsevier, 2011. v. 4p. 1–181.
- ANDREWS, R. J.; LI, J.; LOVENHEIM, M. F. **Quantile treatment effects of college quality on earnings: Evidence from administrative data in texas**. Cambridge, MA: [s.n.].
- AQUINO, J. M. DE; PAZELLO, E. T. Trabalho materno e desempenho educacional das crianças: uma análise da probabilidade de aprovação escolar. **Pesquisa e Planejamento Econômico**, v. 41, n. 1, p. 71–100, 2011.
- BALART, P.; OOSTERVEEN, M.; WEBBINK, D. Test scores, noncognitive skills and economic growth. **Economics of Education Review**, v. 63, p. 134–153, 2018.
- BARRO, R. J. **Human capital and growth: Theory and evidence: A comment**. Carnegie-Rochester Conference Series on Public Policy. **Anais...**1990
- BAUCHMÜLLER, R.; GORTZ, M.; RASMUSSEN, A. W. long-run benefits from universal high-quality preschooling. **Early Childhood Research Quarterly**, v. 29, n. 4, p. 457–470, 2014.
- BECKER, G. S. Investment in human capital: A theoretical analysis. **Journal of political economy**, v. 70, n. 5, Part 2, p. 9–49, 1962.
- BECKER, G. S. **HumanCapital: A Theoretical and Empirical Analysis, with Special Reference to Education**. New York: [s.n.].
- BECKER, S.; ICHINO, A. Estimation of average treatment effects based on propensity scores. **The Stata Journal**, v. 2, n. 4, p. 358–377, 2002.
- BEHRMAN, JERE R; CHENG, YINGMEI; TODD, P. E. Evaluating preschool programs when length of exposure to the program varies: A nonparametric approach. **Review of Economics and Statistics**, v. 86, n. 1, p. 108–132, 2004.
- BENZIL, C.; HANSEN, J.; KRISTENSEN, N. **Estimating Complementarity between education and training**. [s.l.: s.n.].
- BERLINSKI, S.; GALIANI, S.; GERTLER, P. The effect of pre-primary education on primary school performance. **Journal of public Economics**, v. 93, n. 1–2, p. 219–234, 2009.
- BEZERRA, M. E. G. **O trabalho infantil afeta o desempenho escolar no Brasil?** [s.l.] Universidade de São Paulo, 2006.

BEZERRA, M. E. G.; KASSOUF, A. L. **Análise dos fatores que afetam o desempenho escolar nas escolas das áreas urbanas e rurais do Brasil**. XLIV Congresso da Sociedade Brasileira de Economia, Administração e Sociologia Rural. **Anais...**Fortaleza, Ceará: 2006

BEZERRA, M. E. G.; KASSOUF, A. L.; ARENDS-KUENNING, M. **The impact of child labor and school quality on academic achievement in Brazil**. [s.l: s.n.].

BORGHANS, L.; MEIJERS, H.; TER WEEL, B. The role of noncognitive skills in explaining cognitive test scores. **Economic Inquiry**, v. 46, n. 1, p. 2–12, 2008.

BORGHANS, L.; SCHILS, T. **The leaning tower of Pisa: Decomposing achievement test scores into cognitive and noncognitive components**. [s.l: s.n.]. Disponível em: <<https://www.sole-jole.org/13260.pdf>>.

BRASIL. **Decreto lei Nº 12.796, de 4 de abril de 2013**. Disponível em: <http://www.planalto.gov.br/ccivil_03/_ato2011-2014/2013/lei/112796.htm>. Acesso em: 30 maio. 2018.

BRASIL. **Lei nº 11.274, 6 de fevereiro de 2006. Altera a redação dos arts. 29, 30, 32 e 87 da Lei nº 9.394, de 20 de dezembro de 1996, que estabelece as diretrizes e bases da educação nacional, dispondo sobre a duração de 9 (nove) anos para o ensino fundamental, c.** Disponível em: <http://www.planalto.gov.br/ccivil_03/_Ato2004-2006/2006/Lei/L11274.htm%3E>. Acesso em: 29 nov. 2018.

CAMERON, A. C.; TRIVEDI, P. K. **Microeconometrics: Methods and Applications**. New York: Cambridge University Press, 2005.

CAMILLI, G. et al. Meta-analysis of the effects of early education interventions on cognitive and social development. **Teachers College Record**, v. 112, n. 3, p. 579–620, 2010.

CAMPBELL, F. A. et al. Early Childhood Education: Young Adult Outcomes from the Abecedarian Project. **Applied Developmental Science**, v. 6, n. 1, p. 42–57, 2002.

CARNEIRO, P.; CUNHA, F.; HECKMAN, J. J. **The technology of skill formation**. Human Capital, Productivity and the Labour Market. **Anais...**London, Ontario, Canada: 2004

CARNEIRO, P. M.; HECKMAN, J. J. **Human capital policy**. Cambridge, MA: [s.n.].

