

O EFEITO DA EDUCAÇÃO INFANTIL SOBRE O DESEMPENHO ESCOLAR MEDIDO EM EXAMES PADRONIZADOS

Fabiana de Felício
Banco Itaú
fabiana.felicio@inep.gov.br

Lígia Vasconcellos
Instituto Nacional de Estudo e Pesquisas Educacionais
'Anísio Teixeira' – Ministério da Educação
ligia.vasconcellos@itau.com.br

Resumo

Este estudo propôs-se a investigar o efeito de ter freqüentado a educação infantil sobre o desempenho escolar em matemática, observado na quarta série do ensino fundamental da rede pública, aplicando metodologias e dados diferenciados dos já existentes na literatura. No primeiro exercício foram utilizados dados do SAEB 2003 e, aplicando métodos de pareamento (*propensity score matching*) para resolver potenciais problemas de seleção. Os resultados encontrados evidenciam que ingressar na escola antes da primeira série tem efeito positivo e significativo sobre o desempenho na avaliação. Esse efeito é de, em média, cerca de 28% do desvio padrão do SAEB 2003, em matemática. Na estimação utilizando um painel de escolas do SAEB 2003 e Prova Brasil 2005 em que foi aplicado o método de efeitos fixos para corrigir possíveis problemas de endogeneidade, os efeitos encontrados são um pouco maiores. Na estimação com os dados de todo o Brasil obteve-se que freqüentar a EI eleva a proficiência média em cerca de 39% do desvio padrão. Algumas simulações mostraram que ainda seria possível elevar em mais de 11% a proficiência se 100% das crianças freqüentassem a educação infantil. Esses resultados confirmam a existência de uma relação de causalidade entre a freqüência à educação infantil e os resultados escolares e podem ser importantes para subsidiar políticas educacionais.

Palavras-Chave: Educação; educação infantil; desempenho escolar.

Abstract

This paper investigates the preschool attendance impact on test scores performance on mathematics observed in the 4th grade of public primary schools. It applies different methodologies and data as already used in the literature. The first exercise uses data from SAEB 2003 and propensity score matching methodology to solve for potential selection bias problem. The estimated results show that starting school earlier than 1st grade has a positive and significant effect on test scores. Measured by the standard-errors, this effect is equivalent to 28% of the standard-error of the SAEB 2003 test scores. A panel data from SAEB 2003 and Prova Brasil 2005, using the fixed effect estimation to solve for possible endogeneity problems, encounters slightly higher point-estimates. For Brazil preschool attendance increases the mean test scores in 39% of the standard-error. Several simulations show the test scores could increase by 11% if 100% of the children would attend preschool. These results confirm the existence of causality between preschool attendance and school performance and may be important to subsidy educational policies.

Keyword: education; preschool education; school performance

Classificação JEL: I21; J13

Área: Área 11 - Economia Social e Demografia Econômica

1. Introdução

Com o objetivo de compreender os fatores que podem estar relacionados ao sucesso na vida escolar e no mercado de trabalho, um conjunto de estudos recentes tem investigado a importância dos cuidados na primeira infância e seus efeitos sobre resultados na vida futura das crianças. É nessa linha que este estudo se insere buscando contribuir com novos dados e metodologias que tragam mais segurança na interpretação dos resultados encontrados para o Brasil relativamente ao efeito de freqüentar a escola antes dos sete anos sobre o aproveitamento escolar medido por meio de exames padronizados.

Diversos estudos já investigaram o que é possível fazer dentro das escolas para melhorar os resultados escolares e, posteriormente, o desempenho no mercado de trabalho, seja por meio de melhora da infra-estrutura escolar, professores e diretores melhor qualificados e outras atividades oferecidas nas escolas¹.

Também preocupada com o desempenho escolar e resultados futuros dos indivíduos, a literatura investiga o que é possível fazer pelas crianças na primeira infância - como alimentação, saúde e atividades educativas - para que elas obtenham um melhor aproveitamento na escola e em outras atividades.

Cunha *et al* (2005) apresentam modelos de desenvolvimento infantil, e mostram que um ambiente adverso ao desenvolvimento de habilidades pela criança pode ser compensado por intervenções no início do desenvolvimento infantil. Mais que isso, as habilidades adquiridas em um período precoce aumentam a retorno do investimento no período seguinte; e ações tardias são custosas e pouco eficientes.

Feinstein *et al* (1998) mostram, a partir de uma amostra de crianças na Inglaterra, que os melhores resultados aos 7 e aos 11 anos de idade são alcançados se a criança passa o tempo com sua mãe ou em uma pré-escola (ficar aos cuidados de terceiros de forma informal não traria melhores resultados). A entrada cedo na escola não seria, portanto, pré-condição de desenvolvimento de habilidades para todas as crianças, mas crianças que não tem o acompanhamento contínuo da mãe ganhariam muito com a freqüência à escola. Este resultado vale mesmo após controles para educação dos pais, classe social, interesse dos pais na educação dos filhos, e vizinhança.

¹ Sobre esse tema ver Coleman (1966), Hanushek (1986), e Heckman, Layne-Farrar e Todd (1996), entre outros, para conhecer os resultados da literatura internacional, Rivkin, Hanushek e Kain (2005) com os resultados mais recentes internacionais e Ferrão *et al* (2001), Albernaz, Ferreira e Franco (2002), Macedo (2004) para resultados de estudos aplicados aos dados brasileiros.

Com dados de um programa de construção de escolas na Argentina, Berlinski *et al* (2006) usam esta informação para controle de variáveis não-observadas e concluem que um ano de pré-escola aumenta o resultado em testes de aptidão em 8% e melhora o comportamento das crianças, medido por atenção, esforço, participação na classe, disciplina.

Conforme levantado no relatório do Banco Mundial, “Brazil Early Child Development: A Focus on the Impact of Preschools” (2001), vários estudos sobre experiências internacionais trazem evidências de que os cuidados na primeira infância são fatores relevantes na determinação de melhores condições de nutrição e saúde; melhor desempenho em testes de inteligência; maior taxa de atendimento escolar; menor taxa de repetência; menor taxa de evasão; e ainda maior participação das mulheres na força de trabalho.

Desse mesmo relatório constam estimações, para dados do Brasil, do impacto da pré-escola sobre o desempenho escolar, participação no mercado de trabalho e saúde. Os resultados significativos encontrados são de efeito positivo sobre a escolaridade média, redução de repetência, aumento da probabilidade de conclusão das etapas de ensino e aumento do rendimento futuro dos homens.

Ainda para o Brasil, Cafiero *et al* (2007), avaliando o processo de alfabetização em Minas Gerais, mostram que há uma diferença estatisticamente significativa de desempenho dos alunos que freqüentaram a fase introdutória (entrada aos 6 anos de idade).

O estudo de Curi e Menezes-Filho (2006) investiga o efeito da creche e da pré-escola sobre a probabilidade de conclusão das etapas de ensino e sobre os salários futuros, a partir dos dados da PPV, e sobre o desempenho escolar, por meio do SAEB-Inep. Ter freqüentado a creche e/ou a pré-escola tem efeitos positivos e significativos em todos os casos.

A pergunta a ser respondida por este estudo é ‘qual o efeito de ter ingressado na escola antes dos sete anos de idade sobre o desempenho escolar em matemática observado na quarta série do ensino fundamental da rede pública?’. A proposta deste trabalho é contribuir para a literatura na investigação de quão importante é a creche e/ ou a pré-escola para o desempenho escolar futuro dos estudantes.

