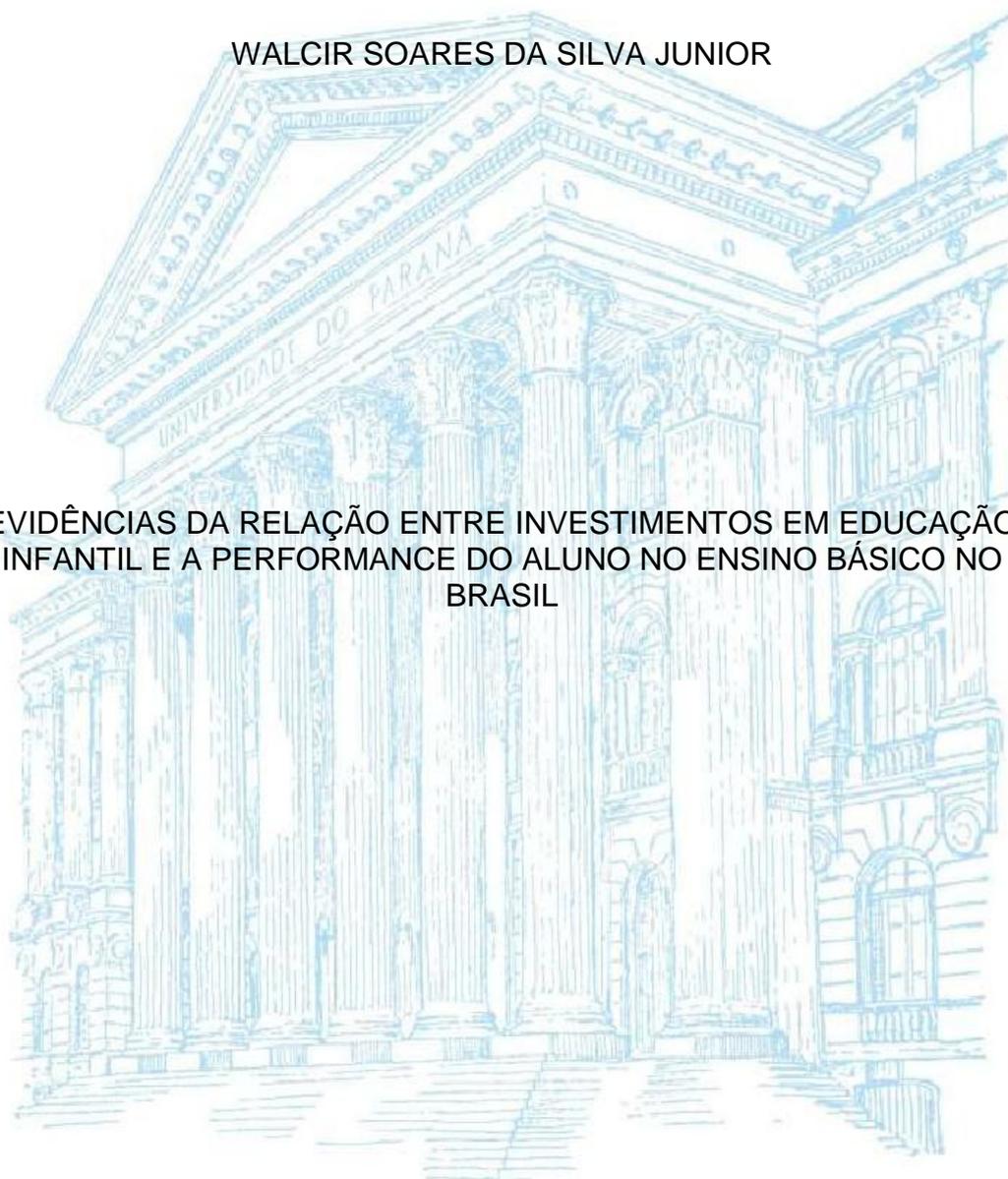


UNIVERSIDADE FEDERAL DO PARANÁ

WALCIR SOARES DA SILVA JUNIOR

EVIDÊNCIAS DA RELAÇÃO ENTRE INVESTIMENTOS EM EDUCAÇÃO
INFANTIL E A PERFORMANCE DO ALUNO NO ENSINO BÁSICO NO
BRASIL



CURITIBA
2013

WALCIR SOARES DA SILVA JUNIOR

EVIDÊNCIAS DA RELAÇÃO ENTRE INVESTIMENTOS EM EDUCAÇÃO
INFANTIL E A PERFORMANCE DO ALUNO NO ENSINO BÁSICO NO
BRASIL

Dissertação apresentada como requisito parcial à obtenção do grau de Mestre em Desenvolvimento Econômico, no Programa de Pós-Graduação em Desenvolvimento Econômico, Setor de Ciências Sociais Aplicadas, Universidade Federal do Paraná.

Orientador: Prof. Dr. Flávio de Oliveira Gonçalves.

CURITIBA
2013

TERMO DE APROVAÇÃO
(EXEMPLO!!! TROCAR PELA VERSÃO COM ASSINATURAS)

WALCIR SOARES DA SILVA JUNIOR

EVIDÊNCIAS DA RELAÇÃO ENTRE INVESTIMENTOS EM EDUCAÇÃO
INFANTIL E A PERFORMANCE DO ALUNO NO ENSINO BÁSICO NO
BRASIL

Dissertação aprovada como requisito parcial para obtenção de grau de Mestre no Programa de Pós Graduação em Desenvolvimento Econômico, Setor de Ciências Sociais Aplicadas, Universidade Federal do Paraná, pela seguinte banca examinadora:

Prof. Dr. Flávio de Oliveira Gonçalves
Orientador – Departamento de Economia, UFPR

Prof. Dr. XXXXXXXXXXXXXXXXXXXX
Departamento de XXXXXX, UFPR

Prof. Dr. XXXXXX
Departamento de XXXX, XXX

Prof. Dr. XXXXXXXX
Departamento de XXX, XXX

Curitiba, XX de Março de 2013.

Dedico este trabalho à minha família, Walcir, Jane e Tiago, e à minha noiva Anne Caroline, pelo apoio incondicional e amor transbordante.

AGRADECIMENTOS

Albert Einstein disse um dia que educação é aquilo que fica depois que você esquece o que a escola ensinou. E acho que a sede de desvendar segredos, a busca pelas respostas, ou mesmo as perguntas corretas podem representar um banho de lucidez a quem quer ter algo próximo à consciência de sua existência nesse mundo. Assim me coloco diante da educação, e de modo mais amplo, do conhecimento. Os que me conhecem sabem que a chegada a esse estágio, de defender uma dissertação de mestrado não foi uma tarefa fácil. Assim como não é fácil para milhares de pessoas como eu, que sem ter condições financeiras favoráveis, ultrapassam as barreiras do impossível para si mesmas, e descobrem que conseguem chegar do outro lado, quando realmente querem.

Assim, faço dessa deixa a oportunidade para dizer que não há nada que possamos conseguir estando sozinhos. Mesmo que o mérito seja merecido por se levantar cedo todos os dias tendo consciência de que o jogo está longe de estar ganho, existe o mérito daqueles que acreditando no impossível, dão forças para torná-lo possível. E aqui, desde o início dos meus atuais 25 anos, estão meus pais, responsáveis por tentar a seu modo responder aos anseios de uma criança curiosa, com sede de conhecimento, manancial de ideias e a sensibilidade artística do desassossego, inquietude e ousadia tímida. Agradeço aos meus queridos pais Walcir e Jane, por apoiarem esse naufrago por opção que geraram, na sua perseverança em alcançar a montanha mais alta. Ao meu irmão Tiago, que apesar de todas as nossas diferenças, sempre transbordou orgulho e admiração por mim, sempre me fez crer estar no caminho correto, e de quem não tenho a mínima dúvida do amor, tanto meu, quanto dele. Agradeço à minha noiva Anne Caroline, pelos incontáveis dias em que não pude estar ao seu lado pra poder corresponder às expectativas de um estudo metódico e eficiente, e por me fazer sentir ser melhor do que imaginava. A todos os meus familiares, que num apoio e incentivos verdadeiros e perenes, me fizeram acreditar no meu potencial, e compreenderam minha ausência. Obrigado a todos os meus professores, de todas as épocas, onde sempre tive suporte para voar, especialmente o meu orientador e especial amigo Flávio Gonçalves, quem compreende e incentiva além de meus anseios acadêmicos, minhas necessidades artísticas como ser humano. Ao Observatório da Educação e à Capes, pelo financiamento deste trabalho, aos professores do PPGDE/UFPR, ao NAPPE, e aos amigos e colegas da Faculdade Santa Cruz. E por fim, a todos os meus amigos que estiveram comigo nessa fase tão importante, especialmente Robson Luiz, Bogdan Skorupa, Emerson Gogola, Denis Cardoso, Priscila Saito, Irene Starepravo, Thiago Sieclick, Pamela Olmedo, Carlos Eduardo Frölich, Joaquim Pereira, Gabriele Vasconcelos, Elidecir Jacques, Ana Kern, Marco Túlio, Rogério Duenhas e Osvaldo Quintella. E por fim, uma frase de Leon Tolstói: “não alcançamos a liberdade buscando a liberdade, mas sim a verdade. A liberdade não é um fim, mas uma consequência”.

A mente verdadeiramente criativa em qualquer campo não é mais que isto: uma criatura humana nascida anormalmente, inumanamente sensível. Para ele, um toque é uma pancada, um som é um ruído, um infortúnio é uma tragédia, uma alegria é um êxtase, um amigo é um amante, um amante é um deus e o fracasso é a morte. Adicione-se a este organismo cruelmente delicado a subjugante necessidade de criar, criar, criar - de tal forma que sem a criação de música ou poesia ou literatura ou edifícios ou algo com significado, a sua respiração lhe é cortada. Ele tem que criar, deve derramar criação. Por qualquer estranha e desconhecida urgência interior, não está realmente vivo a menos que esteja criando.

Pearl Buck

RESUMO

Este trabalho tem por objetivo estimar o efeito do ensino infantil no desempenho dos alunos no ensino básico no Brasil, utilizando dados do Prova Brasil 2011. Supondo um problema de seleção por observáveis na escolha dos pais entre colocar ou não seus filhos no ensino infantil, será utilizada uma metodologia de pareamento, a saber, *Propensity Score Matching*, que tem por objetivo encontrar um grupo ideal de comparação. Para estimar o efeito do ensino infantil na nota do 5º e 9º anos controlando para características que estão ligadas ao desempenho, ainda serão estimadas regressões por Mínimos Quadrados Ponderados pelo *Propensity Score*. Foram encontrados efeitos positivos e significativos nas notas de 5º e 9º anos para os alunos que iniciaram seus estudos no ensino infantil, creche e pré-escola, com ambas as metodologias, atualizando e corroborando as evidências disponíveis de que o ensino infantil pode melhorar a qualidade da educação, quando desenvolve habilidades que impactam no desempenho dos alunos mais tarde. Por fim, um cálculo de retorno econômico mostra um resultado positivo para investimentos no ensino infantil, o que pode ser considerado um resultado subestimado, já que os retornos sociais da educação também podem ser muito altos.

Palavras-chave: ensino infantil, habilidades cognitivas e não-cognitivas, desempenho escolar, *propensity score matching*, mínimos quadrados ponderados pelo *propensity score*, retorno econômico.

ABSTRACT

This study aims to estimate the effect of early childhood education on student performance in basic education in Brazil, using data from Brazil Exam 2011. Assuming a selection problem for the observable parental choice between placing or not their children in kindergarten, a methodology will be used for pairing, namely propensity score matching, which aims to find an ideal comparison group. To estimate the effect of early childhood education in the 5th grade and 9th grade controlling for characteristics that are linked to performance, yet be estimated by least squares regressions weighted by propensity score. We found significant positive effects on grades of 5 and 9 years for students who began their studies in early childhood education, child care and preschool, with both methodologies, updating and confirming the available evidence that early childhood education can improve the quality education, while developing skills that impact on student performance later. Finally, a calculation of economic return shows a positive result for investments in early childhood education, which can be considered a result underestimated, since the social returns to education can also be very high.

Key-words: early childhood education, cognitive and non-cognitive skills, school performance, propensity score matching, weighted least squares by propensity score, economic return.