CASTRO, J. F.; ROLLESTON, C. The contribution of early childhood and schools to cognitive gaps: New evidence from Peru. **Economics of Education Review**, v. 64, p. 144–164, 2018.

CORTÁZAR, A. Long-term effects of public early childhood education on academic achievement in Chile. **Early Childhood Research Quarterly**, v. 32, n. 3, p. 13–22, 2015.

CUNHA, F.; HECKMAN, J. The technology of skill formation. **American Economic Review**, v. 97, n. 2, p. 31–47, 2007.

CUNHA, F.; HECKMAN, J. J.; SCHENNACH, S. M. Estimating the Technology of Cognitive and Noncognitive Skill Formation. **Econometrica**, v. 78, n. 3, p. 883–931, 1 maio 2010.

CURI, A. Z.; MENEZES-FILHO, N. A. A relação entre educação pré-primária, salários, escolaridade e proficiência escolar no Brasil. **Estudos Econômicos**, v. 39, n. 4, p. 811–850, 2009.

CURI, A. Z.; MENEZES-FILHO, N. A. The relationship between school performance and future wages in Brazil. **EconomiA**, v. 15, n. 3, p. 261–274, 2014.

CURRIE, J. Early Childhood Education Programs. **Journal of Economic Perspectives**, v. 15, n. 2, p. 213–238, 2001.

CURRIE, J. Healthy, Wealthy, and Wise: Socioeconomic Status, Poor Health in Childhood, and Human Capital Development. **Journal of Economic Literature**, v. 47, n. 1, p. 87–122, 2009.

CURRIE, J.; VOGL, T. Early-life health and adult circumstance in developing countries. **Annu. Rev. Econ.**, v. 5, n. 1, p. 1–36, 2013.

DE FELÍCIO, F.; VASCONCELLOS, L. **O efeito da educação infantil sobre o desempenho escolar medido em exames padronizados**. XXXV Encontro Nacional de Economia. **Anais...**Recife, Pernambuco: 2007

DE HOLANDA BARBOSA FILHO, F.; DE ABREU PESSÔA, S. Educação e crescimento: o que a evidência empírica e teórica mostra? **Revista EconomiA**, v. 11, n. 2, p. 265–303, 2010.

FIRPO, S. Efficient semiparametric estimation of quantile treatment effects. **Econometrica**, v. 75, n. 1, p. 259–276, 2007.

GARCES, E.; THOMAS, D.; CURRIE, J. Longer-term effects of Head Start. **American economic review**, v. 92, n. 4, p. 999–1012, 2002.

GLEWWE, P.; JACOBY, H. G. Economic growth and the demand for education: is there a wealth effect? **Journal of Development Economics**, v. 74, n. 1, p. 33–51, 2004.

HANUSHEK, E. A. Economic growth in developing countries: The role of human capital. **Economics of Education Review**, v. 37, p. 204–212, 2013.

HANUSHEK, E. A.; KIMKO, D. D. Schooling, labor-force quality, and the growth of nations. **The American economic review**, v. 90, n. 5, p. 1184–1208, 2000.

HANUSHEK, E. A.; WOESSMANN, L. The role of cognitive skills in Economic Development. **Journal of Economic Literature**, v. 46, n. 3, p. 607–668, 2008.

HANUSHEK, E. A.; WOESSMANN, L. The economics of international differences in educational achievement. In: **Handbook of the Economics of Education**. [s.l.] Elsevier B.V., 2011a. p. 89–200.

HANUSHEK, E. A.; WOESSMANN, L. How much do educational outcomes matter in OECD countries? **Economic Policy**, v. 26, n. 67, p. 427–491, 2011b.

HANUSHEK, E. A.; WOESSMANN, L. Do better schools lead to more growth? Cognitive skills, economic outcomes, and causation. **Journal of Economic Growth**, v. 17, n. 4, p. 267–321, 2012.

HECKMAN, J. J. Skill formation and the economics of investing in disadvantaged children. **Science**, v. 312, n. 5782, p. 1900–1902, 2006.

HECKMAN, J. J. The economics, technology, and neuroscience of human capability formation. **Proceedings of the national Academy of Sciences**, v. 104, n. 33, p. 13250–13255, 2007.

HECKMAN, J. J. Schools, Skills, and synapses. **Economic Inquiry**, v. 46, n. 3, p. 289–324, 2008.

HECKMAN, J.; PINTO, R.; SAVELYEV, P. Understanding the mechanisms through which an influential early childhood program boosted adult outcomes. **American Economic Review**,

v. 103, n. 6, p. 2052–2086, 2013.

INSTITUTO BRASILEIRO DE GEOGRAFIA E ESTATÍSTICA, IBGE. **Pesquisa Nacional por Amostra de Domicílios Contínua: educação: 2017**. Disponível em: <https://biblioteca.ibge.gov.br/visualizacao/livros/liv101576_informativo.pdf>. Acesso em: 20 nov. 2018.