Aprofunda-se aqui esta investigação utilizando novos dados (SAEB de 2003, Prova Brasil de 2005, e Censos Escolares de 2003 e 2005) e métodos de estimação que tratam problemas de seleção e de variáveis não-observadas (Pareamento e Efeito-Fixo), que podem ocorrer nas análises já existentes, feitas com base em dados *cross-section*. Desta forma, temos condições de confirmar a existência de uma relação de causalidade entre a freqüência à educação infantil e os resultados escolares.

A preocupação em buscar estimativas mais precisas sobre o tema justifica-se pela observação do aumento dos incentivos ao ingresso das crianças na escola antes dos seis anos, porém com taxas de

crescimento do atendimento entre 0 e 6 anos ainda modestas como poderá ser observado na próxima sessão.

Este estudo está dividido em três sessões além desta Introdução. Na segunda sessão estão apresentados os dados e metodologias utilizados. Os resultados são discutidos na terceira e na quarta sessão estão as considerações finais.

2. Dados e Metodologia

Para avaliar o impacto de freqüentar a educação infantil, creche e pré-escola, sobre o desempenho escolar da quarta série do ensino fundamental medido a partir de exames padronizados, foram utilizadas três bases de dados, Censo Escolar, Sistema de Avaliação da Educação Básica – SAEB e Prova Brasil, todas elaboradas pelo Instituto Nacional de Estudos e Pesquisas Educacionais “Anísio Teixeira” – Inep/MEC e apresentadas abaixo.

2.1 Dados

Censo Escolar

O Censo Escolar é preenchido pelas escolas ou secretarias de educação e coletado anualmente pelo Inep. Dele constam, desde 1992, as matrículas realizadas na educação básica e o movimento das matrículas entre os anos (transferências, aprovação, reprovação e abandono). A partir de 1997, a coleta foi enriquecida com a distribuição das matrículas pelas características dos estudantes e diversas questões sobre infra-estrutura escolar e participação em programas do governo federal.

SAEB

O Sistema Nacional de Avaliação da Educação Básica (SAEB) foi aplicado pelo Inep em 1990 a uma amostra de escolas representativas das redes pública e privada em todo o país, e para cada unidade da federação, produzindo informações sobre a realidade educacional brasileira. Essa avaliação foi repetida em 1993 e, desde então, o SAEB tornou-se um exame bienal de proficiência, em matemática e em Língua Portuguesa (leitura), aplicado em amostras de alunos de 4^a e 8^a séries do ensino fundamental e da 3^a série do ensino médio. Além da avaliação, são aplicados questionários que investigam questões sobre o *background* familiar, sobre a escola, os professores e o diretor.

Desde 1995, é utilizada a técnica de medição do desempenho dos alunos é baseada na Teoria de Resposta ao Item (TRI). Uma das grandes vantagens da TRI sobre a Teoria Clássica das Medidas é que a

primeira nos permite comparar o desempenho entre populações desde que submetidas a provas que tenham alguns itens em comuns, ou ainda, entre indivíduos da mesma população que tenham sido submetidos a provas diferentes. Portanto, é possível comparar o desempenho dos alunos ou das escolas brasileiras ao longo dos anos entre 1995 e 2005, e tentar diagnosticar quais os fatores relevantes à qualidade da educação.

Prova Brasil

Foi aplicada em 2005 uma nova avaliação, a Prova Brasil, uma avaliação em larga escala, como o SAEB, mas desta vez em caráter universal com resultados para todas as escolas públicas e urbanas com mais de 30 alunos nas séries avaliadas - quarta e oitava séries do ensino fundamental.

Outra diferença entre o SAEB e Prova Brasil é que para a última foram reduzidos os questionários a um único com questões referentes aos estudantes e às suas condições sócio-econômicas, o que tornou necessária a utilização do Censo Escolar conjuntamente à Prova Brasil para complementar, ainda que não perfeitamente, as informações referentes às escolas.

As proficiências estimadas na Prova Brasil são comparáveis aos resultados do SAEB, pois as duas avaliações utilizam mesma metodologia, itens comuns e estão divulgados na mesma escala (0 a 500 pontos). Isso viabiliza a utilização das duas bases em conjunto.

2.2 Estatísticas Descritivas e Projeções

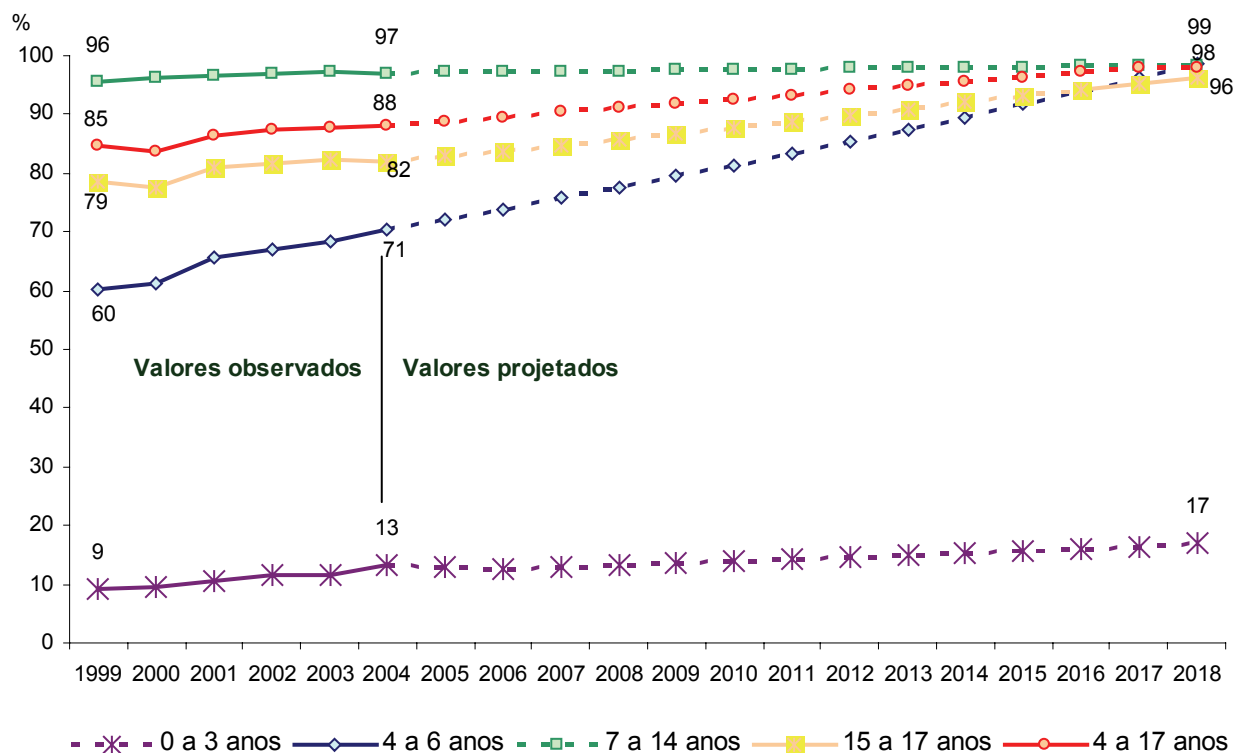
A expansão do atendimento e o aumento do investimento em educação infantil (EI) no mundo – saúde, nutrição, educação – refletem a preocupação com essa etapa da vida na qual se dá, conforme estudos médicos e mais recentemente econométricos, uma fase fundamental do desenvolvimento da criança que pode ser decisivo para seu futuro.