LISTA DE GRÁFICOS E FIGURAS

GRÁFICO 1. TAXAS DE RETORNO PARA OS INVESTIMENTOS EM CAPITAL HUMANO EM CRIANÇAS EM DESVANTAGEM.....	24
GRÁFICO 2: HISTÓRICO DA ESTIMATIVA DO PERCENTUAL DO INVESTIMENTO PÚBLICO TOTAL EM EDUCAÇÃO EM RELAÇÃO AO PIB, POR NÍVEL DE ENSINO – BRASIL 2000-2010.....	25
GRÁFICO 3: HISTÓRICO DA ESTIMATIVA DO INVESTIMENTO PÚBLICO DIRETO EM EDUCAÇÃO POR ESTUDANTE, COM VALORES ATUALIZADOS PARA 2010 PELO IPCA, POR NÍVEL DE ENSINO.....	26
GRÁFICO 4. PROVA BRASIL 2011. DISTRIBUIÇÃO DE FREQUÊNCIAS DAS NOTAS EM PORTUGUÊS E MATEMÁTICA. 5º ANO.....	39
GRÁFICO 5. PROVA BRASIL 2011. DISTRIBUIÇÃO DE FREQUÊNCIAS DAS NOTAS EM PORTUGUÊS E MATEMÁTICA. 9º ANO.....	44

LISTA DE TABELAS

TABELA 1. DESCRIÇÃO DAS VARIÁVEIS UTILIZADAS PARA ALUNOS DO 5º ANO.	39
TABELA 2. DESCRIÇÃO DAS VARIÁVEIS UTILIZADAS PARA ALUNOS DO 9º ANO.	44
TABELA 3. RESULTADOS ESTIMADOS DO MODELO PROBIT – 5º ANO. ...	47
TABELA 4. RESULTADOS ESTIMADOS DO MODELO PROBIT – 9º ANO. ...	49
TABELA 5. ESTIMATIVA DO EFEITO TRATAMENTO NOS TRATADOS (ATT) PELO MÉTODO STRATIFICATION MATCHING, E PELO MÉTODO DOS MÍNIMOS QUADRADOS PONDERADOS PELO PROPENSITY SCORE.....	50
TABELA 6. REGRESSÃO MÍNIMOS QUADRADOS PONDERADOS PELO PROPENSITY SCORE. ESTIMATIVA 1: CRECHE - 5º ANO – PORTUGUÊS.	54

SUMÁRIO

1. INTRODUÇÃO.....	12
2. A RELEVÂNCIA DO ENSINO INFANTIL E O INVESTIMENTO PÚBLICO NO BRASIL.....	18
2.1 O ENSINO INFANTIL COMO CRUCIAL NO DESENVOLVIMENTO DE HABILIDADES	21
2.2 OS NÍVEIS DE INVESTIMENTO EM EDUCAÇÃO, E A COBERTURA DO ENSINO INFANTIL NO BRASIL	24
2.3 MODELO TEÓRICO: VARIÁVEIS QUE CONDICIONAM A FORMAÇÃO DE HABILIDADES NO CICLO DE VIDA.....	27
3. ESTRATÉGIA EMPÍRICA	29
3.1 APRESENTAÇÃO DA METODOLOGIA UTILIZADA NO TRABALHO ..	30
3.2 O ESTIMADOR PROPENSITY SCORE MATCHING (PSM)	31
3.3 O ESTIMADOR DOS MÍNIMOS QUADRADOS PONDERADOS PELO PROPENSITY SCORE	34
4. DESCRITIVAS DO PROVA BRASIL 2011	36
4.1 BASE DE DADOS E ESTATÍSTICAS DESCRITIVAS – PROVA BRASIL 2011	36
4.1.1 DESCRITIVAS PROVA BRASIL 2011 – 5º ANO.....	37
4.1.2 DESCRITIVAS PROVA BRASIL 2011 – 9º ANO	42
5. ANÁLISE DOS RESULTADOS.....	47
5.1 ANÁLISE DOS RESULTADOS OBTIDOS PELO <i>PROPENSITY SCORE</i> 5º ANO	47
5.2 ANÁLISE DOS RESULTADOS OBTIDOS PELO <i>PROPENSITY SCORE MATCHING</i> – 9º ANO	48
5.3 ESTIMAÇÃO DO IMPACTO DE TER FREQUENTADO A EDUCAÇÃO INFANTIL NOS RESULTADOS DO ENSINO BÁSICO.....	50
5.4 ANÁLISE DOS RESULTADOS OBTIDOS PELO MÉTODO DE MÍNIMOS QUADRADOS PONDERADOS PELO <i>PROPENSITY SCORE</i> – 5º ANO	54

6	RETORNO ECONÔMICO	57
7	CONCLUSÃO.....	60
	REFERÊNCIAS.....	63
	APÊNDICES	68

1. INTRODUÇÃO

A preocupação com a primeira infância nos estudos sobre economia da educação tem sido um dos focos mais importantes desta literatura. Como ensino da primeira infância, consideram-se aqui todas as fases que precedem o ensino fundamental, importantes para a formação, mas ainda hoje colocados num patamar de menor importância no ciclo de ensino com relação à cobertura. São essas fases a creche (0 a 3 anos) e pré-escola (4 a 5 anos). Diversos autores apresentam evidências de que o ciclo de ensino se divide em fases, e que a primeira infância, por ser um período onde diversas habilidades são exclusivamente desenvolvidas, é a fase mais importante entre as fases de ensino (Cunha, et. al. 2006, Heckman, 2000 e Carneiro e Heckman, 2003). A despeito dessas evidências, no Brasil os investimentos na área da educação recaem com peso muito maior na fase do ensino fundamental, médio e principalmente superior, onde muitas das habilidades já não possuem tanta maleabilidade quanto na primeira infância. Assim, investindo-se na primeira infância, o desenvolvimento dessas habilidades seria o canal pelo qual a qualidade da educação, medida tradicionalmente pelo desempenho em testes padronizados, aumentaria.

Duas das principais características do capital humano, é exibir auto-produtividade e complementaridade. O primeiro aspecto diz respeito a que a realização de habilidades no primeiro estágio aumenta a realização de habilidades no futuro. O segundo aspecto, complementaridade, diz respeito a que investimentos iniciais em educação facilitam a produtividade de investimentos mais tarde, mas esses investimentos não são produtivos sem aqueles. Para Becker e Tomes (1986) não há trade-off entre equidade e eficiência nos investimentos de capital humano na fase inicial, já que as decisões com respeito à entrada da criança no ensino infantil são tomadas pelos pais, e portanto, uma criança receber investimentos em educação infantil, não significa um desestímulo para as outras. Para investimentos posteriores o trade-off existe, já que dar preferência para um determinado grupo pode criar um desestímulo, já que este precisará se esforçar menos do que os outros. Em todos os ciclos do ensino, habilidades geram mais habilidades. No entanto,

determinadas habilidades possuem um período sensível, em que seu desenvolvimento é muito mais eficiente. Se a duração desse período é limitada, passa a ser chamado de período crítico, a partir do qual a maleabilidade de aprendizado é muito pequena. (Cunha et. al., 2006)

Heckman (2000) e Carneiro e Heckman (2003) reconhecem que tanto as habilidades cognitivas – habilidades voltadas à inteligência como o Q.I. - quanto as não-cognitivas – habilidades como paciência, perseverança, esforço, etc. - possuem um componente genético e outro componente influenciado pelo ambiente. É através do último que as políticas públicas podem agir. Os autores mostram também que as restrições orçamentárias realmente relevantes que as crianças em desvantagem enfrentam estão ligadas aos recursos que seus pais podem lhes proporcionar na fase inicial de ensino. As restrições financeiras posteriores do período de adolescência explicam muito menos do não comparecimento à faculdade. Assim, através do aumento no desenvolvimento das habilidades no primeiro período, é que o comparecimento dessas crianças à escola, e depois à faculdade, aumenta. Cunha et. al. (2006) mostra que mesmo quando as intervenções não aumentam o QI, elas podem aumentar as habilidades não cognitivas, gerando grandes efeitos sobre a escolaridade, o mercado de trabalho e aspectos comportamentais como gravidez na adolescência e marginalidade.

Britto (2012) utilizando um painel de escolas extraído dos microdados da Prova Brasil 2009 e controlando para características específicas dos indivíduos e do *background* familiar encontra que as escolas possuem um desempenho melhor se seus alunos ingressaram no maternal ou na pré-escola, em comparação com escolas em que os alunos entraram diretamente na primeira série. Felício e Vasconcellos (2007) utilizando dados do SAEB 2003 estimaram com base nos dados individuais dos alunos, aplicando métodos de pareamento (*Propensity Score Matching (PSM)*), que frequentar a escola antes da primeira série tem um impacto positivo e significativo sobre a proficiência das crianças em testes padronizados de matemática, no 5º ano.

Assim como Felício e Vasconcellos (2007) este trabalho utilizará um estimador de *Propensity Score Matching (PSM)*, que tem por objetivo controlar a existência de seleção por observáveis que parece existir na escolha entre frequentar a educação infantil ou não, utilizando dados do Prova Brasil 2011.

Uma das contribuições deste trabalho será atualizar a evidência de 2003 para dados de 2011, de que o ensino infantil impacta positivamente na performance dos alunos no ensino fundamental de 5º e 9º anos, em provas de português e matemática. Além disso, este trabalho estimará o efeito da educação infantil no desempenho do aluno, controlando para características do aluno que impactam na sua performance, através de um método duplamente robusto, a saber, Mínimos Quadrados Ponderados pelo *Propensity Score*. Após a estimação dos efeitos, será feita também uma análise do retorno econômico dos investimentos na primeira infância, justificando ainda mais a importância das políticas públicas focarem nessa fase de ensino.

Schweinhart et al (2005) apresentam estudo empírico em que o programa de educação infantil teve efeitos positivos e significativos sobre a conclusão do ensino médio, a renda e o emprego na vida adulta, entre outros, com retorno econômico de US\$17,07 por dólar investido, a preços de 2000. Campbell et al (2002) apresentam os resultados de um experimento avaliando um grupo com educação infantil tradicional e outro com um programa especial. O grupo do programa especial resultou melhor desempenho no ensino fundamental, menor índice de repetência e gravidez na adolescência, além de ter tido maior frequência na conclusão do ensino médio e do ensino superior, rendendo US\$3,78 por dólar investido, a preços de 2000. Reynolds et al (2001) também apresentam evidências através de um experimento de que o grupo que frequentou ensino infantil apresentou melhor desempenho educacional e comportamento social, menor repetência, menor índice de marginalidade, além de maior taxa de conclusão escolar, com retorno para a sociedade de US\$7,10 por dólar investido, a preços de 2000.

A despeito das evidências a favor da primeira infância, no Brasil, do ano de 2000 até 2010, segundo dados do INEP, os investimentos totais em educação infantil ficaram estagnados em 0,4% do Produto Interno Bruto anuais. Sendo que os investimentos totais em educação tiveram um aumento de 23% durante estes dez anos, passando de 4,7% para 5,8% do PIB. Além da estagnação, observa-se também que a educação infantil é o nível de menor cobertura no Brasil.

A educação infantil representa também o nível educacional de menor proporção percentual do Investimento Público Direto, apresentando inclusive

uma redução nessa década em questão de 10%. A educação infantil também continua sendo o menor nível no investimento direto em educação por estudante¹, apresentando de 2000 a 2010 um aumento de aproximadamente 70%, sendo que houve aproximadamente 120% de aumento nos investimentos agregados em educação. Assim, a fase em que a literatura evidencia como uma das mais importantes no ciclo de ensino, tem tido pouco investimento em relação ao ensino fundamental, médio e superior.

Desse modo, justifica-se a importância de se encontrar evidências para o Brasil que corroborem com a literatura à favor do ensino infantil, apoiando e justificando a existência de políticas públicas que contribuam com a melhora nos investimentos em educação infantil. O melhor cenário para pesquisa em economia da educação seria os experimentos naturais ou quase-experimentos, pois garantem a independência das variáveis explicativas, levando em consideração problemas como endogeneidade e seleção, tradicionais para esses tipos de dados. No entanto, esses métodos além de possuírem um alto custo, enfrentam sérios problemas éticos, quando da escolha dos grupos de tratamento e de controle. (Behrman, 1996).

No entanto, utilizando-se um estimador *Propensity Score Matching* (PMS), mesmo sem a existência de um contrafactual, ou seja, um grupo que tenha as mesmas características dos que foram tratados, mas que não tenham recebido o tratamento, pode-se obter dados de um conjunto de unidades de comparação potenciais. Esse grupo de comparação é formado por aqueles indivíduos cujas características observáveis x quando pareadas são iguais em um determinado grau àquelas do grupo tratado. Assim, a média dos resultados dos não tratados pareados com os grupos identifica a média contrafactual resultante no grupo tratado na ausência de tratamento. (Becker e Ichino, 2002).

Portanto, para cumprir o propósito de medir o efeito de ter frequentado a creche ou pré-escola (educação infantil) sobre as habilidades cognitivas no ensino básico (nota no Prova Brasil no 5º e 9º anos), foi estimado um modelo de pareamento usando *propensity score* (PSM), utilizando a base de dados do Prova Brasil 2011, uma base recente e ainda pouco explorada, atualizando as

¹ É preciso salientar que aqui o investimento por estudante utilizado é a média do país. Existe um problema muito mais complexo quanto à desigualdade na distribuição dos recursos, o que faz com que em municípios mais ricos, o investimento seja maior.

evidências da literatura. Para controlar também as características dos alunos que impactam no seu desempenho, serão estimadas regressões utilizando o método dos Mínimos Quadrados Ponderados pelo Propensity Score, um método duplamente robusto, que controla para as características determinantes do desempenho, e através do *Propensity score*, para as características determinantes da seleção. Ambas as metodologias encontraram efeitos positivos e altamente significativos do ensino infantil na nota do aluno do 5º e 9º anos: para o 5º ano, frequentar a educação infantil aumenta em 8,2 pontos a nota do aluno em português e 8,5 em matemática. Para o 9º ano, o aluno que inicia seus estudos no ensino infantil possui uma nota maior em 9 pontos em português e 8,2 pontos em matemática. O efeito da pré-escola é ainda maior, acrescentando em 15,4 e 16,5 pontos em português e matemática, respectivamente, para o 5º ano, e 14,4 e 13,4 pontos em português e matemática, respectivamente, para o 9º ano. Uma média de 11,7 pontos significa que um aluno que esteja na mediana, para um desvio-padrão médio de 46,5, apenas por ter feito o ensino infantil, ultrapassa 10% dos alunos na distribuição. Além das habilidades cognitivas, mais voltadas à inteligência como o QI, evidências mostram que o ensino pré-escolar aumenta também as habilidades não cognitivas, mais difíceis de serem medidas, como esforço, atenção e disciplina (Graces et al., 2002 e Gertler e Fernald, 2004). Corroborando com as evidências encontradas, a análise de retorno econômico apresentou um retorno de R\$1,32 reais (a preços de 2010) por real investido na primeira infância, retorno esse que é subestimado, já que os retornos sociais são difíceis de mensurar, tendo inclusive, evidências de que ultrapassam os retornos privados. (GROSSMAN, 2006, in BARBOSA FILHO e PESSOA, 2009).

Este trabalho está dividido em sete seções, a contar com esta introdução. A segunda apresenta as principais evidências da educação infantil na literatura, números dos investimentos em educação no Brasil, e um modelo teórico expando as relações entre as principais variáveis envolvidas na escolha de um aluno começar ou não seus estudos na educação infantil. A terceira seção apresenta a estratégia empírica adotada, e as metodologias utilizadas para estimar os efeitos, a saber, o *Propensity Score Matching*, e o Mínimos Quadrados Ponderados pelo *Propensity Score*. A quarta seção apresenta

estatísticas descritivas da base de dados do Prova Brasil 2011 para o 5º e 9º anos, e uma breve exposição das variáveis utilizadas nos exercícios econométricos. A quinta seção apresenta a análise dos principais resultados para ambos os modelos estimados, a comparação entre os resultados das diferentes metodologias utilizadas, e também dos resultados encontrados com as evidências da literatura. A sexta seção apresenta uma breve discussão dos retornos econômicos de programas de educação infantil calculados na literatura, além de um cálculo de retorno econômico dos investimentos em educação infantil no Brasil. A sétima seção apresenta as conclusões deste trabalho.