INSTITUTO NACIONAL DE ESTUDOS E PESQUISAS EDUCACIONAIS ANÍSIO TEIXEIRA, INEP. **Relatório do 1º ciclo de monitoramento das metas do PNE: biênio 2014-2016**. Disponível em: <http://portal.inep.gov.br/informacao-da-publicacao/-/asset_publisher/6JYIsGMAMkW1/document/id/626732>. Acesso em: 20 nov. 2018.

_____. **Sistema de Avaliação da Educação Básica SAEB**. Disponível em: <<http://portal.inep.gov.br/web/%0Aguest/educacao-basica/saeb>>. Acesso em: 20 set. 2018a.

_____. **Escala de proficiência de português, 5 ano do ensino fundamental**. Disponível em: <http://download.inep.gov.br/educacao_basica/prova_brasil_saeb/escala/escala_proficiencia/2018/PT_5EF.pdf>. Acesso em: 15 set. 2018b.

_____. **Escala de proficiência de matemática, 5 ano do ensino fundamental**. Disponível em: <http://download.inep.gov.br/educacao_basica/prova_brasil_saeb/escala/escala_proficiencia/2018/MT_5EF.pdf>. Acesso em: 15 set. 2018c.

_____. **Relatório do 2 ciclo de monitoramento das metas do plano nacional de educação-2018**. Brasília-DF: [s.n.].

_____. **Relatório SAEB (ANEB e ANRESC) 2005 - 2015: panorama da década**. Brasília-DF: [s.n.].

LEE, D. W.; LEE, T. H. Human capital and economic growth tests based on the international evaluation of educational achievement. **Economics Letters**, v. 47, n. 2, p. 219–225, 1995.

LOKSHIN, M. M.; GLINSKAYA, E.; GARCIA, M. **The effect of early childhood development programs on women's labor force participation and older children's schooling in Kenya**: 2376. Washington, D.C.: [s.n.].

MINISTÉRIO DA EDUCAÇÃO, MEC. **Educação básica. Crianças terão que ir à escola a partir dos 4 anos de idade**. Disponível em: <<http://portal.mec.gov.br/ultimas-noticias/211-218175739/18563-criancas-terao-de-ir-a-escola-a-partir-do-4-anos-de-idade>>. Acesso em: 1 jun. 2018.

_____. **Base nacional comum curricular**. Disponível em: <http://portal.mec.gov.br/index.php?option=com_docman&view=download&alias=79611-anexo-texto-bncc-aprovado-em-15-12-17-pdf&category_slug=dezembro-2017-pdf&Itemid=30192>. Acesso em: 30 maio. 2018.

MINCER, J. The distribution of labor incomes: a survey with special reference to the human capital approach. **Journal of economic literature**, v. 8, n. 1, p. 1–26, 1970.

MINCER, J. A. Schooling and earnings. In: **Schooling, experience, and earnings**. [s.l.] NBER, 1974. p. 41–63.

OLMSTED, P. P.; WEIKART, D. P. **How nations serve young children: Profiles of child care and education in 14 countries**. [s.l.] High/Scope, Ypsilanti, Mich., US, 1989.

ON THE DEVELOPING CHILD., C. **From best practices to breakthrough impacts: A**

science-based approach to building a more promising future for young children and families Harvard University, Center on the Developing Child Cambridge, MA, , 2016.

ONO, I. B. **O impacto do trabalho infantil no desempenho escolar**. [s.l.] Universidade de São Paulo, 2015.

PARISH, W. L.; WILLIS, R. J. Daughters, Education, and Family Budgets Taiwan Experiences. **The Journal of Human Resources**, v. 28, n. 4, p. 863–898, 1993.

PAXSON, C.; SCHADY, N. Cognitive development among young children in Ecuador the roles of wealth, health, and parenting. **Journal of Human resources**, v. 42, n. 1, p. 49–84, 2007.

ROMER, P. M. Human Capital And Growth: Theory and Evidence. **National Bureau of Economic Research Working Paper Series**, v. No. 3173, n. 3173, p. 1–49, 1989.

RUBIN, D. B. Using propensity scores to help design observational studies: Application to the tobacco litigation. **Health services and outcomes research methodology**, v. 2, n. 3–4, p. 169–188, 2001.

RUSSO, L. X.; DIAS, J. The health influence on returns to education in Brazil: A nonlinear approach. **EconomiA**, v. 17, n. 2, p. 210–220, 2016.

SCHULTZ, T. P. Health human capital and economic development. **Journal of African Economies**, v. 19, n. suppl_3, p. 12–80, 2010.

SCHULTZ, T. W. Investment in human capital. **The American economic review**, v. 51, n. 1, p. 1–17, 1961.

TODD, P. E.; WOLPIN, K. I. **The production of cognitive achievement in children: Home, school and racial test score gaps**. Philadelphia: [s.n.].

WOOLDRIDGE, J. M. **Econometric Analysis of Cross Section and Panel Data**. Cambridge, MA: MIT Press, 2002.