No Brasil essa tendência não é diferente. Recentemente foi aumentado em um ano o tempo mínimo para conclusão do ensino fundamental (passando de oito para nove anos), o que reduziu a idade obrigatória para ingresso na escola para os seis anos. Além disso, trabalha-se para incentivar o aumento de oferta de vagas na creche e na pré-escola e para elevar o ingresso na escola das crianças em idade entre 0 e 6 anos. Para isso incluiu-se, por exemplo, a educação infantil no FUNDEB (Fundo de Valorização do

Magistério e Manutenção da Educação Básica)², o que significa garantir um valor mínimo, dentro de cada estado, de recursos para cada criança matriculada na educação infantil.

De acordo com os dados da PNAD³, apresentados no Gráfico 1, entre 1999 e 2004, a taxa de atendimento das crianças de 0 a 3 anos (creche) cresceu de 9,2 para 13,4% (aumento de 46%) e no caso das crianças de 4 a 6 anos (pré-escola) a cobertura passou de 60 para 71% (aumento de 18%).

Gráfico 1. Expansão do Atendimento na Educação Infantil (1999 a 2004) e projeções seguindo a tendência dos últimos cinco anos (2005 a 2018)



FONTE: Dados PNAD e CENSO 2000 – IBGE, Tabulação e Projeção Inep/MEC

Se a tendência observada nos últimos cinco anos se mantiver constante, no ano de 2014 (dez anos após o último dado observado) o atendimento da pré-escola será de 89,5%, mas o da creche ainda será de 15% o que é muito baixo se considerarmos que freqüentar a escola nesse período pode ser importante no desenvolvimento futuro das crianças.

Outra informação interessante sobre o tema pode ser observada nos questionários preenchidos pelos alunos que fazem o SAEB. Nos anos de 2003 e 2005 foi incluída uma questão em que se pergunta ao

² Fundo que substituiu, em 2007, o antigo FUNDEF (Fundo de Valorização do Magistério e Manutenção do Ensino Fundamental).

³ PNAD – Pesquisa Nacional por Amostra de Domicílios, elaborada pelo Instituto Brasileiro de Geografia e Estatística - IBGE.

estudante “QUANDO VOCÊ COMEÇOU A ESTUDAR?” e as opções oferecidas são (a) *No maternal*; (b) *Na Pré-escola*; (c) *Na primeira série*; (d) *Na segunda série*; (e) *Na terceira série*.

No caso da Prova Brasil a questão é “QUANDO VOCÊ ENTROU NA ESCOLA?”. As respostas possíveis são (a) *No maternal (jardim da infância)*; (b) *Na Pré-escola*; (c) *Na primeira série*.

Pode-se observar na Tabela 1 a distribuição das crianças pela série em que ingressou no ensino fundamental. Para analisar as respostas é preciso levar em consideração que alunos que respondem aos questionários do SAEB e Prova Brasil na quarta série têm entre 10 e 12 anos e que, provavelmente na maioria dos casos têm dificuldade de diferenciar maternal/ creche do que é pré-escola.

Tabela 1. Distribuição dos alunos nos níveis educacionais em que pode ter entrado na escola*

	SAEB 2003 (escolas urbanas e públicas)			Prova Brasil 2005		
Maternal	23,59			36,44		
Pré-escola	52,19			37,95		
1ª série	24,22			25,41		
Não informado	0,00			0,20		
	Distribuição por escolaridade da mãe					
	Até 8 anos	De 9 a 11	12 ou mais	Até 8 anos	De 9 a 11	12 ou mais
Maternal	17,88	30,13	39,98	31,57	41,39	49,38
Pré-escola	52,39	53,46	47,57	39,33	38,33	32,68
1ª série	29,73	16,41	12,45	28,90	20,14	17,79

(*) Dados do SAEB e Prova Brasil - Inep, referentes às escolas públicas e urbanas. Tabulação própria.

A Tabela 1 apresenta ainda as respostas dos estudantes de 4ª série sobre o ingresso à escola, distribuídos por escolaridade das mães. O intuito aqui é mostrar que existe um atendimento desigual na educação infantil o que pode acentuar as desigualdades no caso da EI ser eficaz na elevação do desempenho futuro de seus alunos.

Esses dados justificam uma preocupação em verificar o efeito da educação infantil para subsidiar políticas voltadas à expansão da creche e pré-escola, especialmente voltada a atender filhos de famílias menos favorecidas.

2.3 Métodos e Amostra

Amostra I e Métodos de Pareamento

As primeiras estimativas utilizam os dados do SAEB de 2003 ao nível dos alunos. O banco de dados que consideramos inclui apenas escolas públicas, estaduais e municipais, em áreas urbanas. São 28.832 observações de alunos.

As variáveis de interesse (tratamento) para estimarmos o impacto sobre a proficiência em matemática na 4ª série são:

Educação infantil:	iniciou escola na creche ou pré-escola
Creche:	iniciou na creche
Pré-escola:	iniciou na pré-escola

O impacto sobre o desempenho no SAEB será estimado em relação a um grupo de controle de crianças que iniciou a escola na 1ª série (crianças que declararam terem entrado na escola na 2ª ou 3ª séries foram reclassificadas para entrada na 1ª série). Para medir impactos diferenciados entre entrada na creche e na pré-escola, consideramos também a variável de tratamento “iniciou na creche” comparada ao grupo de controle que iniciou na pré-escola. Para as estimativas consideramos o logaritmo da proficiência como variável dependente.

Acreditamos que crianças que iniciaram na creche provavelmente continuaram freqüentando a educação infantil e, portanto, também freqüentaram uma pré-escola. Desta forma, é esperado que o impacto da creche sobre quem iniciou os estudos na 1ª série seja igual ou maior do que o impacto apenas da pré-escola.

Serão apresentados os impactos estimados por mínimos quadrados ordinários e por pareamento. O segundo método leva em consideração a possibilidade de haver um problema de seleção relacionado à freqüência de educação infantil.

A análise de impacto pretende resolver o problema de não termos o contrafactual para a seguinte questão: como teria sido o desempenho de matemática das crianças de 4ª série caso elas não tivessem freqüentado uma pré-escola ou creche? A estratégia é determinar, dentre as crianças que iniciaram na 1ª série, as mais “parecidas” com as que cursaram pré-escola ou creche. Este processo é realizado com base em variáveis explicativas observadas que, acreditamos, possam influir nos resultados escolares e estar correlacionadas com a freqüência a educação infantil (tratamento).

Os questionários que acompanham o SAEB trazem perguntas contemporâneas, e não retrospectivas. Para o pareamento precisamos de variáveis que expliquem tanto a freqüência a educação infantil, ocorrida

pelo menos 4 anos antes da pesquisa, como a proficiência na 4ª série. Partimos da hipótese que as condições sócio-econômicas relatadas pelas crianças não se alteraram neste período. As variáveis explicativas incluem características familiares e pessoais, condição sócio-econômica da família, e características relacionadas à oferta e qualidade do ensino. Estas variáveis foram compiladas dos questionários respondidos por alunos, professores e diretores, e estão apresentadas no anexo A.