2. A RELEVÂNCIA DO ENSINO INFANTIL E O INVESTIMENTO PÚBLICO NO BRASIL

A educação infantil no Brasil é dividida em creche, dos 0 aos 3 anos, e pré-escola, de 4 a 5 anos². A partir da década de 1990, o acesso a esse nível de ensino teve sua importância aumentada, devido à redemocratização após a Constituição de 1988, que além de garantir um aumento da quantidade educacional às crianças de zero a seis anos, passou a tratar a educação infantil também como um direito da criança. Esse aumento representou um grande avanço, dando acesso educacional às crianças, e permitindo também à mulher, para que tivesse uma profissão e participasse na formação da renda familiar. No entanto, a infraestrutura e o capital humano não tiveram um aumento proporcional, levando a uma queda da qualidade educacional das escolas públicas. Em meados de 1990, junto com a preocupação com a qualidade educacional como um todo, a qualidade da educação infantil também passou a ganhar importância, principalmente após a promulgação da Lei de Diretrizes e Bases da Educação Nacional (LDB), em 1996, que incluiu o atendimento de crianças de 0 a 3 anos nas creches. (Guimarães e Pinto, 2001). Campos (1997) explica que a importância da educação infantil mudou ao longo do tempo para diversos níveis, mas o aumento da literatura sobre educação infantil nos últimos anos, e por conseguinte, seus resultados, mostram que os retornos nessa fase da infância são muito altos.

O processo educacional já não é mais visto como sendo composto por uma só fase, e sim de diversas fases em que o ser humano forma suas habilidades³, cada fase com uma importância diferente da outra em diversos aspectos. Nesse ínterim, a primeira infância possui um papel fundamental, pois nesta fase, se algumas habilidades não forem desenvolvidas, tentar desenvolvê-las através de investimentos mais tarde, no ensino básico e superior, pode ser muito custoso, e muitas vezes pouco eficiente. Dentro desse processo há o reconhecimento das habilidades tanto cognitivas como não-

² Recentemente, o ensino fundamental foi aumentado em um ano, considerando-se então o que era o ensino pré-escolar, como primeiro ano. Assim, a pré-escola passou de 5 a 6 para de 4 a 5 anos de idade.

³ Aqui capacidades e habilidades são usadas como sinônimos, e servem para especificar os conhecimentos adquiridos pelos alunos que fazem com o que o seu desempenho seja melhor ou pior com o passar dos anos.

cognitivas⁴ responsáveis pelo sucesso futuro da criança, habilidades essas que podem ser produzidas tanto pela família – importância especial para o papel da mãe - como pelas ações pessoais do indivíduo. Além disso, a formação dessas habilidades, tanto cognitivas, quanto não cognitivas, possui um componente genético, e um componente advindo do ambiente, que interagem entre si. Assim, políticas públicas podem agir sobre o componente advindo do ambiente, já na fase da primeira infância, aumentando assim a produtividade dos investimentos nos períodos posteriores. (Heckman, 2000, Carneiro e Heckman, 2003 e Cunha et. al. 2006).

Knudsen (2004) separa o processo de aprendizado em dois períodos, o período crítico e o período sensível. Segundo o autor, o cérebro é amplamente influenciado pelas experiências do indivíduo. Quando o efeito de determinada experiência é muito forte em um período limitado, esse é chamado de período sensível. Ou seja, período sensível é aquele em que o cérebro consegue aprender mais facilmente determinadas habilidades. Durante esse período, o cérebro cria conexões estáveis, que podem ser alteradas depois do período sensível, mas a maleabilidade dessas mudanças depende da estrutura criada durante o período sensível. Quando determinada experiência é crucial em um período sensível, para o desenvolvimento normal do indivíduo, e altera o seu desempenho permanentemente, esse período é chamado de período crítico. Knudsen (2004) ainda aponta que para diversas habilidades como o Q.I., o período crítico acaba em torno dos dez anos de idade, o que mostra a extrema importância de seu desenvolvimento antes dessa idade máxima.

Contribuindo com essa visão, Newport (2002) apresenta a evidência de que a idade para exposição da criança à uma nova linguagem é negativamente relacionada à proficiência adquirida em sua língua natal. O autor mostra que o aprendizado de aspectos mais formais da língua fica mais difícil na idade adulta, enquanto a aquisição de vocabulário, por exemplo, pode ser feita mais facilmente. Essa é uma das evidências a favor da hipótese do período crítico e período sensível.

⁴ Habilidades cognitivas basicamente são aquelas ligadas à inteligência e à organização, como por exemplo, memória, capacidade de abstração, etc. Já as não cognitivas são aquelas ligadas às competências socio-emocionais, como motivação, autonomia, equilíbrio emocional, etc.

Para Cunha et. al. (2006), tanto as capacidades cognitivas quanto as não-cognitivas podem sofrer a influência do ambiente, já que em ambas habilidades há um componente genético, e um componente que pode ser adquirido, embora essa flexibilidade seja diferente para as diferentes idades. Segundo Heckman, Stixrud e Urzua (2006), tanto as habilidades cognitivas quanto as não cognitivas afetam o desempenho e performance escolares.

Segundo Cunha et al. (2006) habilidades geram ainda mais habilidades através da complementaridade e auto-produtividade, que são dois conceitos que caracterizam os investimentos em capital humano. O primeiro aspecto diz respeito a que a realização de habilidades no primeiro estágio aumenta a realização de habilidades no futuro. Já a complementaridade diz respeito a que investimentos iniciais facilitam a produtividade de investimentos mais tarde, mas não são produtivos sem eles. Para Becker e Tomes (1986) não há trade-off entre equidade e eficiência nos investimentos de capital humano na fase inicial, já que a decisão de colocar a criança no ensino infantil é tomada pelos pais, e portanto, uma criança receber investimentos em educação infantil, não significa um desestímulo para as outras. Para investimentos posteriores o trade-off existe, já que dar preferência para um determinado grupo pode criar um desestímulo para os outros grupos.

Assim, evidencia-se o argumento de que um tratamento diferente para os anos iniciais é muito importante, já que para que os investimentos mais tarde sejam produtivos, os investimentos iniciais devem ser prioritários. O foco deste trabalho é a atenção voltada aos primeiros anos da infância – ou a falta dela –, para que os retornos mais tarde sejam os maiores possíveis, e evidenciar o resultado de que o fato de frequentar o ensino infantil (pré-escola ou creche) pode influenciar nas notas dos alunos no ensino básico. A metodologia do estimador *Propensity Score Matching (PSM,)* se justifica devido à possível existência de seleção por observáveis, no momento em que a seleção se a criança irá ou não frequentar o ensino infantil. Para se controlar para as variáveis que podem influenciar no desempenho do aluno, será feita também uma estimativa utilizando Mínimos Quadrados ponderados pelo *Propensity Score*. É preciso salientar, no entanto, que um dos limites deste trabalho é que além deste tipo de seleção, pode existir também seleção por não-observáveis, como por exemplo, a proximidade ou inexistência de creches

ou pré-escolas perto do domicílio de algumas famílias, o que pode subestimar a estimativa do efeito, já que estas crianças sem creche são oriundas de zonas onde a oferta de educação é muito baixa ou inexistente, o que faz com que o efeito estimado seja menor do que o real.

2.1 O ENSINO INFANTIL COMO CRUCIAL NO DESENVOLVIMENTO DE HABILIDADES

Carneiro e Heckman (2002, 2003) mostram que as restrições orçamentárias realmente relevantes que as crianças em desvantagem enfrentam estão ligadas ao seu ambiente inicial, e os recursos que seus pais poderiam lhe proporcionar, e não à incapacidade da família de realizar empréstimos que financiem a educação de seus filhos em idade escolar ou na adolescência. Políticas de enriquecimento do ambiente inicial da criança podem aumentar suas habilidades cognitivas e não-cognitivas, e esse fator sim, é o maior responsável pelo comparecimento dessas crianças à escola. Em outras palavras, investir na primeira infância é mais eficiente que investir em políticas de acesso ao ensino superior, e é o desenvolvimento na primeira fase do ensino, seja por habilidades ou por enriquecimento cultural, que mais influenciam no comparecimento das crianças na escola e no ensino superior.

Mesmo quando as intervenções no início da infância não aumentam o Q.I., elas aumentam as habilidades não cognitivas, que também têm grandes efeitos na escolaridade, no mercado de trabalho, e em outros aspectos comportamentais como gravidez na adolescência. E assim, através dos efeitos na escolaridade, essas intervenções podem aumentar a pontuação em testes mesmo que não aumentando o Q.I., o que demonstra a importância crucial das habilidades não-cognitivas. (Cunha et. al, 2006, Gertler e Fernald, 2004 e Graces et. al, 2002).

Para Carneiro e Heckman (2002, 2003) a falta de recursos a que estão sujeitas as crianças, são a mais importante variável na explicação da diferença que existe entre o número de alunos que concluem o ensino médio e o número de matrículas no ensino superior. Os autores controlaram para as habilidades já formadas no início da adolescência, e descobriram que a renda da família na

adolescência tem um papel muito pequeno na explicação dessa diferença nas matrículas do ensino superior. Em outras palavras, a causa desta diferença é menos a restrição orçamentária familiar, e mais a falta de habilidades cognitivas e não-cognitivas que são adquiridas no início do ciclo de vida das crianças, ou seja, os fatores familiares de longo-prazo, que se traduzem nessas habilidades adquiridas no início do ciclo. Cunha et al. (2006) ainda apresentam um exercício em que mostram que a diferença racial e de renda existente em testes de desempenho é significativamente reduzida quando se controla para algumas características como educação materna, habilidades cognitivas e estrutura familiar. Assim, eles concluem que programas de enriquecimento de ambiente inicial para crianças em desvantagem podem parcialmente remediar a diferença entre estas e as crianças mais favorecidas.

Outros estudos têm obtido resultados semelhantes a favor da primeira infância. Um deles, Berlinski et al. (2006), em um estudo de experimento natural sobre um programa de construções de escolas na Argentina, concluem que os testes de aptidão (habilidades cognitivas) apresentam uma melhora de 8% para cada ano de pré-escola, além de melhoras nas variáveis atenção, esforço, participação e disciplina (habilidades não-cognitivas). Feinstein et al. (1998) apresentam o resultado, válido também após controlar educação dos pais, classe social e interesse dos pais e vizinhança, de que crianças dos 7 aos 11 anos demonstram maior desempenho na escola se frequentaram uma pré-escola ou ficaram sob os cuidados da mãe. Esse resultado evidencia a importância da pré-escola, principalmente para aquelas crianças que não possuem acompanhamento materno contínuo. Curi e Menezes-Filho (2006) encontram efeitos positivos e significativos de ter frequentado a creche sobre a probabilidade de conclusão das etapas de ensino e salários futuros.

Britto (2012) utilizando um painel de escolas extraído dos microdados da Prova Brasil 2009 e controlando para características específicas dos indivíduos e do *background* familiar encontra que as escolas possuem um desempenho melhor se seus alunos ingressaram no ensino infantil, em comparação com escolas em que os alunos entraram diretamente no ensino fundamental. Felício e Vasconcellos (2007) utilizando dados do SAEB 2003 estimaram com base nos dados individuais dos alunos, aplicando o método de *Propensity Score Matching (PSM)*, que frequentar a escola antes da primeira série tem um

impacto positivo e significativo sobre a proficiência das crianças em testes padronizados de matemática, no 5º ano.

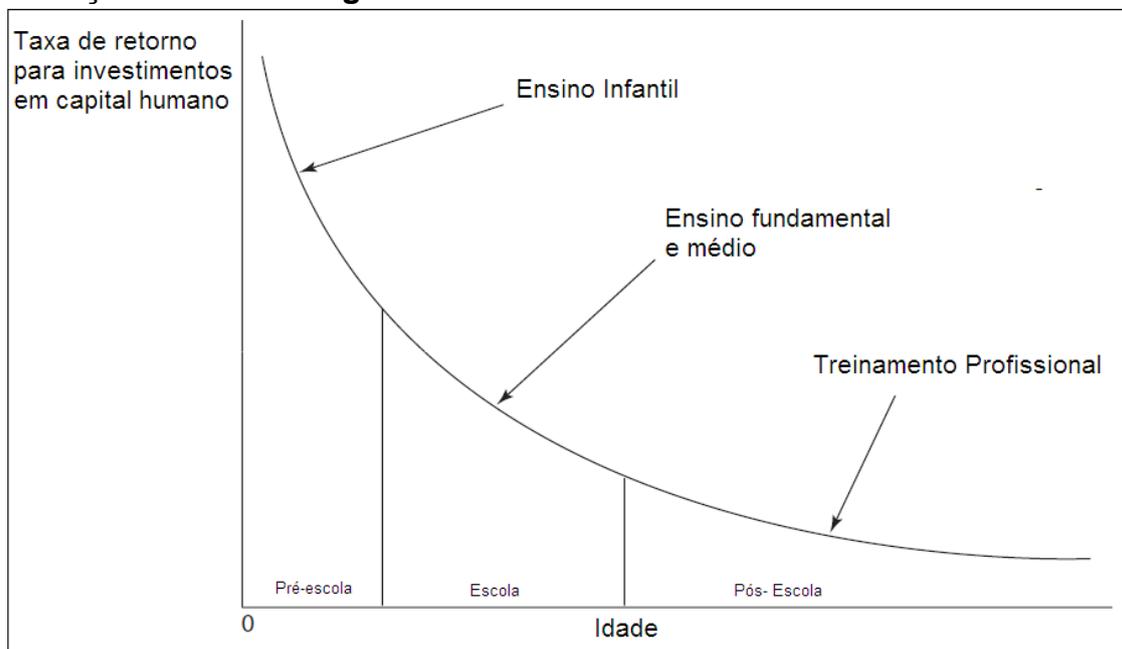
Este trabalho também utilizará um estimador de *Propensity Score Matching (PSM)*, tendo como objetivo atualizar as evidências com dados de 2003 de Felício e Vasconcellos (2007) para dados do Prova Brasil 2011, controlando a seleção por observáveis que parece existir na escolha entre frequentar a educação infantil ou entrar diretamente no ensino fundamental. Além disso, este trabalho estimará o efeito da educação infantil no desempenho do aluno, controlando para características do aluno que impactam na sua performance, através de um método duplamente robusto, a saber, o Mínimos Quadrados Ponderados pelo *Propensity Score*. Adicionalmente, será feita também uma análise do retorno econômico dos investimentos na primeira infância no Brasil, justificando ainda mais a importância das políticas públicas focarem nessa fase de ensino.