APÊNDICES
APÊNDICE A

Tabela 8: Média das covariadas e proporção da redução do viés pelo algoritmo de 5 vizinhos mais próximos em 2010.

Variáveis	Não pareada (N)	Média		% viés	% Redução	t-test
	Pareada (P)	Tratado	Controle		viés	p> t
Sexo	N	0.49138	0.48877	0.5	87	0.059
	P	0.49138	0.49172	-0.1		0.836
Idade	N	4.8131	5.1505	-42.9	90.3	0.000
	P	4.8131	4.7804	4.2		0.000
Cor ou raça	N	0.48885	0.45958	5.9	96.5	0.000
	P	0.48885	0.48782	0.2		0.685
Localização	N	0.75671	0.68704	15.6	76.7	0.000
	P	0.75671	0.74047	3.6		0.000
Rendimento per capita familiar	N	455.93	372.88	8.8	95.5	0.000
	P	455.93	452.19	0.4		0.636
Número de pessoas na família	N	4.455	4.7242	-18.4	86.3	0.000
	P	4.455	4.4918	-2.5		0.000
Escolaridade da mãe	N	7.8851	7.4626	18.2	79.7	0.000
	P	7.8851	7.7992	3.7		0.000
Horas de trabalho da mãe, por semana	N	17.863	16.457	6.6	94.2	0.000
	P	17.863	17.782	0.4		0.495
Idade da mãe	N	31.903	32.134	-2.9	85.9	0.000
	P	31.903	31.87	0.4		0.907
Escolaridade do pai	N	7.3919	7.0193	15.1	81.1	0.000
	P	7.3919	7.3216	2.9		0.000
Horas de trabalho do pai, por semana	N	38.088	37.003	5.2	79.6	0.000
	P	38.088	37.867	1.1		0.000
Oferta de escolas de educação infantil	N	0.00753	0.00757	-0.1	-794.9	0.792
	P	0.00753	0.00792	-1.3		0.775
Famílias monoparentais	N	0.00619	0.00573	0.6	88	0.030
	P	0.00619	0.00624	-0.1		0.329
Norte	N	0.08375	0.11752	-11.2	81.1	0.000
	P	0.08375	0.09014	-2.1		0.000
Nordeste	N	0.32986	0.30998	4.3	61.6	0.000
	P	0.32986	0.32223	1.6		0.000
Sudeste	N	0.35136	0.29038	13.1	85.8	0.000
	P	0.35136	0.34271	1.9		0.000
Sul	N	0.16376	0.19799	-8.9	76.4	0.000
	P	0.16377	0.17183	-2.1		0.000
R de Rubin	N	0.92				
	P	0.78				
B de Rubin	N	52.6				
	P	7.5				

Fonte: Resultados da pesquisa.

