

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

Edi Flores Reyna¹, Maria Micheliana da Costa Silva², Viviani Silva Lírio²

Resumo: 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 5º 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. Para tanto, foram utilizados os bancos de dados de Censo Demográfico de 2010 e o SAEB de 2015. 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 e da regressão quantílica, considerando os pesos obtidos no pareamento. Os resultados encontrados mostram uma relação positiva entre as variáveis de interesse. 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. O coeficiente médio encontrado para pré-escola representa 9.1% do desvio padrão no teste de português e 10,1% no teste de matemática.

Palavras-Chave: Pré-escola; teste de proficiência; capital humano

Abstract: This paper 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. We provide new evidences of the possible effects of child education in the average and along the test grades distribution through a new identification strategy, based in a two steps estimation. For this, we were used the 2010 Demographic Census and 2015 SAEB data. In the first stage, we were identified control and treated groups by propensity score matching. After that, we were obtained the OLS and unconditional quantile regression estimators considering the parity weight. The results show a positive relationship between the variables of interest. With the quantile regressions, it was observed a higher relationship between pre-school and the grades in higher quantiles of SAEB scores, which can reflect more developed cognitive abilities. The average coefficient found in pre-school represents 9.1% of the standard deviation in Portuguese test and 10.1% in mathematics.

Keywords: Pre-school; proficiency test; human capital

Código JEL: A20, I21, I25, O15

1 INTRODUÇÃO

A acumulação de capital humano pode ser entendida como o processo pelo qual o indivíduo adquire um conjunto de habilidades cognitivas e não cognitivas e que após serem adquiridas são indissociáveis dele. Compreender este processo de formação de 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

¹ Doutorando em Economia Aplicada pela Universidade Federal de Viçosa, MG. E-mail: floresreyna.edi@gmail.com

² Professora do Departamento de Economia Rural da Universidade Federal de Viçosa, MG.

durante toda a vida do indivíduo (CARNEIRO; CUNHA; HECKMAN, 2004; CASTRO; ROLLESTON, 2018; CUNHA; HECKMAN, 2007; HECKMAN, 2006, 2007, 2008; HECKMAN; PINTO; SAVELYEV, 2013).

Embora a formação de habilidade seja um processo que acontece durante todo o ciclo de vida dos indivíduos, as bases da arquitetura cerebral são estabelecidas na primeira infância. 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. 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).

A educação na primeira infância, portanto, possui relevância no desenvolvimento de diversas habilidades dos indivíduos. Nesse sentido, a relação a ser explorada nesta pesquisa é a existente entre a pré-escola e o acúmulo de capital humano medido através dos escores dos testes de proficiência do Sistema de Avaliação da Educação Básica (SAEB) dos alunos do 5º ano do ensino fundamental. Especificamente verifica-se os fatores que influenciaram a decisão em frequentar a pré-escola, cinco anos antes da realização do teste; identificar os possíveis efeitos da pré-escola sobre a média e os quantis de notas dos testes de proficiência do SAEB.

Mesmo com diversos estudos sobre a primeira infância na literatura (BAUCHMÜLLER; GORTZ; RASMUSSEN, 2014; BEHRMAN; CHENG; TODD, 2004; BERLINSKI; GALIANI; GERTLER, 2009; CORTÁZAR, 2015; CURI; MENEZES-FILHO, 2009; CURRIE, 2001; DE FELÍCIO; VASCONCELLOS, 2007; GARCES; THOMAS; CURRIE, 2002; HECKMAN; PINTO; SAVELYEV, 2013), 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).

Dentro desse contexto, o presente trabalho contribui para a literatura de duas formas. Propõe-se, primeiramente, uma nova estratégia de identificação que permita utilizar dois bancos de dados diferentes. Esta nova perspectiva traz uma possível solução ao problema de ausência de dados longitudinais. Os dados, aqui utilizados, provêm do Censo Domiciliar de 2010 e do SAEB 2015, e serão juntados por meio de coortes definidas por características fixa como sexo, Unidade da Federação (UF), ano de nascimento e frequentou ou não a pré-escola. A estratégia é composta de duas etapas, detalhadas na seção da metodologia, pois se considera os inputs contemporâneos e passados ao processo educativo do indivíduo.

A análise do efeito da pré-escola sobre a distribuição das notas nos testes é uma lacuna ainda não explorada pela literatura. Analisar o comportamento ao longo da distribuição oferece uma noção mais detalhada dos efeitos da pré-escola sobre os diversos níveis de habilidades desenvolvidas pelas crianças durante os 5 primeiros anos da educação fundamental. Desse modo, a pesquisa contribui também proporcionando evidências sobre os possíveis efeitos da pré-escola, nos diferentes quantis de notas.