Perotti (1996), em estudo sobre os canais pelos quais a desigualdade pode afetar o crescimento econômico, mostra que a desigualdade é associada com baixos níveis de formação de capital humano, o que por sua vez é associado com baixos níveis de crescimento econômico. Galor e Zeira (1988,1993) mostraram que na presença de imperfeições no mercado de crédito, a distribuição de renda também possui um efeito sobre a formação de capital humano, ou seja, a desigualdade pode ser prejudicial à formação de capital humano e o desenvolvimento econômico.

Seguindo o argumento de Heckman e Masterov (2007), é preciso atentar para o fato de que essa discussão em favor das políticas de melhora nos ambientes iniciais das crianças em desvantagem não quer dizer que todas as habilidades e motivações são formadas nos primeiros anos. Muito menos quer dizer que as escolas e empresas não importam na formação das pessoas, ou que os anos iniciais são os únicos determinantes do sucesso. Dessa forma, dizer o argumento principal é defender o papel que o ambiente inicial tem nos resultados posteriores e que sua importância é negligenciada nas políticas atuais. Inclusive, as evidências da tecnologia de formação das habilidades humanas dizem que ambientes inicialmente enriquecidos precisam ser seguidos por boa escolaridade e aprendizagem no local de trabalho, ou seja, a complementaridade está intrínseca no processo de formação das habilidades.

A figura 1 apresenta as taxa de retorno para os investimentos em capital humano para os diversos níveis educacionais.

Gráfico 1. Taxas de retorno para os investimentos em capital humano em crianças em desvantagem

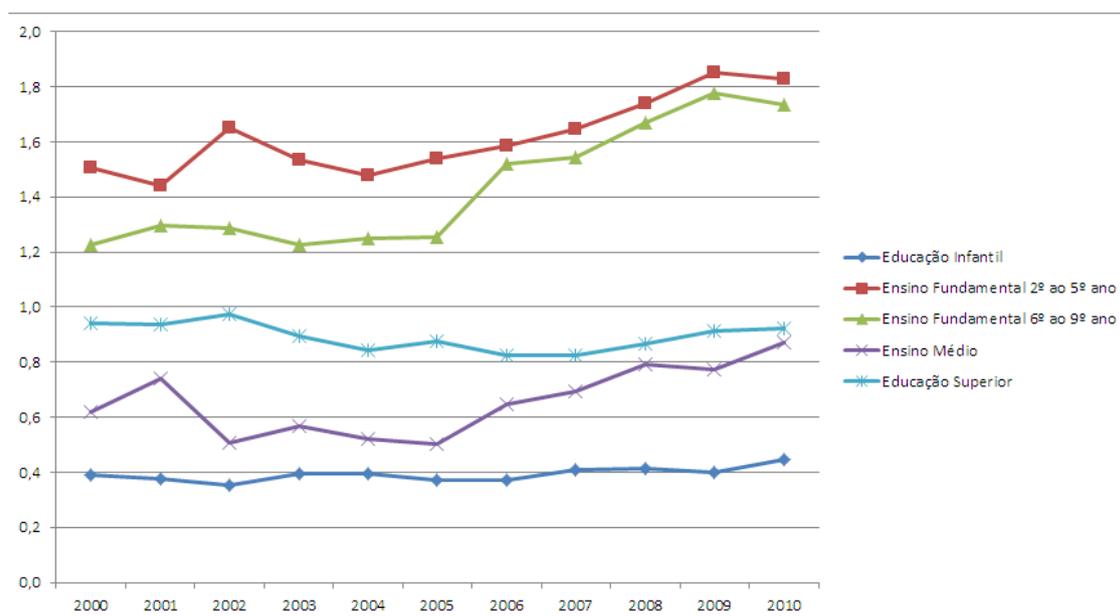


Fonte: Heckman e Masterov (2007)/ Elaboração Própria.

2.2 OS NÍVEIS DE INVESTIMENTO EM EDUCAÇÃO, E A COBERTURA DO ENSINO INFANTIL NO BRASIL

A seguir apresenta-se 2 gráficos com estatísticas a fim de analisar como se deram os investimentos em educação infantil no Brasil nos últimos anos. Apesar das evidências a favor da primeira infância, o gráfico 1 mostra que do ano de 2000 até 2010, os investimentos totais em educação infantil ficaram estagnados em 0,4% do Produto Interno Bruto anuais. Os investimentos totais em educação tiveram um aumento de 23% durante estes dez anos, passando de 4,7% para 5,8% do PIB. Além da estagnação, observa-se também que a educação infantil é o nível de ensino de menor gasto público total.

Gráfico 2: Histórico da estimativa do percentual do Investimento Público Total em educação em relação ao PIB, por nível de ensino – Brasil 2000-2010.



Fonte: elaboração própria. Dados INEP/MEC.

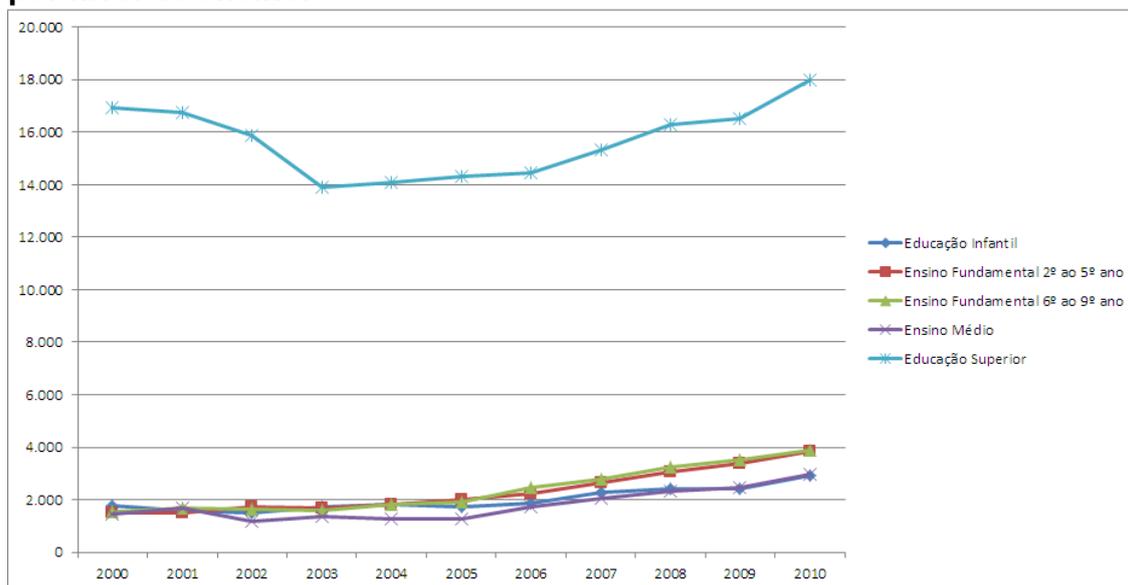
Analisando agora o investimento direto em educação por estudante, o gráfico 2 mostra dados com valores atualizados pelo IPCA para 2010: apesar de apresentar um aumento no período de aproximadamente 70%, esse aumento está muito aquém do aumento de aproximadamente 120% nos investimentos agregados em educação. No entanto, a educação infantil possui um nível investimento por aluno muito parecido com o do ensino médio, evidenciando que o problema da educação infantil no Brasil é mais uma questão de cobertura do que investimento por aluno.

Dados do IBGE de 2010 mostram que para crianças de 0 a 3 anos, pouco mais de 23% frequentavam a creche, e 80% das crianças de 4 e 5 anos frequentavam a pré-escola, em 2010. A cobertura para o ensino básico no Brasil é muito maior, 97% das crianças entre 6 e 14 anos frequentavam a escola em 2010. Entre os jovens de 15 e 17 anos, um percentual de 64% para 1995 cresce para mais de 83% em 2010.

No entanto, mesmo para o ensino básico, ainda que a quantidade educacional seja crescente, Veloso (2011) ressalta dois pontos: o grau de repetência e evasão ainda é muito alto, e embora a universalização do ensino

esteja quase atingida, uma parcela representativa das crianças não conclui o ensino fundamental. Segundo o IBGE (2010), apenas 47% dos jovens entre 15 e 17 anos estavam matriculados adequadamente no ensino médio em 2010, sem defasagem idade-série.

Gráfico 3: Histórico da estimativa do Investimento Público Direto em educação por estudante, com valores atualizados para 2010 pelo IPCA, por nível de ensino.



Fonte: elaboração própria. Dados INEP/MEC.

Apesar do baixo crescimento nos investimentos agregados em educação infantil em comparação com as outras fases de ensino, algumas políticas demonstram certa preocupação com o ensino infantil, mesmo que ainda muito aquém do esperado. O tempo mínimo para conclusão do ensino fundamental foi recentemente aumentado de oito para nove anos, reduzindo a idade de ingresso das crianças para seis anos de idade com a Lei 11.274 de 6 de fevereiro de 2006. Isso também foi percebido pela inclusão do ensino infantil no FUNDEB (Fundo de Valorização do Magistério da Educação Básica), que substituiu em 2007 o FUNDEF, garantindo um valor mínimo de recursos para cada aluno matriculado. No entanto, apesar dessa redistribuição em um patamar mínimo, o problema de distribuição ainda existe, com os municípios mais ricos gastando mais e municípios pobres gastando menos.

Assim, diversas reformas, programas e políticas públicas nas últimas duas décadas têm tido sucesso no aumento da escolaridade como um todo. No

entanto, mesmo a qualidade educacional sendo o principal desafio destas políticas, os investimentos em quantidade da educação infantil ainda são muito baixos, justificando a verificação empírica dos efeitos que o ensino infantil pode ter sobre o desempenho das crianças no ensino básico, e portanto, na qualidade educacional.

2.3 MODELO TEÓRICO: VARIÁVEIS QUE CONDICIONAM A FORMAÇÃO DE HABILIDADES NO CICLO DE VIDA

Considere um caso de uma família representativa em que a única forma de herança é educação. Para simplificar o problema, suponha que existam apenas dois tipos de educação: a educação infantil e a educação primária, p_1 e p_2 , respectivamente. Os indivíduos altruístas acumulam capital humano com a educação recebida, de acordo com a seguinte função:

$$H_i(p_1, p_2) = a^{p_1} p_2 \quad (1)$$

A educação primária é a única obrigatória, assim, p_2 é uma variável exógena, e os indivíduos escolhem se alocam seus recursos em p_1 . A educação infantil é discreta, e pode ser 2 ou 0. Se os pais decidem alocar $p_1 = 2$, então aumentam o capital humano dos filhos em a^2 . Se alocam $p_1 = 0$, então o capital humano dos filhos será determinado apenas por p_2 , ou seja, somente a educação primária. A acumulação do capital humano determina a renda dos indivíduos $Y(H_i(p_1, p_2))$. A renda pode ser alocada entre consumo e o gasto no capital humano de seus filhos, conforme a equação (2):

$$Y(H_p(p_1, p_2)) = u(c) + H_f(p_1, p_2), \quad (2)$$

onde p e f representam a função para os pais e a função para os filhos, respectivamente. Assim, a resolução do problema de maximização da utilidade do indivíduo fica:

$$u(c) + H_f(p_1, p_2) - \lambda Y(H_p(p_1, p_2)) \quad (3)$$

Resolvendo o lagrangiano:

$$\frac{dL}{dc} = u'(c) \quad (4)$$

$$\frac{dL}{dp_1} = H'_f(p_1, p_2) - \lambda Y'(H_p(p_1, p_2)) = 0 \quad (5)$$

$$H'_f(p_1, p_2) = \lambda Y'(H_p(p_1, p_2)) \quad (6)$$

A equação (6) mostra que a acumulação do capital humano herdado pelo filho depende do capital humano que recebeu do pai e da renda que esse capital humano lhe permite receber. Consideremos agora o caso do pai que recebeu educação infantil (7), e o caso do pai que recebeu apenas a educação primária (8):

$$H'_f(p_1, p_2) = \lambda a^{p_1-1} p_1 p_2 \quad (7)$$

$$H'_f(p_1, p_2) = \lambda p_2 \quad (8)$$

Claramente, a equação (8) é menor do que a equação (7). Em outras palavras, o pai que recebeu uma educação menor, neste caso, não fez o ensino infantil, tende a ter uma renda menor, o que fará com que no equilíbrio da maximização da sua utilidade, este não invista na educação infantil dos seus filhos, gerando um círculo vicioso, em que famílias com baixa escolaridade tenderão a ter filhos com escolaridade e, por conseguinte, renda, baixas. Assim, faz-se necessária a intervenção por meio da política pública de enriquecimento de ambiente inicial, já que a renda e a escolaridade dos pais estão intimamente ligadas à seleção das crianças no ensino infantil. A próxima seção apresentará a estratégia empírica utilizada neste trabalho, controlando o problema de seleção pela renda e escolaridade dos pais.