APÊNDICE B

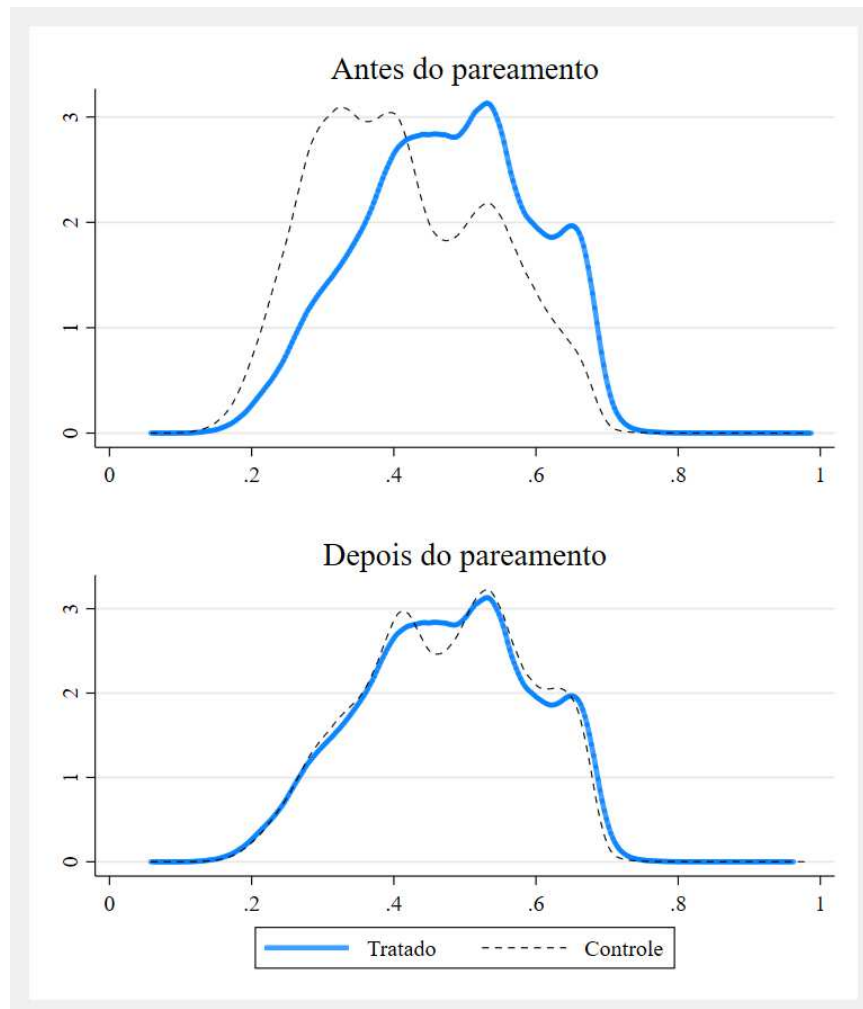


Figura 6: Distribuição do propensity score estimados antes e depois do pareamento pelo algoritmo de kernel.

Fonte: Resultado da pesquisa.

APÊNDICE C

Tabela 9: Média das covariadas e proporção da redução do viés pelo algoritmo de Kernel em 2010.

Variáveis	Não pareada (N)	Média		% viés	% Redução viés	t-test p> t
		Pareada (P)	Tratado			
Sexo	N		0.49138	0.48877	0.5	0.059
	P		0.49138	0.49049	0.2	66.1
Idade	N		4.8131	5.1505	-42.9	0.000
	P		4.8131	4.8118	0.2	99.6
Cor ou raça	N		0.48885	0.45958	5.9	0.000
	P		0.48885	0.48028	1.7	70.7
Localização	N		0.75671	0.68704	15.6	0.000
	P		0.75671	0.73536	4.8	69.4
Rendimento per capita familiar	N		455.93	372.88	8.8	0.000
	P		455.93	406.88	5.2	40.9
Número de pessoas na família	N		4.455	4.7242	-18.4	0.000
	P		4.455	4.5084	-3.6	80.2
Escolaridade da mãe	N		7.8851	7.4626	18.2	0.000
	P		7.8851	7.7391	6.3	65.4
Horas de trabalho da mãe, por semana	N		17.863	16.457	6.6	0.000
	P		17.863	17.402	2.2	67.2
Idade da mãe	N		31.903	32.134	-2.9	0.000
	P		31.903	31.816	1.1	62.3
Escolaridade do pai	N		7.3919	7.0193	15.1	0.000
	P		7.3919	7.2625	5.3	65.3
Horas de trabalho do pai, por semana	N		38.088	37.003	5.2	0.000
	P		38.088	37.725	1.7	66.5
Oferta de escolas de educação infantil	N		0.00753	0.00757	-0.1	0.792
	P		0.00753	0.00742	0.3	-139.6
Famílias monoparentais	N		0.00619	0.00573	0.6	0.030
	P		0.00619	0.00651	-0.4	29.2
Norte	N		0.08375	0.11752	-11.2	0.000
	P		0.08375	0.0914	-2.5	77.3
Nordeste	N		0.32986	0.30998	4.3	0.000
	P		0.32986	0.32326	1.4	66.8
Sudeste	N		0.35136	0.29038	13.1	0.000
	P		0.35136	0.332	4.2	68.2
Sul	N		0.16377	0.19799	-8.9	0.000
	P		0.16377	0.17767	-3.6	59.4
R de Rubin	N		0.92			
	P		0.64			
B de Rubin	N		52.6			
	P		7.1			

Fonte: Resultados da pesquisa.

APÊNDICE D

Tabela 10: Modelos estimados por MQO para as notas de matemática e português dos alunos do 5 ano de ensino fundamental em 2015

Variáveis	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	Português				Matemática			
Pré-escola	7.21409*** (0.08229)	6.43105*** (0.15416)	6.36471*** (0.15914)	6.46401*** (0.16432)	7.11305*** (0.08181)	6.57173*** (0.15437)	6.49675*** (0.15883)	6.51137*** (0.16370)
sexo		8.38007*** (0.15596)	8.32543*** (0.16097)	8.33393*** (0.16614)		-5.39480*** (0.15613)	-5.40882*** (0.16062)	-5.45528*** (0.16546)
Cor_raça		6.35300*** (0.15833)	5.46104*** (0.16482)	4.98539*** (0.17029)		6.80905*** (0.15854)	5.56098*** (0.16449)	5.00414*** (0.16963)
idade		2.13692*** (0.14559)	2.16036*** (0.15041)	2.16162*** (0.15524)		0.79336*** (0.14578)	0.89004*** (0.15011)	0.93975*** (0.15463)
Mora com a mãe		5.30942*** (0.30523)	4.82982*** (0.31563)	4.57792*** (0.32629)		6.20802*** (0.30560)	5.56105*** (0.31498)	5.33804*** (0.32501)
Escolaridade da mãe		3.13823*** (0.07557)	3.38643*** (0.07809)	3.49017*** (0.08068)		2.04857*** (0.07566)	2.37230*** (0.07792)	2.50678*** (0.08035)
Mora com o pai		3.68565*** (0.17378)	3.62149*** (0.17952)	3.51057*** (0.18537)		3.92640*** (0.17399)	3.89089*** (0.17915)	3.83486*** (0.18464)
Escolaridade do pai		1.55479*** (0.07114)	1.66201*** (0.07352)	1.76155*** (0.07590)		0.50032*** (0.07123)	0.69480*** (0.07337)	0.80780*** (0.07560)
Pais assistem às reuniões		7.77364*** (0.33229)	7.54774*** (0.34269)	7.42655*** (0.35297)		7.02490*** (0.33264)	6.74359*** (0.34197)	6.56462*** (0.35157)
Incentivam a estudar		18.16061*** (0.75858)	17.32352*** (0.78361)	16.61074*** (0.81291)		14.49119*** (0.75900)	13.76422*** (0.78127)	12.80619*** (0.80949)
Incentivam fazer dever de casa		13.49877*** (0.55281)	13.39714*** (0.57184)	13.26683*** (0.59086)		9.38213*** (0.55367)	8.86341*** (0.57077)	8.51594*** (0.58886)
Lê livros		10.61044*** (0.38823)	10.64888*** (0.40066)	10.41608*** (0.41297)		9.55255*** (0.38871)	9.58613*** (0.39986)	9.31119*** (0.41144)
Lê revistas		4.33113*** (0.16387)	4.51138*** (0.16912)	4.55916*** (0.17450)		3.28501*** (0.16407)	3.49400*** (0.16877)	3.60053*** (0.17381)