Os principais resultados encontrados confirmam a existência de uma correlação positiva entre frequentar a pré-escola e o desempenho acadêmico dos alunos do 5º ano do ensino fundamental, tanto na prova de português quanto na prova de matemática do SAEB. A aplicação da regressão quantílica não condicional mostrou que o coeficiente de correlação é maior nos

quantis superiores da distribuição, o qual captura maiores níveis de habilidades cognitivas e não cognitivas.

Este trabalho está organizado em cinco seções. Além desta introdução, apresenta-se o referencial teórico utilizado nesta pesquisa na seção dois. Na terceira seção mostra-se em maior detalhe a estratégia de identificação utilizada. Os resultados da pesquisa são expostos na quarta seção. Finalmente, as conclusões e as limitações do trabalho são indicadas na seção derradeira.

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) e Castro e Rolleston (2018).

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:

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

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 $f(\cdot)$ 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 estágios superiores são mais efetivos em indivíduos que possuem grandes estoques de habilidades. Logo, os estudantes com altos níveis de capital humano na infância demonstram maior eficiência na aquisição de novas habilidades.

Com relação à auto-produtividade, entende-se que o vetor de A_t permite a criação de maiores níveis de habilidades no período A_{t+1} com grande facilidade. Portanto, é possível escrever que $\frac{\partial f_t}{\partial A_t} > 0$. Esta, conjuntamente como a complementariedade dinâmica, mostra como o investimento na primeira infância gera maiores retornos e são mais eficientes do que em períodos posteriores. Estas duas características chaves geram um efeito multiplicador que é chamando de “*skill begets skill*”.

Nessa perspectiva, esta pesquisa levará em consideração um processo de acumulação de capital de dois estágios, seguindo a modelo apresentado por Heckman (2008) e Cunha e Heckman (2007). A equação que denota o processo é designada como:

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

A equação (2) pode ser escrita assumindo que a relação entre os investimentos apresenta a forma funcional de uma CES (*Constant Elasticity of Substitution*), permitindo assim relacionar os investimentos em dois períodos 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 parâmetro ϕ denota o grau de substituição entre os investimentos feitos no período 1 e 2. O parâmetro 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:

$$A_{i2} = \min\{I_1; I_2\} \quad (4)$$

Desse modo, para valores pequenos de ϕ observa-se que intervenções tardias não podem compensar a falta de investimento em capital humano de estágios iniciais da vida. Define-se o preço do investimento no período 1 igual a R\$ 1, no período 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:

$$\frac{I_1}{I_2} = \left[\frac{\gamma}{(1 - \gamma)(1 + r)} \right]^{\frac{1}{1-\phi}} \quad (5)$$

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 causal 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. Portanto, 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 das 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 viesadas. 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 indivíduo 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. 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, 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:

$$ATE = E \left[\frac{(D_i - p(X_i))Y_i}{(1 - p(X_i))P(D_i)} \right] \quad (6)$$

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. 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. Assume-se, portanto, 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 (7).

$$\ln f_{2010i} = X'_{2010i}\beta + \epsilon_i \quad (7)$$

Na equação (7) 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 nacional³.

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. 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).

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.

$$A_{i2015} = X_{if2015}\beta + X_{is2015}\alpha + X_{it2015}\gamma + \omega Educ_i + \mu_i \quad (8)$$

A equação (8) 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 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, o qual permitiria 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 potenciais⁴. Logo, o QTT pode ser escrito como:

$$QTT = \Delta_{\tau|T=1} = q_{1,\tau|T=1} - q_{0,\tau|T=1} \quad (9)$$

Na equação (9) 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 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

³ 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.

⁴ Nota do aluno dado que frequentou ou não a pré-escola.

divisão dos quantis de notas será feita de acordo com a escala de proficiências apresentada pelo SAEB (INEP, 2018a, 2018b). Portanto a equação (9) pode ser reescrita da seguinte maneira:

$$QTT = \left\{ \inf_q P[f(A_i; X^* = x_1^*; T = 1) \leq q] = \tau \right\} - \left\{ \inf_q P[f(A_i; X^* = x_1^*; T = 0) \leq q] \geq \tau \right\} \quad (10)$$

Onde X^* contém todos os vetores de controles identificados na equação (8).

3.2 Fontes de Dados

Os dados utilizados nesta pesquisa são provenientes de duas fontes. Do Censo Demográfico de 2010 disponibilizado pelo Instituto Brasileiro de Geografia e Estatística (IBGE) foram considerados domicílios em que exista pelo menos uma criança com idade entre 4 e 6 anos. Por sua vez do SAEB de 2015 disponibilizado pelo Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira (INEP), foram considerados os alunos que fizeram o 5º ano em 2015, que correspondem às crianças ingressantes na educação infantil em 2010. As variáveis que serão utilizadas no trabalho são apresentadas no Quadro 1.