3. ESTRATÉGIA EMPÍRICA

Para estimar o efeito de o aluno ter entrado no ensino infantil em sua nota no 5º ano, a primeira metodologia a ser utilizada será o método de *Propensity Score Matching* (PSM). O primeiro passo para o cálculo do PSM é a estimação de um modelo *probit* que busca identificar através de variáveis observáveis a probabilidade do aluno ter frequentado o ensino infantil. Assim, o tratamento nesta metodologia é o aluno ter iniciado os estudos no ensino infantil, e o controle, o aluno ter iniciado diretamente no ensino fundamental.

O ensino infantil pode ser dividido em duas partes: aqueles alunos que entraram no ensino de zero a 3 anos de idade, aqui chamado creche, e os que entraram na pré-escola, de 4 a 5 anos. Assim, serão três estimativas, e portanto, três variáveis de interesse: *inicio_creche*, *inicio_pre*, e *inicio_ensino_inf*, explicadas abaixo:

- a) *inicio_creche*: onde o grupo tratamento será o de alunos que iniciaram no ensino na creche, de 0 a 3 anos, e o grupo de controle, os alunos que iniciaram no ensino diretamente no ensino fundamental;
- b) *inicio_pre*, onde o grupo de tratamento será o de alunos que iniciaram na pré-escola, de 4 a 5 anos, e o de controle o grupo iniciou seus estudos no ensino fundamental;
- c) *inicio_prim_inf*, onde o grupo de tratamento será o de alunos que iniciaram no ensino infantil, ou seja, creche e pré-escola, e o grupo de controle serão os alunos que iniciaram diretamente no ensino fundamental.

Assim, serão feitos três pareamentos, controlando as variáveis que podem influenciar na seleção do aluno para o tratamento, a saber, a condição socioeconômica e a escolaridade dos pais. Após cada pareamento, serão estimados os impactos de ter feito o tratamento no desempenho de português e depois no desempenho de matemática. Isso para cada variável dependente, a saber, a nota de português no 5º ano, a

nota de matemática no 5º ano, a nota de português no 9º ano e a nota de matemática no 9º ano.

O segundo método utilizado, o Mínimos Quadrados Ponderados pelo *Propensity Score*, é uma regressão onde a variável dependente é o desempenho do aluno, o tratamento é uma das explicativas, e as outras são controles que podem influenciar no seu desempenho. Essa estimativa é considerada duplamente robusta, pois pondera a regressão de acordo com os *propensity scores*, ou seja, a probabilidade de o aluno estar no grupo de tratamento dadas suas características observáveis de seleção. Da mesma maneira como na metodologia anterior, serão feitas três estimativas, uma para a creche, uma para a pré-escola e uma para o ensino infantil como um todo. E também, para cada variável dependente, a saber, a nota de português no 5º ano, a nota de matemática no 5º ano, a nota de português no 9º ano e a nota de matemática no 9º ano.

3.1 APRESENTAÇÃO DA METODOLOGIA UTILIZADA NO TRABALHO

Dado o objetivo deste trabalho, de avaliar o impacto de ter frequentado a escola na primeira infância na nota do aluno no ensino básico, será preciso fazer uma avaliação de tratamento. Para tanto, será utilizada uma ferramenta econométrica, especificamente, o *propensity score matching*, que consiste em um método de pareamento. O objetivo do *matching* é encontrar um grupo ideal de comparação, já que a escolha de ter frequentado a escola na primeira infância pode ser influenciada por características observáveis. No *propensity score* serão controladas variáveis que podem influenciar na seleção do tratamento, ou seja, na escolha dos pais entre colocar seus filhos no ensino infantil ou diretamente no ensino fundamental. Após a estimativa do *propensity score*, serão estimados os efeitos do ensino infantil no desempenho dos alunos em português e matemática no 5º e 9º anos. Ainda, para controlar para variáveis que podem influenciar no desempenho dos alunos, serão estimadas regressões com o estimador de Mínimos Quadrados Ponderados pelo

Propensity Score, que possui a vantagem de ser um estimador duplamente robusto. Ambas as metodologias são brevemente apresentadas a seguir.

3.2 O ESTIMADOR *PROPENSITY SCORE MATCHING* (PSM)

Considerando a possibilidade de haver problemas de seleção por observáveis na frequência da educação infantil, utilizar-se-á um método de pareamento. O método de *propensity score*, é um método de pareamento (*matching*) inexato muito popular. Ao invés de fazer o pareamento nos regressores, ele faz nos escores ponderados, chamados *propensity scores*. Neste caso, um método de pareamento exato não é possível, então a unidade de comparação é aquela onde os *propensity scores* tendem suficientemente às unidades tratadas. Assim, nesse método, a medida utilizada para o pareamento é a probabilidade condicional de receber o tratamento dado x , que pode ser calculada através de um modelo *probit*. (Rosenbaum and Rubin, 1983).

Utilizando um estimador *Propensity Score Matching* (PSM), mesmo sem a existência de um contrafactual para estudos de efeito de um determinado tratamento sobre as variáveis, pode-se obter dados de um conjunto de unidades de comparação potenciais que não são necessariamente retiradas da mesma população das unidades tratadas, mas daqueles cujas características observáveis x pareadas com aquelas do grupo tratado são iguais em um determinado grau. Assim, a média dos resultados dos não tratados pareados com os grupos identifica a média contrafactual resultante no grupo tratado na ausência de tratamento. (Becker e Ichino, 2002).

Em outras palavras, a medida contrafactual seria saber como seria o desempenho dos alunos no 5º e 9º anos se não tivessem frequentado o ensino infantil. O que o método faz é identificar grupos com crianças de características muito próximas, mas que diferem na questão de ter frequentado o ensino infantil, ou começado os estudos diretamente no ensino fundamental.

Seja o grupo de comparação para o caso tratado i com características x_i denotado pelo conjunto $A_j(x) = \{j \mid x_j \in c(x_i)\}$, onde $c(x_i)$ são as características vizinhas de x_i . Seja NC o número de casos no grupo de comparação e $\omega(i, j)$ o

peso dado para o j -ésimo caso que faz uma comparação com o i -ésimo caso tratado, $\sum_j \omega(i,j) = 1$. Então a fórmula geral para o pareamento estimado é:

$$\Delta^M = \frac{1}{N_T} \sum_{i \in \{D=1\}} [y_{1,i} - \sum_j \omega(i,j) y_{0,j}], \quad (1)$$

Onde $0 < \omega(i,j) \leq 1$, e $\{D=1\}$ é o conjunto dos indivíduos tratados, e j é um elemento do conjunto de unidades de comparação casadas. A diferença dos estimadores de pareamento é geralmente pela variação da escolha de $\omega(i,j)$. (CAMERON et al. 2005).

O pareamento simples compara as células com exatamente o mesmo x discreto,

$$\Delta^M = \sum_k \omega_k [\bar{y}_{1,k} - \bar{y}_{0,k}], \quad (2)$$

Onde \bar{y}_1 é o resultado médio dos tratados e \bar{y}_0 é o resultado médio dos não-tratados e ω_k é o peso da k -ésima célula (isto é, a fração de observações na célula k). (CAMERON et al. 2005).

O estimador utilizado aqui será o *stratification* ou *interval matching*, que é baseado na ideia de dividir o intervalo de variação dos *propensity scores* em intervalos tais que dentro de cada intervalo as unidades de controles e tratados tenham, na média, o mesmo escore de propensão. Pode-se usar os mesmos blocos identificados no cálculo do PSM. Então calcula-se a diferença entre a média dos resultados do grupo dos tratados e de controle. O ATET (Efeito médio do tratamento nos tratados) é a média ponderada destas diferenças, com os pesos sendo determinados pela distribuição das unidades tratadas entre os blocos. Uma das desvantagens deste método é que ele descarta observações nos blocos em que cada unidade tratada ou controle esteja ausente.

Seja b os blocos definidos sobre os intervalos dos escores de propensão. Então o efeito tratamento dentro do b -ésimo bloco é definido como:

$$ATET_b^S = (N_b^T)^{-1} \sum_{i \in I(b)} Y_{1i} - (N_b^C)^{-1} \sum_{j \in I(b)} Y_{0j}, \quad (3)$$

onde:

$I(b)$ = conjunto de unidades no bloco b ;

N_b^T = é o número de unidades tratadas no b -ésimo bloco;

N_b^C = é o número de unidades de controle no b -ésimo bloco.

Então, o efeito tratamento baseado na estratificação é definido como:

$$ATET^S = \sum_{b=1}^B ATET_b^S \times \left[\frac{\sum_{i \in I(b)} D_i}{\sum D_i} \right], \quad (4)$$

onde o termo entre colchetes é o peso de cada bloco dado pela fração correspondente das unidades tratadas, e B é o número total de blocos.

Segundo Heckman e Navarro-Lozano (2003), esses métodos de pareamento não exigem exogeneidade das variáveis explicativas, o que para métodos menos sofisticados é uma suposição muito importante. Aliás, para dados relacionados à economia da educação, isso é um problema bastante corriqueiro, já que a suposição de exogeneidade das variáveis é muito forte. Para cumprir o propósito de medir o efeito de ter frequentado a creche ou pré-escola (educação infantil) sobre as habilidades cognitivas no ensino básico (nota no Prova Brasil no 5º e 9º anos), irá ser utilizada a base de dados do Prova Brasil 2011, para estimar um modelo de pareamento usando *propensity score* (PSM). A justificativa na utilização desta metodologia é a possível existência de seleção por observáveis – como a renda familiar, características familiares como morar com a mãe, e a escolaridade dos pais – na escolha entre frequentar ou não o ensino infantil. Assim, calcular-se-á as probabilidades (*propensity scores*) com um modelo *probit*, depois serão selecionadas as

crianças pareadas utilizando o estimador *Stratification Matching*. Assim, as médias dos dois grupos poderão ser comparadas.

3.3 O ESTIMADOR DOS MÍNIMOS QUADRADOS PONDERADOS PELO PROPENSITY SCORE

Segundo Imbens e Wooldridge (2008), estimativas de efeitos controlando para características observáveis com o estimador de Mínimos Quadrados Ponderados pelo *Propensity Score*, contornam o problema de má especificação, por ser uma combinação do *propensity score* com a regressão, que tem o benefício de ser “duplamente robusto”. Esse problema pode estar na equação da regressão ou do *propensity score*, mas nunca em ambos. Assim, estes métodos chamados como duplamente robustos, removeriam o efeito direto das variáveis omitidas na regressão, e reduziriam a correlação entre as variáveis omitidas e incluídas, reponderando-as.

Segundo Becker e Ichino (2002), para a realização do Mínimos Quadrados Ponderados pelo *Propensity Score*, são necessários dois procedimentos: o primeiro é o cálculo do *propensity score*, que pode ser $Y_i = \alpha + \tau \cdot W_i + \varepsilon_i$, feito através de um modelo probit ou logit, como já explicado na seção anterior. No segundo procedimento, estima-se o Mínimos Quadrados Ponderados pelos *propensity score* propriamente dito.

Segundo Imbens (2004), a regressão estimada é com pesos iguais a:

$$\lambda_i = \sqrt{\frac{W_i}{e(X_i)} + \frac{1-W_i}{1-e(X_i)}} \quad , \text{ onde} \quad (1)$$

Y_i = resultado

W_i = variável de participação no programa (0 ou 1)

$e(X_i)$ = *propensity score*

X_i = características observadas que explicam a participação no programa

λ_i = peso

O coeficiente τ estimado é um estimador consistente para o efeito médio do tratamento (ATE). Essa regressão além de diminuir o viés, aumenta a precisão da estimativa, sendo assim com os pesos dados pelo *propensity score* o estimador de mínimos quadrados se torna consistente. Sendo \hat{e} um estimador consistente do *propensity score* e garantindo que cada termo dentro da raiz quadrada some 1, o estimador por mínimos quadrados ponderados é consistente para o ATE:

$$\hat{\lambda}_i = \sqrt{\frac{W_i / \hat{e}(X_i)}{\sum_{i=1}^N W_i / \hat{e}(X_i)} + \frac{(1-W_i) / (1-\hat{e}(X_i))}{\sum_{i=1}^N (1-W_i) / (1-\hat{e}(X_i))}} \quad (2)$$

em que $\hat{e}(X_i)$ é o *propensity score* estimado.

Para se estimar o efeito tratamento sobre os tratados (ATT), o ponderador

$$\hat{\lambda}_i = \sqrt{\frac{W_i}{\sum_{i=1}^N W_i} + \frac{(1-W_i) \cdot \hat{e}(X_i) / (1-\hat{e}(X_i))}{\sum_{i=1}^N (1-W_i) \cdot \hat{e}(X_i) / (1-\hat{e}(X_i))}} \quad (3)$$

produz o estimador do coeficiente τ que é consistente para o ATT.

4. DESCRITIVAS DO PROVA BRASIL 2011

Dada a importância evidenciada da educação infantil para o ciclo educacional como um todo, este capítulo tem por objetivo utilizar os dados da Prova Brasil 2011 para estimar o impacto que a educação infantil tem sobre o desempenho dos alunos no ensino básico. A Prova Brasil é uma base de dados muito completa para esse fim, pois apresenta além do desempenho do aluno em determinado momento no tempo, suas condições sociais.

4.1 BASE DE DADOS E ESTATÍSTICAS DESCRITIVAS – PROVA BRASIL 2011

Desenvolvida pelo Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira (INEP/MEC), a Prova Brasil e o Sistema Nacional de Avaliação da Educação Básica (SAEB) têm o objetivo de avaliar a qualidade do ensino no Brasil através de testes padronizados e questionários socioeconômicos. Esses testes são aplicados no quinto e nono anos apenas das escolas públicas, nas redes estaduais, municipais e federais, de área rural e urbana, nas escolas que possuíam no mínimo 20 alunos matriculados em cada ano avaliado. Os testes levam em consideração os conhecimentos dos alunos em matemática e língua portuguesa, e os questionários abrangem informações socioeconômicas que ajudam no processo de relacioná-las com o desempenho destes. É importante salientar que os questionários são respondidos pelos próprios alunos, que podem desconhecer ou responder erroneamente alguma questão mais complexa, uma das limitações deste tipo de banco de dados.