Formalmente, sejam dois resultados potenciais (Y_0, Y_1), onde 1 indica resultado com tratamento. Seja $D = 1$ se ocorre tratamento e $D = 0$ caso contrário. Sejam X variáveis observadas que determinam a participação no tratamento e seu resultado. Nós restringiremos nossa análise ao efeito-tratamento sobre tratados (ATT), ou seja, focamos no diferencial de resultado obtido pelas crianças que efetivamente freqüentaram pré-escola.

Podemos escrever o ATT como:

$$E(Y_1 - Y_0 | X, D=1) \quad (1)$$

Como não temos o contrafactual $E(Y_0 | X, D=1)$, o problema de seleção surge quando queremos utilizar a diferença das médias amostrais observadas para estimar o efeito do tratamento:

$$E(Y_1 | X, D=1) - E(Y_0 | X, D=0) \quad (2)$$

No caso do ATT, o viés gerado equivale à diferença entre (1) e (2):

$$\begin{aligned} & [E(Y_1 | X, D=1) - E(Y_0 | X, D=0)] - [E(Y_1 - Y_0 | X, D=1)] = \\ & = E(Y_0 | X, D=1) - E(Y_0 | X, D=0) \end{aligned} \quad (3)$$

As hipóteses de identificação, ou seja, as hipóteses que garantem que a diferença em (3) é nula, são:

- (a) $Y_0 \perp D | X$, isto é, independência de Y_0 em relação a D dado X ,
- (b) $0 < \Pr(D=1 | X) = P(X) < 1$, isto é, existem observações tratadas e não-tratadas.

Rosenbaum e Rubin (1983) mostram que, dados, (a) e (b), também vale:

- (c) $Y_0 \perp D | P(X)$, o que reduz a dimensão necessária para resolver o pareamento.

A hipótese (a) poderia ser simplificada para a hipótese de independência de médias, $E(Y_0 | X, D=0) = E(Y_0 | X)$. Neste caso, porém, a condição (c) não segue imediatamente, precisando ser considerada uma hipótese adicional.

Se, por um lado, a hipótese de identificação pode ser forte em casos em que nem todas as variáveis relevantes são observáveis, o *pareamento*, como observa Heckman e Navarro-Lozano (2003), tem a vantagem de não exigir exogeneidade das variáveis explicativas (se o *pareamento* é bom, as variáveis não observadas relevantes estão igualmente distribuídas nos grupos de tratados e de controle, e seu efeito, portanto, se cancela) e não exigir restrição de exclusão (que entre as variáveis que explicam a participação no programa haja variáveis que não são correlacionadas com o resultado escolar). A hipótese de identificação depende, portanto, de não haver variáveis não-observadas que afetem os resultados das crianças de forma diferenciada nos grupos de tratamento e controle.

O uso de $P(X)$ em vez de X é conhecido como método de *propensity score*. Estimamos as probabilidades (*propensity scores*) com um modelo *probit*, e escolhemos as crianças “pareadas” (ou *matched*) utilizando um estimador *kernel* de uma regressão linear local. Para garantir que as crianças são realmente pareadas, utilizamos um suporte comum de probabilidades, tal que apenas as crianças não-tratadas com *propensity score* dentro da faixa de *propensity score* das crianças tratadas são consideradas.

Escolhido o grupo de controle, e sendo bom o *pareamento* realizado, as médias dos resultados escolares dos dois grupos poderiam ser diretamente comparadas. Preferiu-se também utilizar uma regressão de mínimos quadrados ponderados pelo *propensity score*, incluindo apenas o suporte comum de probabilidades e controlando pelas variáveis explicativas observadas. Como as variáveis relacionadas à escola que explicam desempenho são posteriores à decisão do tratamento, o primeiro estágio do *pareamento* pode não ser suficiente para eliminar todo o viés de seleção.

Para definir se o *pareamento* foi bem feito comparamos os histogramas das probabilidades estimadas entre tratados e pareados. Esta comparação deve mostrar distribuições similares no caso de o *pareamento* ter sido bem feito. Foram estimadas densidades *kernel*.

Para verificar a robustez dos resultados estimados apresentamos também outro método de estimação, o método de efeito fixo.

Amostra II e Painel de Dados

Para o segundo exercício foi utilizado um painel de dados de dois anos por escola em que foram unidos e compatibilizados os dados do SAEB 2003 e da Prova Brasil 2005 e, aos dados das escolas que constavam de ambas avaliações, foram adicionados os dados do Censo Escolar 2003 e 2005, respectivamente. Da união das duas avaliações resultaram 1.335 escolas observadas nas duas avaliações, todas públicas e urbanas devido à restrição ao universo da Prova Brasil.

A vantagem de se utilizar a Prova Brasil de 2005 e não o SAEB 2005 se deve ao fato da Prova Brasil ser universal e, portanto, contar com um número de escolas comuns ao SAEB 2003 muito maior que o SAEB 2005, que tem uma amostra definida de maneira independente da amostra de 2003.

Devido a uma diferença entre as respostas dadas à questão sobre ingresso na escola no SAEB 2005 e na Prova Brasil 2005 (não apresentado), as respostas de maternal e pré-escola foram trabalhadas em conjunto neste segundo exercício, ou seja, a variável de interesse representa a proporção de estudante que ingressou na escola antes da 1ª série.

Foram estimadas regressões em painel com dados de 2003 (SAEB e Censo Escolar) e de 2005 (Prova Brasil e Censo Escolar), agregados por escola. Aplicou-se o método de efeitos fixos para fazer o controle por características não observáveis que podem ser efeitos específicos de cada escola correlacionados à probabilidade de ingressar na escola antes dos sete anos e aos resultados escolares.⁴ Dessa forma, resolvemos de outra forma o problema de endogeneidade. A hipótese de identificação subjacente é que as variáveis não-observadas relativas diretamente às escolas, mas também relativas às características médias de seus alunos, são efetivamente fixas no tempo.

A variável a ser explicada é o logaritmo da média das escolas na avaliação de matemática. Da mesma forma que no exercício anterior, a variável de interesse refere-se à entrada na escola, no entanto, neste segundo exercício, ela é a proporção de alunos que respondeu ter ingressado no maternal ou pré-escola⁵.

Também foram incluídas nas estimações variáveis de controle que foram retiradas dos questionários dos alunos das duas pesquisas, além de informações sobre professores e infra-estrutura escolar, obtidas no Censo Escolar. O modelo estimado pode ser representado conforme equação abaixo.

$$Y_{it} = \alpha_{it} + I_{it}\beta_1 + A_{it}\beta_2 + E_{it}\beta_3 + v_i + u_{it} \quad (4)$$

$i = 1, 2, \dots, 1.335^6$

$t = 2003; 2005$

Em que:

Y_{it} - logaritmo do desempenho médio em matemática da escola i no ano t ;

I_{it} - proporção de crianças que ingressou na escola antes da primeira série

⁴ Foi realizado o Teste de Hausman para verificar qual método de estimação seria o mais apropriado para o painel de escolas analisado, efeito aleatório ou efeito fixo. O resultado do teste confirmou a hipótese de que os efeitos específicos não observados estariam correlacionados com as variáveis explicativas e a proficiência.