Tempo (TV e internet)	3.91575*** (0.06126)	3.73159*** (0.06334)	3.58093*** (0.06543)	4.35141*** (0.06133)	4.12136*** (0.06320)	3.96685*** (0.06516)
Trab. doméstico (tempo)	-3.29156*** (0.07472)	-3.27167*** (0.07712)	-3.20901*** (0.07965)	-2.60611*** (0.07479)	-2.59719*** (0.07694)	-2.57027*** (0.07931)
Trab. fora de casa	-2.3e+01*** (0.27936)	-2.2e+01*** (0.28868)	-2.2e+01*** (0.29844)	-1.9e+01*** (0.27974)	-1.8e+01*** (0.28813)	-1.7e+01*** (0.29729)
Faz dever de port.	1.28965*** (0.16893)	1.45417*** (0.17493)	1.48354*** (0.18097)			
Renda familiar (<i>proxy</i>)	0.26853*** (0.01896)	0.12698*** (0.01971)	0.03591* (0.02044)	0.67189*** (0.01898)	0.49247*** (0.01967)	0.38758*** (0.02036)
Sexo do professor		3.96172*** (0.23767)	3.66663*** (0.24595)		3.96116*** (0.24396)	3.41892*** (0.25193)
Idade do professor		0.02850** (0.01129)	0.00376 (0.01169)		0.02168* (0.01126)	-0.00742 (0.01164)
Cor do professor		7.15810*** (0.15988)	5.55561*** (0.16786)		9.89821*** (0.15960)	8.03770*** (0.16719)
Escolaridade do professor		2.21431*** (0.17077)	1.76807*** (0.17764)		2.64749*** (0.17018)	2.16963*** (0.17668)
Lic. em port.		0.11516 (0.22643)	-0.40969* (0.23355)			
Salário do professor		0.67356*** (0.03694)	0.45731*** (0.03849)		0.57026*** (0.03690)	0.33960*** (0.03837)
Exp. professor		-0.00270 (0.01274)	0.02540* (0.01319)		0.01910 (0.01271)	0.04891*** (0.01313)
Situação do prof. na escola		0.19901*** (0.07274)	0.29253*** (0.07551)		0.30771*** (0.07258)	0.42008*** (0.07520)
Frequenta a biblioteca		1.16570*** (0.20353)	0.53627** (0.21029)		1.30862*** (0.20271)	0.65997*** (0.20919)
Passa dever de casa		1.42887*** (0.09606)	1.56419*** (0.09953)		1.90560*** (0.09496)	2.03113*** (0.09820)

Escola possui computadores				0.68522*				0.61623*
				(0.35848)				(0.35708)
Há acesso à internet				3.55510***				4.78364***
				(0.24208)				(0.24119)
Biblioteca				0.87299***				0.19158
				(0.21160)				(0.21088)
Quadra de esporte				4.88001***				6.63518***
				(0.19505)				(0.19434)
Lab. de informática				1.05621***				0.64240**
				(0.29614)				(0.29495)
Lab. de ciência				0.98398***				-1.09029***
				(0.23767)				(0.23698)
Alunos pegam emprestados os livros da escola				4.51586***				4.73841***
				(0.33638)				(0.33478)
Professores pegam emprestados os livros da escola				0.37572				1.54257***
				(0.32087)				(0.31939)
Faz dever de mat.						0.74543***	1.07606***	1.07010***
						(0.19143)	(0.19784)	(0.20483)
Lic. em matemática							-0.05624	-1.04076***
							(0.27101)	(0.28004)
Constante	214.564 ***	112.087 ***	85.62702***	79.48739***	225.454 ***	152.809 ***	120.677 ***	113.286 ***
	(0.05069)	(1.81222)	(2.10404)	(2.18327)	(0.05039)	(1.81695)	(2.10257)	(2.17755)
R ²	0.00597	0.11417	0.12469	0.13261	0.00588	0.09953	0.11708	0.12878
BIC	13400000	3342817	3096084	2872697	13400000	3355009	3105541	2880516
Estatística F	7685.96	2313.97	1524.42	1181.86	7560.07	1990.46	1423.49	1146.26
Prob.>F	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000