Mais do que informações oriundas dos alunos, a Prova Brasil coleta informações relacionadas aos professores e diretores, como condições de trabalho, dados demográficos e perfil profissional. Além de auxiliar no aprimoramento da qualidade da educação e redução das desigualdades, as médias de desempenho subsidiam o cálculo do Índice de Desenvolvimento da Educação Básica (Ideb), e estão disponíveis para o público, o que pode ser muito útil já que os dados são comparáveis ao longo do tempo e são ampliados a cada dois anos.

4.1.1 Descritivas Prova Brasil 2011 – 5º ano

Antes das estimações, algumas informações adquiridas dos dados brutos podem ser úteis para dar um panorama das características do aluno de escola pública no Brasil. Isso pode ser importante para se traçar um perfil do aluno que frequenta o ensino público no Brasil, já que a análise deste deve levar em conta essas diferenças.

Para o 5º ano, o banco de dados apresenta informações de 1.792.242 alunos, 5.296 municípios e 40.799 escolas. Quanto à variável sexo, o banco de dados é dividido em 50,61% do gênero masculino e 49,39% do gênero feminino. Apenas 10,41% dos alunos se auto declaram negros. Uma parcela significativa dos alunos (40,48%) iniciou seu ensino na creche, ou seja, antes dos 3 anos de idade, 41,01% dos alunos iniciaram na pré-escola, de 4 a 5 anos de idade, e somente 18,51% dos alunos iniciaram diretamente no ensino fundamental.

Um percentual de 8,54% disseram que não moram com a mãe, e 4,79% disseram que suas mães não sabem ler, mas mais que isso, 10,43% dos alunos disseram não verem suas mães lendo. O pai parece ser uma presença um pouco mais afastada: 31,27% disseram não morar com o pai, que também possui uma alfabetização menor que a da mãe, já que 7,05% dos pais não sabem ler, e praticamente 20% dos alunos não vêem seus pais lendo.

Quanto ao grau de instrução das mães, 10,29% das mães completaram o ensino superior, 12,16% completaram o ensino médio, 12% completaram o oitava série, 14,36% completaram a quarta série e 9,5% não completaram nem a quarta série. 2,62% das mães nunca estudaram e 36% dos alunos não souberam responder ao questionário, o que torna a variável de má qualidade. No caso dos pais, 9,41% concluíram o ensino superior, 9% completaram o ensino médio, 10% completaram a oitava série, 10% completaram a quarta série, 8% não completaram nem a quarta série, e 3,67% nunca estudaram. No entanto, 42,32% dos alunos não souberam responder ao questionário.

Quando questionadas se os pais incentivam a estudar, apenas 2,42% disseram que não. No entanto 8,74% dos alunos disseram que seus pais não comparecem nunca ou quase nunca às reuniões escolares. O perfil típico de aluno do 5º ano não costuma ler jornais (apenas 14% disseram ler sempre ou

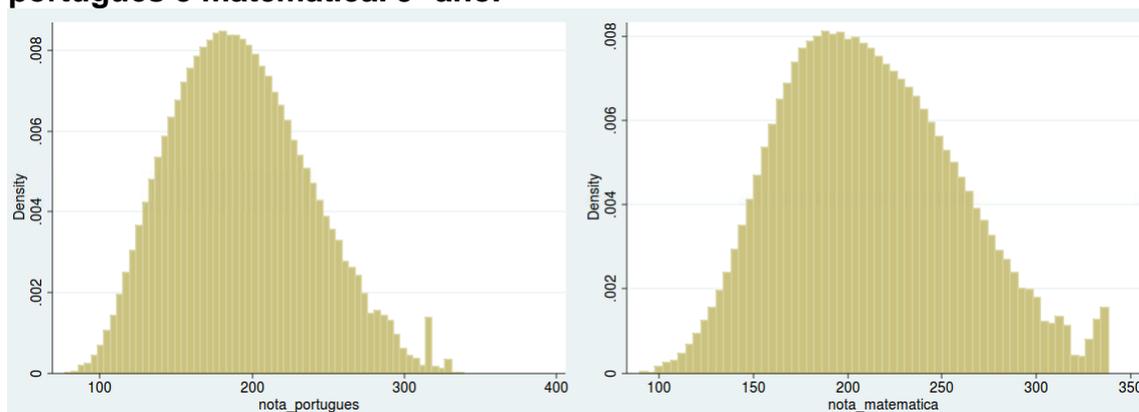
quase sempre), apenas 42% admitem terem o hábito de lerem livros em geral. 20% dos alunos admitem não lerem quase nunca livros de literatura. Apenas 24% frequentam a biblioteca da escola sempre ou quase sempre, e mais de 30% admitem frequentarem a biblioteca nunca ou quase nunca. 30% dos alunos afirmaram gastar mais de 4 horas por dia assistindo TV, navegando na internet ou jogando jogos eletrônicos. Praticamente 80% dos alunos disseram que seus professores corrigem as lições de português e matemática sempre ou quase sempre.

Além de questões sobre as características da família – se o aluno mora com a mãe, a quantidade de pessoas que moram na mesma casa, se o aluno vê a mãe lendo, se a mãe tem o ensino superior completo e se o pai sabe ler – o Prova Brasil 2011 contém questões sobre o contexto socioeconômico dos alunos– se o aluno possui empregada em casa, se possui televisão, carro, freezer, geladeira, máquina de lavar, e variáveis que podem influenciar no desempenho dos alunos, como se o aluno trabalha fora, se já abandonou a escola alguma vez, se já reprovou, se é negro, se frequenta a biblioteca, se lê livros, se faz o dever de português e de matemática, e se sempre estudou em escola pública.

No grupo de variáveis de contexto socioeconômico, 5% dos alunos disseram não ter televisão em casa, 3,08% disseram não ter uma geladeira, e 78% disseram não ter um freezer separado da geladeira. A grande maioria, 55,13% disse não ter carro, e 48% dos alunos disseram não possuir computador em casa. 30% dos alunos disseram não ter máquina de lavar em casa, e mais de 90% não possuem uma empregada.

No grupo das variáveis que podem influenciar o desempenho do aluno, 12% dos alunos disseram trabalhar fora, 6,5% disseram já ter abandonado a escola ao menos uma vez, 78% dos alunos disseram sempre ter estudado sempre em escola pública e 28% disseram ter reprovado ao menos uma vez, o que é um número expressivo para alunos do 5º ano.

Gráfico 4. Prova Brasil 2011. Distribuição de frequências das notas em português e matemática. 5º ano.



Fonte: Prova Brasil 2011/Elaboração própria.

O gráfico 3 apresenta a distribuição de frequências das notas em português e matemática do Prova Brasil 2011 para os alunos da 4ª série. A distribuição das notas de português apresenta um histograma suavemente assimétrico para a esquerda, indicando uma média não centrada de 192 pontos na escala única do SAEB – o Prova Brasil utiliza a mesma metodologia do SAEB, baseada na Teoria de Resposta ao Item (TRI), o que permite a comparação entre as notas –, máximo de 339 pontos e mínimo de 77 pontos. O desvio-padrão para a nota de português é de 45 pontos. Já a distribuição de notas em matemática apresenta um histograma também suavemente assimétrico para esquerda, representando um desempenho médio melhor dos alunos nessa disciplina, com média de 211 na escala única do SAEB, máximo de 338 pontos e mínimo de 90 pontos. O desvio-padrão das notas de matemática é de 46 pontos.

As variáveis utilizadas nas estimativas do 5º ano estão descritas na tabela abaixo:

Tabela 1. Descrição das variáveis utilizadas para alunos do 5º ano.

Variável	Descrição	Média	Desvio Padrão	Nº de Obs.
nota_portugues	Proficiência do aluno em Língua Portuguesa calculada na escala única do SAEB.	190.7	45.4	1.792.242
nota_matematica	Proficiência do aluno em Matemática calculada na	209.8	47.0	1.792.242

Variável	Descrição	Média	Desvio Padrão	Nº de Obs.
	escala única do SAEB, com média = 250 e desvio = 50			
inicio_creche	Se o aluno iniciou os estudos na creche ou no ensino fundamental. Creche = 1, Ensino fundamental = 0.	0.68	0.46	972.034
inicio_pre	Se o aluno iniciou os estudos na pré-escola ou no ensino fundamental. Pré = 1, Ensino fundamental = 0.	0.63	0.48	1.078.160
apenas_creche	Se o aluno iniciou os estudos no creche ou no pré. Creche = 1, Pré = 0.	0.49	0.49	1.350.050
inicio_prim_inf	Se o aluno iniciou os estudos no ensino infantil (creche ou pré) ou no ensino fundamental. Ensino Infantil = 1, Ensino fundamental = 0.	0.81	0.38	1.655.402
mora_mae	Se o aluno mora com a mãe. Sim = 1, Não = 0.	0.91	0.27	1.772.100
Mae_le	Se o aluno vê sua mãe lendo. Sim = 1, Não = 0.	0.89	0.30	1.749.216
mae_superior_completo	Se a mãe completou o ensino superior. Sim = 1, Não = 0.	0.16	0.37	1.093.586
pai_sabe_ler	Se o pai do aluno é alfabetizado. Sim = 1, Não = 0.	0.93	0.25	1.569.140
mae_sabe_ler	Se a mãe do aluno é alfabetizada. Sim = 1, Não = 0.	0.95	0.20	1.730.303
Empregada	Se o aluno possui empregada em casa. Sim = 1, Não = 0.	0.09	0.29	1.779.659
Computador	Se o aluno possui computador em casa. Sim = 1, Não = 0.	0.50	0.49	1.778.768
tv_em_casa	Se o aluno possui TV em casa. Sim = 1, Não = 0.	0.95	0.20	1.792.242
trab_fora	Se o aluno trabalha fora de	0.12	0.33	1.743.760

Variável	Descrição	Média	Desvio Padrão	Nº de Obs.
	casa. Sim = 1, Não = 0.			
freezer_em_casa	Se o aluno possui freezer em casa (separado da geladeira). Sim = 1, Não = 0.	0.22	0.41	1.698.793
mora_ate_5	Se na casa do aluno possui até 5 pessoas. Sim = 1, Não = 0.	0.76	0.42	1.767.013
radio_em_casa	Se o aluno possui rádio em casa. Sim = 1, Não = 0.	0.82	0.37	1.792.242
sempre_escola_publica	Se o aluno estudou sempre em escola pública. Sim = 1, Não = 0.	0.77	0.41	1.755.795
abandonou_escola	Se o aluno abandonou a escola ao menos uma vez. Sim = 1, Não = 0.	0.071	0.25	1.763.020
maquina_de_lavar_em_casa	Se o aluno possui máquina de lavar em casa. Sim = 1, Não = 0.	0.69	0.45	1.769.494
geladeira_em_casa	Se o aluno possui geladeira em casa. Sim = 1, Não = 0.	0.96	0.17	1.792.242
Carro	Se o aluno possui ao menos 1 carro. Sim = 1, Não = 0.	0.44	0.49	1.774.896
Reprovou	Se o aluno já reprovou ao menos uma vez. 1 = Sim, Não = 0.	0.28	0.44	1.755.550
Negro	Se o aluno se considera negro. 1 = Sim, Não = 0.	0.10	0.30	1.568.979
Freq._biblioteca	Se o aluno respondeu frequentar a biblioteca sempre ou quase sempre. 1 = Sim, Não = 0	0.26	0.44	1.646.964
Le_livros_sempre	Se o aluno respondeu ler livros em geral sempre ou quase sempre. 1 = Sim, Não = 0.	0.45	0.50	1.652.614
Faz_dever_port_sempre	Se o aluno respondeu fazer sempre ou quase sempre o dever de português. 1 = Sim,	0.74	0.43	1.751.694

Variável	Descrição	Média	Desvio Padrão	Nº de Obs.
	Não = 0.			
Faz_dever_mat_sempre	Se o aluno respondeu fazer sempre ou quase sempre o dever de matemática. 1 = Sim, Não = 0.	0.78	0.42	1.760.924

4.1.2 Descritivas Prova Brasil 2011 – 9º ano

Os dados do 5º e 9º anos tratam de alunos que responderam o questionário no mesmo ano de 2011 e, portanto, não mostram a evolução dos mesmos alunos no tempo. O ideal seria a existência de dados longitudinais para uma melhor análise. No entanto, as descritivas podem apontar uma boa aproximação das mudanças no perfil dos alunos ao longo do tempo.

Para o 9º ano, o banco de dados apresenta informações de 1.984.181 alunos, 5.458 municípios e 31.599 escolas. Quanto à variável sexo, o banco de dados é dividido em 46,94% do gênero masculino e 53,06% do gênero feminino. Apenas 11,03% dos alunos se autodeclararam negros. 33% dos alunos iniciaram seu ensino na creche, um número menor que os 40,48% do 5º ano. Sendo os alunos do 9º ano mais velhos que os do 5º ano, pode-se supor que ou o acesso à creche aumentou, ou os pais estão colocando os filhos na escola mais cedo. 46% dos alunos iniciaram na pré-escola, de 4 a 5 anos de idade, e 19% dos alunos iniciaram diretamente no ensino fundamental. Os outros 2% iniciaram depois do ensino fundamental.

Na base de dados dos alunos do nono ano, 5,16% disseram não morar com a mãe, 6,04% disseram que suas mães são analfabetas. 14,26% disseram não verem suas mães lendo, um número maior que os 10,43% do 5º ano, o que pode representar uma piora no comportamento dos pais com o passar dos anos da criança na escola. 33,40% disseram não morar com o pai, e 9,04% disseram que seu pai é analfabeto. Um número significativo, 29,83% disseram não ver o pai lendo. Por fim, 33,87% dos alunos disseram terem reprovado ao menos uma vez.