⁵ Assim como no primeiro exercício, as crianças que declararam terem entrado na escola na 2ª ou 3ª séries, opções existentes apenas no SAEB, foram reclassificadas para entrada na 1ª série.

⁶ O número de escolas públicas incluídas no painel.

A_{it} - vetor de variáveis com características que representam o histórico familiar e escolar dos alunos da escola i no ano t ;
 E_{it} - vetor de variáveis com características da escola i no ano t ;
 α_{it} - constante;
 V_i - denota o efeito específico não observável de cada escola i ;
 u_{it} - termo aleatório.

Feitas as estimações por efeitos fixos, foram simuladas quais seriam as proficiências médias que seriam obtidas caso todas as características se mantivessem constantes exceto a variável de interesse, ou seja, simulando as seguintes situações: (i) nenhuma criança ingressou antes da 1ª série – EI=0%; (ii) todas as crianças ingressaram na educação infantil – EI=100%. Esses resultados foram comparados ao resultado estimado para a proporção observada de crianças que declarou ter ingressado na educação infantil e estão apresentados na sessão seguinte.

Além disso, para fim de comparação também foi feita a estimação, para o mesmo banco de dados, utilizando o método de MQO, com mesma especificação, incluindo apenas uma variável *dummy* de Ano 2005 e quatro variáveis *dummy* de macrorregião. Neste caso, serão apresentados na sessão 3 apenas os coeficientes de interesse.

Por fim, a mesma metodologia foi aplicada separadamente para cada região geográfica brasileira para obter um melhor ajuste da estimação e avaliar possíveis diferenças regionais no efeito da EI.

3. Resultados

3.1 Pareamento

Os gráficos abaixo mostram o resultado do pareamento para a frequência em educação infantil. O primeiro gráfico mostra a distribuição de probabilidades estimadas para o grupo de controle antes e depois do pareamento, e o segundo compara, depois do pareamento, a distribuição de probabilidades dos grupos de tratamento e controle. Vemos que o pareamento foi bom, praticamente igualando as distribuições dos dois grupos. Os gráficos para creche e pré-escola em separado (não apresentados) também mostram um bom pareamento.

Educação infantil: Probabilidades estimadas para grupo de tratamento e de controle

Gráfico 2. Grupo de controle: antes e depois do pareamento

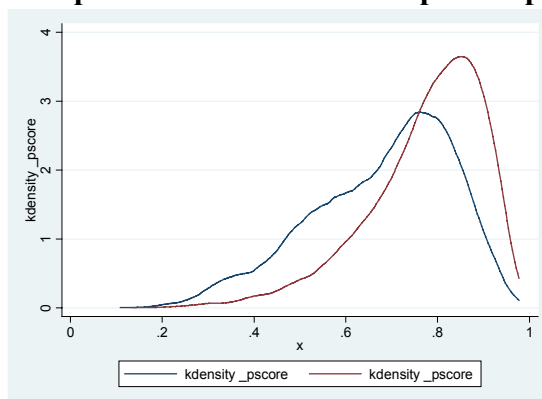
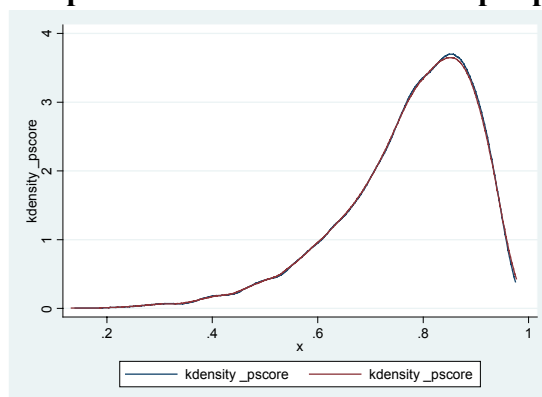


Gráfico 3. Grupo de tratamento e controle: após pareamento



A tabela 2 traz o número de observações usadas no pareamento, praticamente todas as observações entram no suporte comum de probabilidades.

Os resultados de impacto da frequência da educação infantil sobre proficiência em matemática na 4ª série são todos significativos⁷, como vemos na Tabela 3. Apresentamos os resultados por MQO e por pareamento. Por pareamento, apresentamos apenas a diferença de médias entre o grupo de tratamento e o grupo pareado, as estimativas após controle de observadas no 2º estágio (nosso resultado preferido), e, para efeito de comparação com os resultados por efeito fixo (apresentados na próxima seção), com uma lista de variáveis explicativas mais restrita (variáveis também disponíveis na Prova Brasil e variáveis fixas no tempo).

A Tabela 3 abaixo traz os resultados de impacto. As estimativas por pareamento incluindo todas as variáveis observadas disponíveis (coluna C) mostram que ter frequentado a educação infantil aumenta em 6,6% a média da proficiência em matemática dos estudantes de 4ª série. Quando consideramos

⁷ Seguindo Imbens (2004), fizemos também estimativas usando mínimos quadrados ponderados pela recíproca de probabilidade do tratamento, e chegamos a estimativas similares, apesar de o pareamento ser pior.

separadamente creche e pré-escola, a primeira aumenta a proficiência em 8,6%, e a segunda, em 6,3%. Comparando-se creche e pré-escola diretamente, medimos impacto de 1,0%. Todos os resultados são significativos a 5%.

Tabela 2. Suporte comum de probabilidades

	tratados	controle	fora do suporte
Infantil	17.367	5.841	13
Creche	5.749	5.841	18
Pré-escola	11.602	5.841	11
pré-escola vs creche	5.763	11.613	4

Tabela 3 – Estimativas por pareamento (ATT)

Impacto (log da proficiência em matemática)	A	B	C	D
	MQO	ATTsem	ATT/MQP	ATTpai
Infantil	0,067	0,068	0,066	0,080
p-valor	0,000	0,000	0,000	0,000
Creche	0,085	0,083	0,086	0,096
p-valor	0,000	0,000	0,000	0,000
pré-escola	0,066	0,062	0,063	0,069
p-valor	0,000	0,000	0,000	0,000
pré-escola vs creche	0,015	0,008	0,010	0,005
p-valor	0,011	sign. 95%	0,013	0.233

Notas: ATT – efeito médio do tratamento sobre tratados; B – diferença de médias, sem controles no 2º estágio; C – mínimos quadrados ponderados com controles no segundo estágio; D – menos controles, similar ao painel. Todos coeficientes significativos a 5%.

Como vemos, é pequena a diferença entre os resultados estimados por MQO e os resultados por pareamento. Este resultado indica que não há problema de seleção na definição de quem frequenta ou não a educação infantil (lembramos que não entraram em nossa análise as escolas particulares e escolas em áreas rurais), o que explicaria os resultados semelhantes.

3.2 Painel de Dados

Resultados das estimações em Painel de Escolas

Na Tabela 4, estão apresentados os coeficientes da variável proporção de crianças que ingressou na escola antes da primeira série, estimados para o painel de escolas do SAEB 2003 e Prova Brasil 2005 pelo método de efeito fixo. Também estão apresentados os coeficientes de MQO para efeito de comparação.