Fonte: Resultados da pesquisa

Nota: Erros padrão em parênteses. * $p < .1$, ** $p < .05$, *** $p < .01$. Erros padrão clusterizado por município.

APÊNDICE E

Tabela 11: Coeficientes estimados por MQO para as notas de matemática e português dos alunos do 5 ano de ensino fundamental em 2015 após a aplicação da estratégia de identificação-estimação utilizando o peso do pareamento por Kernel

Variáveis	(1)	(2)	(3)	(4)	(5)	(6)
	Português			Matemática		
Pré-escola	6.0819*** (.06559)	4.38356*** (1.0137)	4.1894*** (0.8978)	6.6276*** (.7119)	4.6614*** (1.1845)	4.47156*** (0.8373)
Constante	208.2287*** (4.4023)	-98.9219 (77.5833)	-64.7159 (67.8398)	217.9416*** (4.5160)	-92.9294 (72.4306)	-57.7946 (72.1227)
Controles (características)						
Indivíduo e da família	Não	Sim	Sim	Não	Sim	Sim
Professores	Não	Sim	Sim	Não	Sim	Sim
Escola	Não	Não	Sim	Não	Não	Sim
R ²	0.0313	0.6554	0.7196	0.0336	0.6514	0.7334
BIC	5492.447	4880.326	4793.445	5555.98	4950.109	4824.74
Estatística F	20.80***	41.17***	42.56***	22.39***	40.44***	45.63***
Prob.>F	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000

Fonte: Resultados da pesquisa

Nota: Erros padrão em parênteses. * $p < .1$, ** $p < .05$, *** $p < .01$. Erros padrão clusterizados por UF e estimados por bootstrap.

ANEXOS

ANEXO I

CRITÉRIO DE CLASSIFICAÇÃO ECONOMIA DA DISTRIBUIÇÃO DE CLASSES:
SISTEMA DE PONTOS.

	Quantidades				
	0	1	2	3	4 ou +
Banheiros	0	3	7	10	14
Empregados domésticos	0	3	7	10	13
Automóveis	0	3	5	8	11
Microcomputador	0	3	6	8	11
Lava louças	0	3	6	6	6
Geladeira	0	2	3	5	5
Freezer	0	2	4	6	6
Máquina de lavar	0	2	4	6	6
DVD	0	1	3	4	6
Micro-ondas	0	2	4	4	4
Motocicleta	0	1	3	3	3
Secador de roupa	0	2	2	2	2

Grau de instrução do chefe de família	
Analfabeto/ fundamental I incompleto	0
Fundamental I completo/ Fundamental II incompleto	1
Fundamental II completo/ Médio incompleto	2
Médio Completo/ Superior incompleto	4
Superior completo	7

Soma dos pontos		Renda média familiar por mês (R\$) ²¹
Classe	Pontos	
A	45-100	20.888
B1	38-44	9.254
B2	29-37	4.852
C1	23-28	2.705
C2	17-22	1.625
D-E	0-16	768

Fonte: ABEP (2018)

²¹ A aproximação da renda familiar foi realizada pela ABEP com os dados da Pesquisa Nacional por Amostra de Domicílio (PNAD) de 2014.