Quanto à escolaridade da mãe, 7,20% dos alunos disseram que suas mães completaram o ensino superior. 21,53% disseram que suas mães completaram o ensino médio, 15,87% a oitava série, 20% a quarta série. 13% das mães não completaram a quarta série, 3% nunca estudaram, e 16,71% dos alunos disseram não saber a escolaridade de suas mães – um número bem menor do que os 36% de não conhecedores da situação materna no 5º ano, o que torna esta variável mais representativa que aquela. Quanto à escolaridade do pai, 5% dos alunos disseram que seus pais completaram o ensino superior, 16,17% o ensino médio, 13,43% a oitava série, 15,20% a quarta série. 12% dos pais não completaram a quarta série, 4% nunca estudaram, e 23% dos alunos disseram não saber a escolaridade de seus pais.

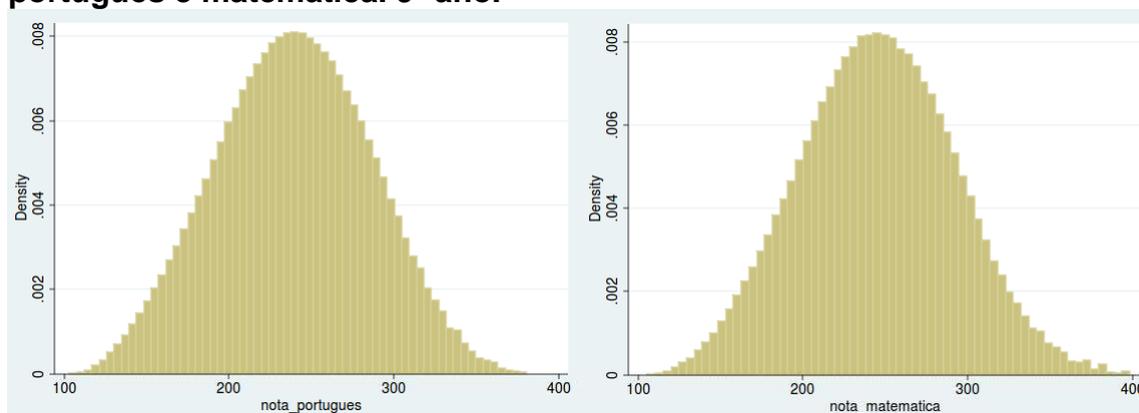
Quando questionadas se os pais incentivam a estudar, apenas 1,40% disseram que não, mas 7,48% dos alunos disseram que seus pais não comparecem nunca ou quase nunca às reuniões escolares. O perfil típico de aluno do 9º ano parece mostrar uma piora com relação aos alunos do 5º ano: o aluno de 9º ano não costuma ler jornais (apenas 11,46% disseram ler sempre ou quase sempre), e apenas 27,60% admitem terem o hábito de lerem livros em geral sempre ou quase sempre, um número abaixo dos 42% que admitem ter esse hábito no 5º ano. 28,97% dos alunos admitem não lerem quase nunca livros de literatura, um aumento com relação aos 20% do 5º ano.

Apenas 12,28% frequentam a biblioteca da escola sempre ou quase sempre (no 5º ano 24% admitem ter esse hábito), e mais de 37% admitem frequentarem a biblioteca nunca ou quase nunca. Mais de 30% dos alunos afirmaram gastar mais de 4 horas por dia assistindo TV, navegando na internet ou jogando jogos eletrônicos. Mais de 80% dos alunos disseram que seus professores corrigem as lições de português e matemática sempre ou quase sempre.

Quanto ao contexto socioeconômico, quase 3% dos alunos disseram não ter televisão em casa, 2,36% disseram não ter uma geladeira, e quase 80% disseram não ter um freezer separado da geladeira. A grande maioria, 54,07% disse não ter carro, e 41,29% dos alunos disseram não possuir computador em casa. Quase 33% dos alunos disseram não ter máquina de lavar em casa, e mais de 90% não possuem uma empregada. Com relação aos alunos do 5º ano, o perfil socioeconômico dos alunos do 9º ano medido por

essas proxies não apresenta grande variação, e em geral demonstram um perfil de alunos em um ambiente em desvantagem.

Gráfico 5. Prova Brasil 2011. Distribuição de frequências das notas em português e matemática. 9º ano.



Fonte: Prova Brasil 2011/Elaboração própria.

O gráfico 4 apresenta a distribuição de frequências das notas em português e matemática do Prova Brasil 2011 para os alunos do 9º ano. A distribuição das notas de português e matemática se mostram muito mais simétricas com relação à distribuição das notas para o 5º ano. A nota em português apresenta uma média de 239 pontos na escala única do SAEB, com máximo em 380 pontos e mínimo de 103 pontos. O desvio-padrão para a nota de português é de 46,7 pontos. Já a distribuição de notas em matemática uma média de 245 pontos na escala única do SAEB, máximo de 398 pontos e mínimo de 105 pontos. O desvio-padrão em matemática é de 47 pontos.

As variáveis utilizadas nas estimativas do 9º ano estão descritas na tabela abaixo:

Tabela 2. Descrição das variáveis utilizadas para alunos do 9º ano.

Variável	Descrição	Média	Desvio Padrão	Nº de Obs.
nota_portugues	Proficiência do aluno em Língua Portuguesa calculada na escala única do SAEB.	239.0	46.7	1.984.181
nota_matematica	Proficiência do aluno em Matemática calculada na escala única do SAEB.	245.4	47.0	1.984.181

Variável	Descrição	Média	Desvio Padrão	Nº de Obs.
inicio_creche	Se o aluno iniciou os estudos na creche ou no ensino fundamental. Creche = 1, Ensino fundamental = 0.	0.32	0.46	1.049.898
inicio_pre	Se o aluno iniciou os estudos na pré-escola ou no ensino fundamental. Pré-escola = 1, Ensino fundamental = 0.	0.68	0.46	1.307.632
Apenas_creche	Se o aluno iniciou os estudos na creche ou no pré. Creche = 1, Pré-escola = 0.	0.19	0.39	1.524.378
Inicio_prim_inf	Se o aluno iniciou os estudos no ensino infantil (creche ou pré) ou no primario. Ensino Infantil = 1, Ensino fundamental = 0.	0.80	0.39	1.892.103
Pai_superior_completo	Se o pai do aluno concluiu o ensino superior. Sim = 1, Não = 0.	0.07	0.26	1.289.514
Mae_superior_completo	Se a mãe do aluno concluiu o ensino superior. Sim = 1, Não = 0.	0.08	0.28	1.555.848
pai_sabe_ler	Se o pai do aluno é alfabetizado. Sim = 1, Não = 0.	0.90	0.28	1.647.308
Mae_sabe_ler	Se a mãe do aluno é alfabetizada. Sim = 1, Não = 0.	0.93	0.23	1.893.689
Empregada	Se o aluno possui empregada em casa. Sim = 1, Não = 0.	0.09	0.29	1.965.865
Tv_em_casa	Se o aluno possui TV em casa. Sim = 1, Não = 0.	0.97	0.14	1.953.835
Freezer_em_casa	Se o aluno possui freezer em casa (separado da geladeira). Sim = 1, Não = 0.	0.20	0.40	1.881.834
Carro	Se o aluno possui ao menos um carro. Sim = 1, Não = 0.	0.46	0.49	1.924.157
K7_dvd_em_casa	Se o aluno possui aparelho de vídeo-cassete ou DVD em casa. Sim = 1, Não = 0.	0.89	0.30	1.914.259
Geladeira.	Se o aluno possui geladeira em casa. Sim = 1, Não = 0.	0.97	0.15	1.924.736
mora_ate_5	Se na casa do aluno possui até 5	0.83	0.37	1.965.419

Variável	Descrição	Média	Desvio Padrão	Nº de Obs.
	peças. Sim = 1, Não = 0.			
Radio_em_casa	Se o aluno possui rádio em casa. Sim = 1, Não = 0.	0.86	0.33	1.960.860
Sempre_escola_publica	Se o aluno estudou sempre em escola pública. Sim = 1, Não = 0.	0.89	0.30	1.949.576
Trab_fora	Se o aluno trabalha fora. Sim = 1, Não = 0.	0.19	0.39	1.912.891
Abandonou_escola	Se o aluno abandonou a escola ao menos uma vez. Sim = 1, Não = 0.	0.05	0.23	1.955.834
Reprovou	Se o aluno reprovou ao menos uma vez. Sim = 1, Não = 0.	0.33	0.47	1.951.616
Negro	Se o aluno se considera negro. Sim = 1, Não = 0.	0.11	0.31	1.831.128
Freq._biblioteca	Se o aluno respondeu frequentar a biblioteca sempre ou quase sempre. Sim = 1, Não = 0.	0.13	0.34	1.819.790
Le_livros_sempre	Se o aluno respondeu ler livros em geral sempre ou quase sempre. Sim = 1, Não = 0.	0.30	0.46	1.831.366
Faz_dever_port_sempre	Se o aluno respondeu fazer o dever de português sempre ou quase sempre. Sim = 1, Não = 0.	0.58	0.49	1.948.543
Faz_dever_mat_sempre	Se o aluno respondeu fazer o dever de matemática sempre ou quase sempre. Sim = 1, Não = 0.	0.56	0.49	1.949.039

5. ANÁLISE DOS RESULTADOS

As tabelas 3 e 4 apresentam os resultados do modelo *Probit*, resultados esses, que são utilizados no cálculo do *propensity Score*. A análise da significância e sinal dos coeficientes verifica se as variáveis escolhidas para controlar a seleção por observáveis, tanto para a condição socioeconômica, quanto a escolaridade dos pais, realmente influenciam na probabilidade do aluno iniciar seus estudos no ensino infantil.

5.1 ANÁLISE DOS RESULTADOS OBTIDOS PELO *PROPENSITY SCORE* 5º ANO

Como explanado na estratégia empírica, três estimativas feitas para cada questionário do Prova Brasil, do 5º ano, e do 9º ano serão apresentadas a seguir. A tabela 3 apresenta os resultados estimados do modelo *probit* para o cálculo do *propensity score* para as três estimativas: 1) creche, 2) pré-escola e 3) primeira infância, para o 5º ano. No apêndice deste trabalho, são apresentados além das estimativas separadas por modelo, também os blocos de agrupamento de comparação entre os grupos de controle e de tratamento. O número de blocos, determinado pelo algoritmo de estimação do escore de propensão, assegura que a média do escore de propensão não seja diferente para tratados e controles em cada bloco.

Tabela 3. Resultados estimados do modelo *Probit* – 5º ano.

Estimativa	Variáveis	Coef.	Erro padrão	Z	P> z	Intervalo de confiança de 95%	
Creche	mae_superior_completo	0.297	0.005	56.85	0.000	0.287	0.307
Creche	Empregada	0.163	0.006	26.01	0.000	0.151	0.175
Creche	Computador	0.213	0.0039	54.58	0.000	0.206	0.221
Creche	tv_em_casa	0.233	0.0095	24.52	0.000	0.214	0.252
Creche	pai_sabe_ler	0.125	0.007	17.56	0.000	0.111	0.139
Creche	freezer_em_casa	-0.081	0.004	-18.31	0.000	-0.089	-0.072
Creche	_cons	-0.005	0.0111	0.48	0.631	-0.016	0.027
Pré-escola	Maquina_de_lavar_em_casa	0.015	0.0038	4.11	0.000	0.008	0.023
Pré-escola	Geladeira_em_casa	0.227	0.009	23.31	0.000	0.208	0.246

Estimativa	Variáveis	Coef.	Erro padrão	Z	P> z	Intervalo de confiança de 95%	
Pré-escola	Pai_sabe_ler	0.237	0.006	36.84	0.000	0.224	0.249
Pré-escola	Radio_em_casa	0.064	0.004	13.83	0.000	0.055	0.073
Pré-escola	Mae_superior_completo	0.077	0.005	15.78	0.000	0.067	0.087
Pré-escola	Tv_em_casa	0.205	0.008	24.45	0.000	0.189	0.222
Pré-escola	_cons	-0.330	0.0127	-26.01	0.000	-0.355	-0.305
Ensino Infantil	Empregada	0.092	0.0054	16.97	0.000	0.081	0.102
Ensino Infantil	Pai_sabe_ler	0.115	0.0060	19.00	0.000	0.103	0.127
Ensino Infantil	Mae_le	0.152	0.0054	28.24	0.000	0.142	0.163
Ensino Infantil	Mae_superior_completo	0.191	0.0045	42.17	0.000	0.182	0.200
Ensino Infantil	Carro	0.088	0.0032	26.93	0.000	0.082	0.095
Ensino Infantil	Geladeira_em_casa	0.154	0.0091	16.95	0.000	0.136	0.172
Ensino Infantil	Tv_em_casa	0.178	0.0078	22.65	0.000	0.163	0.194
Ensino Infantil	Radio_em_casa	0.134	0.0043	3.11	0.002	0.004	0.021
Ensino Infantil	Mora_ate_5	0.156	0.0037	42.09	0.000	0.149	0.164
Ensino Infantil	_cons	0.138	0.0125	11.04	0.000	0.114	0.163

Fonte: Dados estimados a partir do Prova Brasil 2011 por meio do *software stata 11*.
Elaboração própria.

Os resultados do modelo *Probit* da tabela 3 mostram que todas as variáveis incluídas no modelo apresentam-se altamente significativas, com exceção da variável *radio_em_casa* para a estimativa 3 do ensino infantil, significante a 5%. Os sinais dos coeficientes mostram haver uma relação positiva entre a condição socioeconômica e a escolaridade dos pais na probabilidade de entrar no ensino infantil, seja na creche ou na pré-escola.

5.2 ANÁLISE DOS RESULTADOS OBTIDOS PELO *PROPENSITY SCORE MATCHING* – 9º ANO

A tabela 4 apresenta os resultados estimados do modelo *probit* para o cálculo do *propensity score* para as três estimativas: 1) creche, 2) pré-escola e 3) primeira infância, para o 9º ano. No apêndice deste trabalho, são

apresentados além das estimativas separadas por modelo, também os blocos de agrupamento de comparação entre os grupos de controle e de tratamento. O número de blocos, determinado pelo algoritmo de estimação do escore de propensão, assegura que a média do escore de propensão não seja diferente para tratados e controles em cada bloco.

Tabela 4. Resultados estimados do modelo *Probit* – 9º ano.