Tabela 4. Coeficientes estimados para EI em MQO e Efeitos Fixos

	B ₁ MQO	B ₁ Efeitos Fixos
BRASIL: Coeficiente de Ingresso na EI	0,1364*	0,0962*

<i>(Erro Padrão)</i>	<i>0,0118</i>	<i>0,0187</i>
Estimações por Macrorregiões		
N: Coeficiente de Ingresso na EI	0,0447***	0,0876**
<i>(Erro Padrão)</i>	<i>0,0267</i>	<i>0,0448</i>
NE: Coeficiente de Ingresso na EI	0,1471*	0,1102*
<i>(Erro Padrão)</i>	<i>0,0211</i>	<i>0,0325</i>
SE: Coeficiente de Ingresso na EI	0,3117*	0,1937*
<i>(Erro Padrão)</i>	<i>0,0355</i>	<i>0,601</i>
S: Coeficiente de Ingresso na EI	0,0924*	0,1074**
<i>(Erro Padrão)</i>	<i>0,0268</i>	<i>0,0514</i>
CO: Coeficiente de Ingresso na EI	0,2065*	0,1353*
<i>(Erro Padrão)</i>	<i>0,0299</i>	<i>0,0485</i>

Nota: (*), (**), (***): significante ao nível de 1%, 5% e 10%, respectivamente.

O resultado da estimação por efeito fixo, para o Brasil, é um coeficiente semelhante ao estimado pelo método de pareamento e aos demais ao nível do aluno, porém é mais baixo que o de MQO aplicado ao painel de escolas, evidência de que as estimações para dados agregados podem gerar coeficientes enviesados quando não controlados os efeitos específicos.

Na metodologia de efeitos fixos não é possível observar as diferenças regionais existentes. Para avaliar as diferenças entre as regiões do Brasil foram estimadas as mesmas especificações desta vez para cada macrorregião. Os coeficientes também estão apresentados na Tabela 4 e surpreendem para algumas regiões. Na região Sudeste observa-se o maior efeito da EI, $B_1=0,19$, e na região Centro-Oeste o coeficiente é de 0,14. Já a região Norte têm o coeficiente mais baixo, $B_1=0,09$. Uma explicação possível para a variação regional é a diferença de qualidade da EI oferecida. Outra possibilidade é que o número de anos frequentando EI seja relevante para o efeito observado no desempenho da quarta série e que o tempo médio que as crianças frequentam creche/ pré-escola seja diferente entre as regiões.

Resultados das simulações

As simulações são uma forma simples de avaliar os efeitos possíveis de serem conseguidos através de políticas de incentivo à educação infantil. Da comparação entre a proficiência média estimada para a proporção observada de alunos que frequentaram EI e o caso em que foi simulado zero ou 100% das crianças tendo ingressado antes da primeira série, é possível avaliar o potencial da educação infantil para elevar o desempenho das crianças ao chegarem à quarta série.

Gráfico 4. Proficiências estimadas variando a proporção de crianças que ingressaram na escola na Educação Infantil

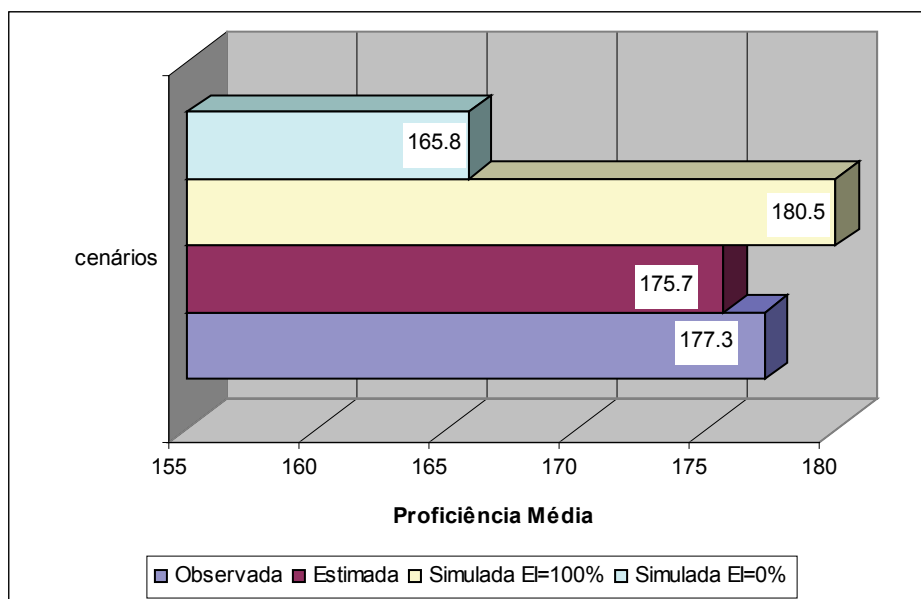
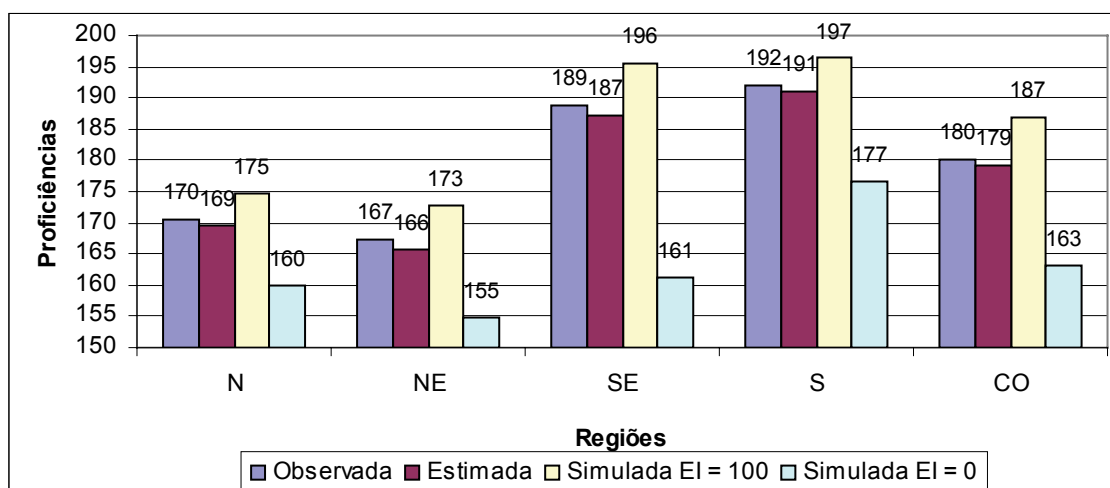


Gráfico 5. Proficiências estimadas por macrorregião variando a proporção de crianças que ingressaram na escola na Educação Infantil



É importante destacar que essa simulação não aponta o efeito máximo da EI, pois o tempo dedicado à EI pode afetar diferentemente os resultados futuros. No caso das simulações por região, por exemplo, o que se tem, provavelmente, é uma diferença de número médio de anos que as crianças frequentam a EI. Assim, estimar o máximo acréscimo de pontos do SAEB possível de ser obtido por expansão da EI, só seria possível se a estimativa fosse feita considerando o número de anos que cada criança frequentou a educação infantil e isso não é possível extrair das avaliações do Inep.