Quadro 1: Variáveis utilizadas nas estimações

Dimensões	Características das crianças	Características das famílias	Características das escolas e professores
Variáveis	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	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)	Lab. de informática Lab. de ciências Tipo de escola (pública=1) Oferta de educação infantil ⁵ 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
Variável de interesse	Pré-escola (sim=1)		
Variáveis dependentes	Notas nos testes de português (prof_p) e matemática (prof_mat)		

Fonte: Elaboração dos autores.

⁵ 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

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, a fim de superar essa limitação, recorre-se a uma variável proxy, construída com base na metodologia da Associação Brasileira de Empresas de Pesquisa (ABEP). 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).

4 RESULTADOS E DISCUSSÕES

Seguindo a estratégia de identificação supracitada, esta seção apresenta os principais resultados da pesquisa. A subseção 4.1 mostra os resultados relativos à criação dos grupos de tratamento e controle. Os resultados da relação de interesse são apontados na subseção 4.2.

4.1 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. 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 (INEP, 2016a). 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.

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 (INEP, 2016b). Quando se considera a diferença entre o urbano e rural⁶, 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), 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.

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 despender 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 à escola.

⁶ Neste trabalho adota-se a diferenciação entre urbano e rural feito pelo Censo Demográfico.

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 ⁷	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.

⁷ Número de horas trabalhadas no trabalho principal.

Ainda na Tabela 1, observa-se que o aumento da probabilidade da criança frequentar a pré-escola está correlacionada com a escolaridade dos pais. 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. A diferença do número de crianças que frequentam a pré-escola entre os 25% mais ricos e os 25% mais pobres vem diminuindo (INEP, 2016b).

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 1. O algoritmo utilizado para o pareamento foi: 5 vizinhos mais próximos. Resultados análogos são encontrados com o algoritmo de Kernel. Na figura observa-se uma sobreposição das densidades do *propensity score* dos grupos após o pareamento, o que evidencia balanceamento satisfatório. Além disso, a estatística R^8 (0,64) e B^9 (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.

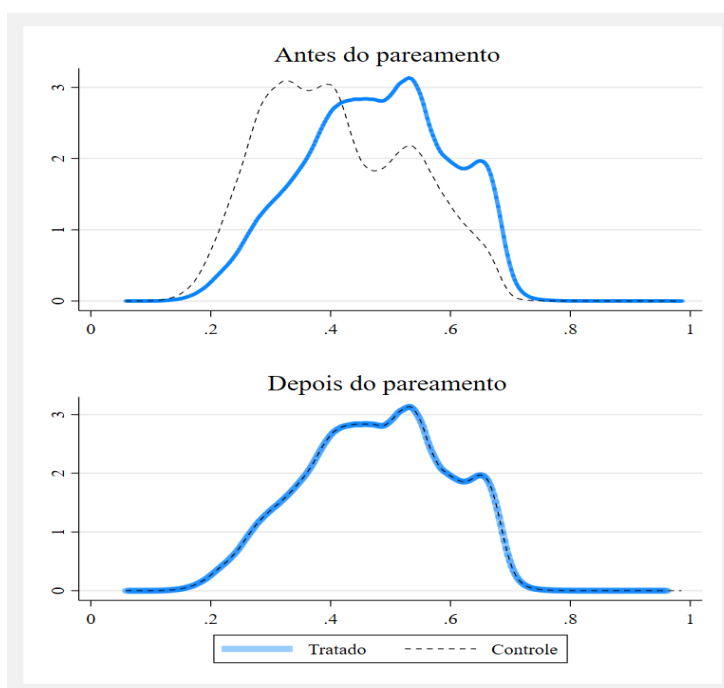


Figura 1: Distribuição do *propensity score* estimados antes e depois do pareamento.

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

⁸ Mostra a razão entre as variâncias do *propensity score* dos grupos após o pareamento.

⁹ Mostra a diferença bruta e padronizada das médias do *propensity score* entre os grupos após o pareamento.

4.2 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 ente 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 e português dos alunos do 5º ano de ensino fundamental em 2015, sem pareamento

	matemática			Português		
	(1)	(2)	(3)	(4)	(5)	(6)
Pré-escola	7.11305*** (0.08181)	6.49675*** (0.15883)	6.51137*** (0.16370)	7.21409*** (0.08229)	6.36471*** (0.15914)	6.46401*** (0.16432)
Constante	225.454*** (0.05039)	120.9004*** (2.11107)	113.2855*** (2.17756)	214.5642*** (0.05069)	85.62702*** (2.10404)	79.48739*** (2.18327)
Controles						
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	0.00588	0.11662	0.1288	0.00597	0.12469	0.13264
BIC	13400000	3105541	2880516	13400000	3096084	2872697
Estatística F	7560.07***	1359.56***	1146.26***	7685.96***	1524.42***	1181.86***
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 clusterizado por municípios; os controles são diversas características apresentadas no quadro I.