ANEXO II

Quadro 12: Escala de proficiência do SAEB

Escala de proficiência de língua portuguesa 5º ano do ensino fundamental		Escala de proficiência de matemática 5º ano do ensino fundamental	
Nível 0 <125 pontos	-Não há item que avaliam as habilidades deste nível -Os estudantes localizados abaixo do nível 125 requerem atenção especial, pois não demonstram sequer habilidades muito elementares.	Nível 0 <125 pontos	-Não há item que avaliam as habilidades deste nível -Os estudantes localizados abaixo do nível 125 requerem atenção especial, pois não demonstram sequer habilidades muito elementares.
Nível 1 125≤pontos<150	-Localizar informações explícitas em textos narrativos curtos. -Identificar o tema do texto -Localizar elementos como a personagem principal.	Nível 1 125≤pontos<150	-Determinar a área de figuras desenhadas em malhas quadriculadas por meio de contagem.
Nível 2 150≤pontos<175	-Localizar informações explícitas em contos. -Inferir características de personagens em fábulas. -Interpretar linguagem verbal e não verbal em tirinhas.	Nível 2 150≤pontos<175	- Resolver problemas do cotidiano envolvendo adição de pequenas quantias de dinheiro. -Localizar informações, relativas ao maior ou menor elemento, em tabelas ou gráficos
Nível 3 175≤pontos<200	-Reconhecer relação de causa e efeito em poemas, contos e tirinhas. - Inferir o sentido de palavra, o sentido da expressão ou o assunto em cartas, contos, tirinhas e histórias com o apoio de linguagem verbal e não verbal.	Nível 3 175≤pontos<200	-Reconhecer dentre um conjunto de polígonos, aquele que possui o maior número de ângulos. -Converter uma quantia, dada uma ordem das unidades de real, em seu equivalente em moedas. - Associar a fração $\frac{1}{4}$ a uma de suas representações gráficas.
Nível 4 200≤pontos<225	- Identificar assuntos comuns a duas reportagens. -Reconhecer sentido de expressão, elementos de narrativas e opinião de reportagens. - Inferir sentido decorrente da utilização de sinais de pontuação e sentido de expressões em poemas, fábulas e contos. -Inferir efeito de humor em tirinhas e histórias de quadrinhos.	Nível 4 200≤pontos<225	- Reconhecer a planificação de uma pirâmide dentre um conjunto de planificações. -Determinar o total de uma quantia a partir da quantidade de moedas de 25 e/ou 50 centavos que a compõe, ou vice-versa. Converte uma hora em minutos. -Reconhecer o princípio do valor posicional do sistema de numeração decimal. - Associar a metade de um total ao seu equivalente em porcentagem.

Nível 5 225 ≤ pontos < 250	- Identificar assunto e opinião em reportagens e contos. - Identificar informação explícita em letras de músicas e contos. - Reconhecer assuntos em poemas e tirinhas. - Reconhecer sentido de conjunções e de locuções adverbiais em verbetes, lendas e contos. - Diferenciar opinião de fato em reportagens.	Nível 5 225 ≤ pontos < 250	- Reconhecer a planificação de um cubo dentre um conjunto de planificações. - Determinar a área de um terreno retangular representado em uma malha quadriculada. - Resolver problemas, no sistema monetário nacional, envolvendo adição e subtração de cédulas e moedas. - Reconhecer uma função como representação da relação parte-todo, com apoio de um polígono dividido em 8 partes ou mais.
Nível 6 250 ≤ pontos < 275	- Inferir assunto principal e sentido de expressão em poemas, fábulas, contos, crônicas, reportagens e tirinhas. - Inferir informações em contos e reportagens. - Identificar informações explícita em reportagens com ou sem o auxílio de recursos gráficos.	Nível 6 250 ≤ pontos < 275	- Reconhecer o m² como unidade de medida de área. - Determinar porcentagens simples (25%, 50%) - Reconhecer uma fração como representação da relação parte-todo, sem apoio de figuras. - Resolver problemas que envolvem soma e subtração de valores monetários.
Nível 7 275 ≤ pontos < 300	- Reconhece elementos das narrativas em fábulas. - Inferir informação e efeito de sentido decorrente do uso de sinais gráficos em reportagens e em letras de músicas. - Interpretar linguagem verbal e não verbal em histórias em quadrinhos.	Nível 7 275 ≤ pontos < 300	- Interpretar a movimentação de um objeto utilizando referencial diferente do seu. - Determinar o perímetro de um retângulo desenhado em malha quadriculada, com as medidas de comprimento e largura explicitados. - Resolver problemas envolvendo conversão de quilograma para grama. - Resolver problemas que envolvem a divisão exata ou a multiplicação de número naturais.
Nível 8 300 ≤ pontos < 325	- Reconhecer relação de causa e consequência e relação entre pronomes e seus referentes em fábulas e reportagens. - Reconhecer assunto comum entre textos de gêneros diferentes.	Nível 8 300 ≤ pontos < 325	- Reconhecer uma linha paralela a outro dado como referência em um mapa. - Determinar a razão entre as áreas de duas figuras desenhadas numa malha quadriculada. - Converter medidas lineares de comprimento (m/cm) - Resolver problemas envolvendo divisão de número naturais com resto.

			-Associar 50% à sua representação na forma de fração.
Nível 9 325≤pontos	- Identificar opinião em fábulas e reconhecer sentido de advérbios em cartas do leitor.	Nível 9 325≤pontos,350	-Reconhecer a planificação de uma caixa cilíndrica. -Resolver problemas que envolvem a conversão entre unidades de comprimento. -Reconhecer frações equivalentes. - Comparar número racionais com quantidades diferentes de casas decimais. -Reconhecer o gráfico de linhas correspondente a uma sequência de valores ao longo do tempo (com valores positivos e negativos)
		Nível 10 350≤pontos	- Reconhecer dentre um conjunto de quadriláteros, aquele que possui lados perpendiculares e com a mesma medida. -Converter uma medida de comprimento, expressando décimos e centímetros, para milímetros.

Fonte: INEP (2018b, 2018c)