Os resultados das simulações de variação na proporção de alunos que ingressaram na escola antes da primeira série estão apresentados nos gráficos 4 e 5. No Gráfico 4, está o resultado da estimativa para o Brasil. Entre a situação simulada como se nenhuma criança tivesse frequentado a EI e a situação atual

existe uma diferença de 10 pontos do SAEB, 23%, aproximadamente, do desvio padrão (DP) em matemática no ano de 2003 (que é de 44 pontos). Entre a situação atual estimada e a simulada para 100% das crianças tendo ingressado antes da primeira série a diferença é de cinco pontos, 11% do DP.

No Gráfico 5 observam-se as variações para cada estimativa feita para as regiões brasileiras. Da comparação entre as estimativas para o quadro atual e a situação de 100% terem frequentado a EI, observa-se que o crescimento obtido é de seis pontos para as regiões Norte e Sul, de sete pontos para a região Nordeste e de oito pontos para o Sudeste e o Centro-Oeste, variação entre 14 e 18% do DP.

4. Considerações Finais

Este estudo propôs-se a investigar o efeito de ter frequentado a educação infantil sobre o desempenho escolar em matemática, observado na quarta série do ensino fundamental da rede pública, aplicando metodologias e dados diferenciados dos já existentes na literatura.

Utilizando dados do SAEB 2003 e aplicando métodos de pareamento para comparar indivíduos semelhantes que teriam mesmas probabilidades de terem frequentado a educação infantil, os resultados encontrados são de que ingressar na escola antes da primeira série tem efeito positivo e significativo sobre o desempenho em exame padronizado. Esse efeito, em média, é de 6,6% da proficiência, ou cerca de 27% do desvio padrão do SAEB 2003, em matemática.

Na estimativa utilizando um painel de escolas do SAEB 2003 e Prova Brasil 2005 em que foi aplicado o método de efeito fixo para corrigir possíveis problemas de endogeneidade existentes nas análises em *cross-section*, os resultados são um pouco maiores. Na estimativa com os dados de todo o Brasil obteve-se que frequentar a EI eleva a proficiência média das escolas em cerca de 9,6%, ou cerca de 38,6% do desvio padrão. Esses resultados variam entre 8,7 e 19,4% da proficiência média entre as regiões brasileiras, apontando, possivelmente, uma diferença de qualidade média da educação infantil oferecida nas regiões do país, ou variação na média de tempo que as crianças frequentam a EI, o que geraria impacto maior da educação infantil.

Dos resultados das simulações é possível interpretar que a EI é responsável por 6% da proficiência média obtida no SAEB 2003 e, conforme simulado, ainda pode elevar em 3% a média de matemática. Os resultados regionais são ainda mais expressivos, chegando a potenciais 4,8% de acréscimo da média na região Sudeste.

Esse estudo contribui para confirmar a existência de uma relação de causalidade entre a frequência à educação infantil e os resultados escolares. Os resultados podem ser importantes para subsidiar políticas educacionais.

Anexo A

Tabela A1. Variáveis explicativas SAEB 2003.

Variáveis (médias / proporções)	creche	pré-escola	1ª série
proficiência em matemática	178.70	175.67	160.39
aluno:			
Idade	10.66	10.83	11.47
Homem	0.47	0.50	0.53
número de pessoas	5.46	5.60	6.11
Branco	0.39	0.39	0.36
bolsa-escola	0.24	0.24	0.27
Lê nunca ou quase nunca	0.01	0.01	0.02
Lê sempre ou quase sempre	0.56	0.54	0.55
Lê de vez em quando	0.43	0.45	0.43
faz lição de casa	0.77	0.75	0.70
bens no domicílio:			
livros em casa	0.88	0.85	0.78
número de banheiros	1.24	1.11	0.99
Carro	0.39	0.35	0.25
máquina de lavar	0.60	0.63	0.54
Computador	0.19	0.13	0.09
água da torneira	0.96	0.94	0.91
Escolaridade da mãe:			
ensino primário incompleto	0.10	0.14	0.23
ensino fundamental incompleto	0.22	0.27	0.26
ensino médio incompleto	0.15	0.14	0.11
ensino médio completo	0.13	0.10	0.06
ensino superior, completo ou não	0.15	0.10	0.06
Sem informação	0.25	0.26	0.29
Escolaridade do pai:			
ensino primário incompleto	0.08	0.13	0.17
ensino fundamental incompleto	0.15	0.19	0.18
ensino médio incompleto	0.11	0.11	0.10
ensino médio completo	0.11	0.08	0.05
ensino superior, completo ou não	0.15	0.09	0.06
Sem informação	0.40	0.40	0.43
mora com a mãe	0.91	0.90	0.86
mora com o pai	0.65	0.68	0.64
cidade grande	0.72	0.58	0.59
escola estadual	0.53	0.50	0.47
diretor da escola:			
Salário	4.21	4.28	4.17
há laboratório	0.23	0.18	0.15
há biblioteca	0.75	0.74	0.68
há problemas financeiros	0.15	0.17	0.15
há internet para alunos	0.17	0.15	0.13
tempo de experiência	1.67	1.62	1.61
professor da turma:			
usa computador	0.39	0.37	0.35

formação nos últimos 2 anos	0.88	0.87	0.87
Salário	3.10	3.10	3.04
tempo de experiência	1.84	1.84	1.80
ensino superior	0.68	0.66	0.63
número de observações	7189	14240	7403

Nota: hábito de leitura (revistas, inclusive quadrinhos, livros ou jornais), idade (8 ou menos, 9,10,11,12,13,14, 15 ou mais), faz lição de casa (sempre ou quase sempre); número de banheiros (0,1,2,3 ou mais), computador, carro, máquina de lavar, há livros; tamanho da cidade (região metropolitana ou população grande em relação a cidades pequenas), tempo de experiência na escola (menos de 2 anos, 3 a 15 anos; mais de 15 anos); salário (múltiplos do salário-mínimo: de 1 a mais de 16).

As estimativas apresentadas na coluna D utilizam a seguinte lista de variáveis explicativas: branco, carro, máquina de lavar, homem, computador, lição de casa, escolaridade da mãe e do pai, mora com a mãe, mora com o pai, laboratório, biblioteca, acesso a Internet para alunos na escola, professor com ensino superior, idade, escola estadual, cidade grande, e *dummies* de estado.

Tabela A2. Impacto do educação infantil sobre proficiência em matemática (4ª série) após pareamento

Estimativa após pareamento por regressão linear local				
log(proficiência)	coeficiente	erro-padrão	estatística t	p-valor
educação infantil	0.066	0.003	24.070	0.000
aluno:				
água da torneira	0.014	0.007	2.060	0.040
número de pessoas	-0.010	0.001	-12.770	0.000
Branco	-0.006	0.003	-1.980	0.047
escola estadual	0.003	0.003	1.130	0.260
bolsa-escola	-0.035	0.003	-10.090	0.000
Lê sempre ou quase sempre	0.065	0.014	4.520	0.000
Lê de vez em quando	0.086	0.014	5.950	0.000
livros em casa	0.032	0.004	7.990	0.000
Idade	-0.020	0.001	-16.310	0.000
cidade grande	0.005	0.003	1.680	0.093
número de banheiros	0.011	0.002	4.280	0.000
Carro	0.009	0.003	2.700	0.007
Máquina de lavar	0.003	0.004	0.890	0.375
Homem	0.028	0.003	10.040	0.000
Computador	0.013	0.004	3.000	0.003
faz lição de casa	0.040	0.003	12.010	0.000
Escolaridade da mãe:				
ensino fundamental incompleto	0.011	0.005	2.300	0.022
ensino médio incompleto	0.004	0.006	0.750	0.456
ensino médio completo	0.019	0.006	3.030	0.002
ensino superior, completo ou não	0.012	0.006	1.970	0.049
Sem informação	0.000	0.005	-0.080	0.935
Escolaridade do pai:				
ensino fundamental incompleto	0.008	0.005	1.520	0.128
ensino médio incompleto	-0.001	0.006	-0.220	0.829
ensino médio completo	0.007	0.007	1.110	0.265
ensino superior, completo ou não	-0.009	0.007	-1.370	0.171
Sem informação	0.004	0.005	0.760	0.447
mora com a mãe	0.004	0.005	0.830	0.406
mora com o pai	-0.010	0.003	-3.170	0.002
diretor da escola:				
salário	0.003	0.002	2.010	0.044
há laboratório	0.003	0.004	0.770	0.440