Para o teste de português, obteve-se também o sinal positivo entre a pré-escola e o desempenho do aluno. Na Tabela 2, 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 destes alicerces podem 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.

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 a fim 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 3. 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 3.

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.

Tabela 3: 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		
	(7)	(8)	(9)	(10)	(11)	(12)
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						
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*.

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 2003¹⁰ 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%. Observa-se então uma diferença entre esses coeficientes e os coeficientes expostos na tabela 3.

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

¹⁰ Em 2003, a proficiência média de matemática para os alunos da 4 série do ensino fundamental era de 177 pontos.

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, 2018c). 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 e 4 da Tabela 2. 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.

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 4.1, 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 testes¹¹. A proficiência média desses 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 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 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

¹¹ 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.

pesquisa mostram que os coeficientes da pré-escola sobre os testes representam 0.1 do desvio padrão¹².

Na Tabela 3, nota-se 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. 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 habilidade 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 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-escola teria efeitos na vida adulta dos indivíduos de duas maneiras. A primeira seria através das habilidades não cognitivas e a 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 cognitiva (INEP, 2018d). Na Tabela 4, 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 4, 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.

¹² O desvio padrão das notas de proficiência em português e matemática é de 46 e 44 pontos respectivamente.

Tabela 4: 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					
Pré-escola	Q10	Q25	Q50	Q75	Q90
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 condicionada¹³ para calcular os coeficientes da pré-escola. Nas estimativas apresentadas nas Tabela 5 e 6, foram utilizados os cortes definidos anteriormente e os pesos 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, nas tabelas 5 e 6, que o coeficiente da variável de interesse aumenta conforme os quantis de nota aumentam¹⁴. O coeficiente da pré-escola passa de 2.68 no quantil Q10 para 6.88 no quantil Q90 nas notas de matemática (Tabela 5). Para a disciplina de português (Tabela 6), 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, 2018d; MEC, 2018).

¹³ Para encontrar as estimativas foi utilizado o comando `rifreg` do `stata`.

¹⁴ Foi realizado também um teste de diferenças estatísticas entre os coeficientes das regressões quantílicas.

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

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

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

	Q10	Q25	Q50	Q75	Q90
Pré-escola	1.83453 (2.14534)	4.55524*** (1.50815)	5.16925*** (1.62239)	6.34673*** (1.68459)	6.47932*** (1.92809)
Constante	-141.4577 (132.62722)	-87.99665 (63.97466)	41.10918 (64.48779)	68.99111 (51.01243)	51.42508 (78.43099)
Controles					
Indivíduos e família	Sim	Sim	Sim	Sim	Sim
Professores	Sim	Sim	Sim	Sim	Sim
Escolas	Sim	Sim	Sim	Sim	Sim
R ²	0.33532	0.46343	0.56492	0.44726	0.23921
Estatística F	4.26***	15.86***	39.86***	14.00***	2.55***
Prob.>F	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. Diversos estudos mostram os efeitos do cuidado e da educação durante a primeira infância sobre o 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 SAEB aplicados nos alunos do 5º ano do ensino fundamental no Brasil.

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. 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*. 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 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. 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.

Pode-se concluir que 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 de notas dos testes. Logo, a primeira infância prepara as crianças para a vida na escola, pois 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

- 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.
- 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, 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.
- 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, 2009a.
- 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, 2009b.
- 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.].

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**. 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.

CAMILI, 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**: working paper nº 821. 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.

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

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.

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.

INEP, I. N. DE E. E P. E. A. T. **Sistema de Avaliação da Educação Básica-Edição 2015-Resultados**. Disponível em: <http://portal.inep.gov.br/artigo/-/asset_publisher/B4AQV9zFY7Bv/content/inep-apresenta-resultados-do-saeb-prova-brasil-2015/21206>. Acesso em: 22 nov. 2018a.

INEP, I. N. DE E. E P. E. A. T. **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. 2018b.

INEP, I. N. DE E. E P. E. A. T. **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. 2018a.

INEP, I. N. DE E. E P. E. A. T. **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. 2018b.

INEP, I. N. DE E. E P. E. A. T. **Relatório do 2 ciclo de monitoramento das metas do plano nacional de educação-2018**. Brasília-DF: [s.n.].

INEP, I. N. DE E. E P. E. A. T. **Relatório SAEB (ANEB e ANRESC) 2005 - 2015: panorama da década**. Brasília-DF: [s.n.].

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.].

MEC, M. DE E. **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.

CENTER ON THE DEVELOPING CHILD. **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.

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.

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.