há biblioteca	0.004	0.004	1.120	0.264
há problemas financeiros	0.000	0.004	0.110	0.916
há internet para alunos	0.011	0.004	2.490	0.013
tempo de experiência	0.003	0.003	0.900	0.367
professor da turma:				
usa computador	-0.001	0.003	-0.260	0.791
formação nos últimos 2 anos	-0.009	0.004	-2.220	0.027
salário	0.005	0.002	3.330	0.001
tempo de experiência	0.013	0.002	5.940	0.000
ensino superior	0.010	0.003	3.100	0.002
constante	5.065	0.026	192.310	0.000

inclui dummies de estado; - MQO ponderados de acordo com pareamento definido em primeiro estágio; número de observações = 18.837; R2 ajustado = 0.194

Anexo B

Tabela B1. Estimativa com método de efeitos fixos

Variável dependente: log(proficiência)	coeficiente	erro-padrão	p-valor
Proporção que ingressou na Educação Infantil aluno:	0.0962	0.0187	0.0595
prop de brancos	-0.0150	0.0165	-0.0473
prop que tem carro	0.0933	0.0201	0.0539
prop que tem máquina de lavar	0.0319	0.0142	0.0041
prop de homens	-0.0220	0.0200	-0.0613
prop não tem lição de casa	-0.0722	0.0316	-0.1341
prop que não tem computador	-0.1112	0.0166	-0.1438
prop que tem computador com Internet	-0.1748	0.0379	-0.2493
prop que mora com pai e mãe	0.0023	0.0203	-0.0375
Escolaridade da mãe:			
prop de alunos cuja mãe tem EF incompleto	0.0396	0.0344	-0.0279
prop de alunos cuja mãe tem EM incompleto	0.0696	0.0373	-0.0035
prop de alunos cuja mãe tem ensino superior, completo ou não	0.1159	0.0495	0.0189
Sem informação	-0.0013	0.0275	-0.0553
prop de alunos cujo pai tem EF incompleto	0.0093	0.0379	-0.0650
prop de alunos cujo pai tem EM incompleto	0.0425	0.0415	-0.0388
prop de alunos cujo pai tem ensino superior, completo ou não	-0.0068	0.0481	-0.1012
Sem informação	-0.0108	0.0256	-0.0611
Professor:			
prop com ensino médio incompleto	-0.0579	0.0361	-0.1287
prop com ensino médio completo	-0.0186	0.0081	-0.0344
Escola:			
há biblioteca	0.0076	0.0060	-0.0042
há laboratório de ciências	0.0037	0.0115	-0.0188
há laboratório de informática	0.0142	0.0077	-0.0010
Constante	5.1366	0.0180	5.1013

Referências Bibliográficas

- ALBERNAZ, Ângela; FERREIRA, Francisco H. G.; FRANCO, Creso. *Qualidade e Equidade na Educação Fundamental Brasileiro*. PPE, v. 33 No.3. 2002.
- BANCO MUNDIAL. Brazil Early Child Development: A focus on the impact of preschools. **Relatório No. 22851-BR**. set 2001.
- BERLINSKI, S., GALIANI, S., GERTLER, P., The effect of pre-primary education on primary school performance fev 2006 (**mimeo**).
- CAFIERO, D, ROCHA, G. e SOARES, F. Avaliação do ciclo inicial de alfabetização em Minas Gerais: o que indicam os primeiros resultados, **Língua Escrita**, Belo Horizonte, n.1, jan/abr 2007
- COLEMAN, James S. et al. *Equality of Educational Opportunity*. Washington, 1966.
- CUNHA, F., HECKMAN, J., LOCHNER, L., MASTEROV, D., *Interpreting the evidence on life cycle skill formation*, **Working Paper** 11331, NBER Working Paper Series mai 2005
- CURI, A., MENEZES-FILHO, N., Os efeitos da pré-escola sobre os salários, a escolaridade e a proficiência escolar. **Anais da ANPEC 2006**.
- DEHEJIA, R., WAHBA, S. Causal effects in non-experimental studies: re-evaluating the evaluation of training programs. **NBER Working Paper 6586**. 1998.
- FEINSTEIN, L. ROBERTSON, D. e SYMONS, J. *Pre-school Education and Attainment in the NCDS and BCS*. Março de 1998. **CEP Discussion Paper**.
- FERRÃO, M. E. et al. O SAEB – Sistema Nacional de Avaliação da Educação Básica: objetivos, características e contribuições na investigação da escola eficaz. **Revista Brasileira de Estudos de População**, v. 18, n.1/2, jan/dez, 2001.
- HANUSHEK, Eric. The Economics of Schooling: Production and Efficiency in Public Schools. *Journal of Economic Literature*, v. 24, No. 3, p 1141-1177, 1986.
- HECKMAN, James; LAYNE-FARRAR, Anne; TODD, Petra. *Does Measured School Quality Really Matter? An Examination of the Earnings-Quality Relationship*. In: G. Burtless (Ed.) **Does Money Matter? The Effect of School Resources on Student Achievement and Adults Success**. Washington, DC: Brooking Institution Press, p192-289, 1996.
- HECKMAN, J., NAVARRO-LOZANO. S. Using matching, instrumental variables and control functions to estimate economic choice models. **NBER Working Paper 9497**. Fev 2003.
- IMBENS, G. Nonparametric estimation of average treatment effects under exogeneity: a review. **The Review of Economics and Statistics**, 86 (1) fev 2004
- Instituto Nacional de Estudos e Pesquisas Educacionais ‘Anísio Teixeira’, do Ministério da Educação – Inep/MEC. Site para consulta das informações www.inep.gov.br
- MACEDO, G. A. *Fatores associados ao rendimento escolar de alunos da 5ª série (2000) – uma abordagem do valor adicionado*. Trabalho apresentado no **XIV Encontro Nacional de Estudos Populacionais**, ABEP, realizado em Caxambú-MG – Brasil, de 20- 24 de Setembro de 2004.
- RIVKIN, Steven G.; HANUSHEK, Eric A.; e KAIN, John F. *Teachers, Schools, and Academic Achievement*. **Econometrica**. vol. 73, No.2, p417-58. 2005.

ROSENBAUM, P., RUBIN. D. The central role of the propensity score in observational studies for causal effects. **Biometrika** 70:1. 1983.

WOOLDRIDGE, J. Econometric Analysis of cross section and panel data. **The MIT Press**, 2002.