Estimativa	Variáveis	Coef.	Erro padrão	Z	P> z	Intervalo de confiança de 95%	
Creche	Mora_ate_5	0.192	0.0037	51.63	0.000	0.185	0.200
Creche	Mae_sabe_ler	0.374	0.0057	64.74	0.000	0.363	0.385
Creche	Carro	0.047	0.0029	15.94	0.000	0.041	0.052
Creche	Empregada	0.205	0.0049	41.59	0.000	0.195	0.215
Creche	Mae_superior_completo	0.499	0.0054	90.87	0.000	0.488	0.510
Creche	_cons	-0.302	0.0061	-49.35	0.000	-0.314	-0.290
Pré-escola	Radio_em_casa	0.102	0.0043	23.36	0.000	0.093	0.110
Pré-escola	Freezer_em_casa	-0.099	0.0034	-28.76	0.000	-0.106	-0.092
Pré-escola	Tv_em_casa	0.352	0.0092	37.94	0.000	0.334	0.370
Pré-escola	Mae_superior_completo	0.339	0.0058	57.84	0.000	0.327	0.350
Pré-escola	Pai_sabe_ler	0.338	0.0047	70.66	0.000	0.328	0.347
Pré-escola	_cons	-0.227	0.0104	-21.79	0.000	-0.247	-0.206
Ensino Infantil	Pai_superior_completo	0.325	0.0056	57.16	0.000	0.314	0.336
Ensino Infantil	Empregada	0.170	0.0047	36.12	0.000	0.160	0.179
Ensino Infantil	Tv_em_casa	0.146	0.0090	16.20	0.000	0.128	0.164
Ensino Infantil	K7_dvd_em_casa	0.264	0.0044	59.63	0.000	0.256	0.273
Ensino Infantil	Geladeira_em_casa	0.332	0.0084	39.43	0.000	0.315	0.348
Ensino Infantil	_cons	0.124	0.0104	11.95	0.000	0.104	0.145

Fonte: Dados estimados a partir do Prova Brasil 2011 por meio do *software stata 11*.

Elaboração própria.

Os resultados do modelo *Probit* da tabela 4 mostram que todas as variáveis incluídas no modelo do 9º ano apresentam-se altamente significativas. Os sinais dos coeficientes, assim como para o 5º ano, mostram haver uma relação positiva entre a condição socioeconômica e a escolaridade dos pais na probabilidade do aluno iniciar seus estudos no ensino infantil, seja na creche ou

na pré-escola. Isso evidencia que a seleção por observáveis como renda e escolaridade dos pais realmente se verifica empiricamente.

5.3 ESTIMAÇÃO DO IMPACTO DE TER FREQUENTADO A EDUCAÇÃO INFANTIL NOS RESULTADOS DO ENSINO BÁSICO

A análise do impacto de ter frequentado a educação infantil na nota de português e matemática será feita em três aspectos: 1) a magnitude do valor estimado para o Efeito Tratamento nos Tratados, 2) o sinal e 3) a significância estatística. A metodologia de pareamento utilizada será a *Stratification Matching*, onde o efeito tratamento nos tratados é mensurado utilizando uma média ponderada dos blocos específicos do efeito de tratamento. Assim, tanto a estimativa do efeito tratamento nos tratados quanto do número final de blocos gerados no escore de propensão são baseados no princípio da estratificação. A título de comparação, o efeito estimado pelo método dos Mínimos Quadrados Ponderados pelo *Propensity Score* também será apresentado na mesma tabela. Na próxima seção, os resultados desse último método serão mais detalhados.

Tabela 5. Estimativa do Efeito Tratamento nos Tratados (ATT) pelo método *Stratification Matching*, e pelo método dos Mínimos Quadrados Ponderados pelo *Propensity Score*.

Estimativa	Número de Tratados	Número de controles	ATT - PSM	Erro padrão	T	ATT - MQP	t
Efeito creche na nota de português – 5º ano.	349.081	573.450	4.921	0.137	35.906	5.063	34.04
Efeito creche na nota de matemática – 5º ano.	349.081	573.450	4.989	0.141	35.279	5.735	36.99
Efeito pré-escola na nota de português – 5º ano.	373.786	588.624	15.404	0.121	127.637	11.213	83.33
Efeito pré-escola na nota de	373.786	588.624	16.451	0.125	131.801	12.497	88.88

Estimativa	Número de Tratados	Número de controles	ATT - PSM	Erro padrão	T	ATT - MQP	t
matemática – 5º ano.							
Efeito Ensino Infantil na nota de português – 5º ano.	719.218	217.627	8.159	0.123	66.086	7.176	53.51
Efeito Ensino Infantil na nota de matemática – 5º ano.	719.218	217.627	8.527	0.127	67.046	8.101	57.90
Efeito creche na nota de português – 9º ano.	496.409	1.023.064	5.405	0.110	49.063	4.094	35.84
Efeito creche na nota de matemática – 9º ano.	496.409	1.023.064	4.141	0.111	37.425	3.521	30.60
Efeito pré-escola na nota de português – 9º ano.	592.659	660.621	14.362	0.109	131.453	11.010	96.00
Efeito pré-escola na nota de matemática – 9º ano.	592.659	660.621	13.402	0.111	120.978	10.261	88.36
Efeito Ensino Infantil na nota de português – 9º ano.	974.813	260.246	9.020	0.108	83.448	6.479	57.88
Efeito Ensino Infantil na nota de matemática – 9º ano.	974.813	260.246	8.166	0.110	74.452	5.819	51.35

Fonte: Dados estimados a partir do Prova Brasil 2011 por meio do *software stata 11*.
Elaboração própria.

A tabela 5 apresenta as estimativas do efeito tratamento nos tratados (ATT) – pelo método *stratification matching*, do *Propensity Score Matching*

(PSM), e a título de comparação, também apresenta o impacto pelo método dos Mínimos Quadrados Ponderados pelo *Propensity Score* (MQP-PS) – ensino infantil nas notas de português e matemática dos alunos do 5º e 9º anos. A primeira estimativa resultou efeitos positivos e significativos de ter frequentado a creche nas notas de português e matemática dos alunos do 5º ano. Na nota de português o efeito é de 4,9 pontos e na nota de matemática, de 5 pontos⁵ para a estimativa do PSM e 5 pontos para português e 5,7 para matemática para a estimativa do MQP-PS. 5 pontos na nota, para um desvio padrão de 46,5⁶, significam que um indivíduo na mediana da distribuição, ultrapassaria 4% dos alunos em sua nota apenas por ter frequentado a creche.

Para a estimativa do efeito da pré-escola, o efeito se mostrou positivo e altamente significativo, aumentando em 15,4 pontos a nota em português e 16,4 pontos a nota em matemática para estimativa do PSM, e 11,2 e 12,5 pontos em português e matemática, respectivamente, para estimativa do MQP-PS, uma média de 14 pontos que significam passar à frente de 12% da distribuição para um aluno na mediana.

A terceira estimativa tratou o ensino infantil como um só, e inclui no grupo de tratamento todos os alunos que iniciaram o ensino antes do ensino fundamental, ou seja, tanto na creche, quanto na pré-escola. O efeito se mostrou positivo e altamente significativo. Ter frequentado o ensino infantil aumenta a nota do aluno em 8,2 pontos em português e em 8,6 pontos em matemática na estimativa do PSM. Na estimativa no MPQ-PS, esses valores passam para 7,2 e 8,1 pontos, respectivamente.

A tabela 5 também apresenta a estimativa do efeito de ter frequentado o ensino infantil na nota dos alunos do 9º ano. Alunos que iniciaram seus estudos na creche apresentaram um efeito de 5,4 pontos na nota de português e 4,2 pontos na nota de matemática, para a estimativa do PSM. Para a estimativa do MQP-PS, os efeitos foram de 4,1 e 3,5, respectivamente.

Para os alunos que iniciaram seus estudos na pré-escola, o efeito na nota de português no 9º ano foi de 14,4 pontos, e para matemática foi de 13,4 pontos para a estimativa por PSM. A estimativa por MQP-PS resultou em 11 e 10,2 pontos, respectivamente. O efeito para o ensino infantil como um todo foi

⁵ Pontos na escala do SAEB.

⁶ Média do desvio-padrão real nas notas de português e matemática de 5º e 9º anos.

de 9 pontos para português e 8,2 para matemática para estimativa por PSM. Para a estimativa por MQP-PS, o resultado foi de 6,5 e 5,8 pontos, respectivamente.

As estimativas dos efeitos estimados por Felício e Vasconcellos (2007), mostram um aumento de 6,6% na média da proficiência em matemática no 5º ano, para os que frequentaram o ensino infantil com um todo. Apenas para a creche, esse efeito foi de 8,6%, e para a pré-escola, de 6,3%. Em termos percentuais, o efeito estimado na nota do 5º ano, em matemática, para o Prova Brasil 2011 foi de 2,33% para creche, 7,86% para o pré e 4,05% para o ensino infantil. A estimativa atualizada apresentou efeitos menores, provavelmente devido à melhor especificação em relação ao trabalho anterior. Felício e Vasconcellos (2007) incluíram no pareamento variáveis que acreditavam explicar tanto a frequência – como características familiares e pessoais e condição socioeconômica – quanto a proficiência dos alunos – características relacionadas à oferta e qualidade do ensino. Neste trabalho, o pareamento foi realizado apenas com variáveis que explicam a seleção, dando atenção além da condição socioeconômica, a escolaridade dos pais, e apenas na metodologia duplamente robusta foram incluídos controles que impactam no desempenho dos alunos. Assim, a estimativa para o Prova Brasil 2011 apresentou resultados menores, mas ainda assim altos e positivos no desempenho para aqueles alunos que fizeram o ensino infantil, atualizando a evidência disponível na literatura.

Além do retorno através do melhor desempenho dos alunos (o que reflete na qualidade das escolas), os investimentos em educação, especialmente na primeira infância, apresentam altos retornos econômicos e sociais. Retornos econômicos podem ser mensurados, já os outros tipos de retornos e externalidades positivas que a educação pode proporcionar são difíceis de medir. No entanto, há diversas evidências de que esses retornos também são muito altos (GROSSMAN, 2006, in BARBOSA FILHO e PESSOA, 2009). A penúltima seção deste trabalho apresentará algumas evidências a respeito dos retornos econômicos de programas de investimento no ensino infantil, e uma análise de retorno econômico para os investimentos em educação infantil no Brasil.

5.4 ANÁLISE DOS RESULTADOS OBTIDOS PELO MÉTODO DE MÍNIMOS QUADRADOS PONDERADOS PELO *PROPENSITY SCORE* – 5º ANO

A tabela 6 apresenta os resultados estimados da regressão pelo método de Mínimos Quadrados Ponderados pelo *Propensity Score*. Os resultados estão divididos em 3: creche, pré-escola e ensino infantil, e cada divisão possui dois resultados, um para a nota em português e outro para nota em matemática como variáveis dependentes. As variáveis independentes agora não estão relacionadas à seleção, e sim ao desempenho dos alunos, e os resultados detalhados para cada variável de controle estão no apêndice deste trabalho.

De uma maneira geral, todas as variáveis resultaram altamente significantes, exceto pela variável *sempre_escola_publica*, com nível de significância de 2% e 11% (a tabela com a estimativa completa encontra-se no apêndice). De uma maneira geral, trabalhar fora, ter abandonado a escola e ser negro, apresentam um efeito negativo e altamente significativo sobre o desempenho do aluno. O aluno que lê livros frequentemente ou faz os deveres de português e matemática também estão relacionados a notas maiores. Alguns resultados como o aluno que sempre estudou em escola pública apresentar um sinal positivo com relação ao seu desempenho, e o aluno que frequenta bibliotecas sempre ou quase sempre apresentar um sinal negativo, pode ser uma evidência da má qualidade dos dados, ou um problema de multicolinearidade, mas que, de maneira geral não compromete o exercício. Dessa forma, os únicos resultados que se alteraram para as estimativas com nota em português e nota em matemática – para as notas em matemática do 5º ano, ler livros frequentemente apresentam relação negativa com o desempenho, e para as notas do 9º ano em matemática, além de ler livros, fazer o dever de português também apresenta um impacto negativo na nota em matemática – não representam um problema para o exercício. Na tabela 6 estão relacionados os efeitos para as variáveis de interesse.

Tabela 6. Regressão Mínimos Quadrados Ponderados pelo *Propensity Score*. Estimativa 1: Creche - 5º ano – Português.

Estimativa	Nome da variável de Interesse	Coef.	Erro padrão	T	P> t	Intervalo de confiança de 95%
------------	-------------------------------	-------	-------------	---	------	-------------------------------

Estimativa	Nome da variável de Interesse	Coef.	Erro padrão	T	P> t	Intervalo de confiança de 95%	
Estimativa 1 – Creche – nota em português – 5º ano.	Inicio_creche	5.06	0.148	34.04	0.000	4.77	5.35
Estimativa 1 – Creche – nota em matemática – 5º ano.	Inicio_creche	5.73	0.217	-56.92	0.000	-12.82	-11.97
Estimativa 2 – Pré-escola – nota em português – 5º ano.	Inicio_pre	11.21	0.134	83.33	0.000	10.950	11.477
Estimativa 2 – Pré-escola – nota em matemática – 5º ano.	Inicio_pre	12.49	0.140	88.88	0.000	12.22	12.77
Estimativa 3 – Ensino Infantil – nota em português – 5º ano.	Inicio_prim_inf	7.17	0.134	53.51	0.000	6.91	7.43
Estimativa 3 – Ensino Infantil – nota em matemática – 5º ano.	Inicio_prim_inf	8.10	0.139	57.90	0.000	7.82	8.37
Estimativa 1 – Creche – nota em português – 9º ano.	Inicio_creche	4.09	0.114	35.84	0.000	3.87	4.31
Estimativa 1 – Creche – nota em matemática – 9º ano.	Inicio_creche	3.52	0.115	30.60	0.000	3.29	3.74
Estimativa 2 – Pré-escola – nota em português – 9º ano.	Inicio_pre	11.01	0.114	96.00	0.000	10.78	11.23
Estimativa 2 – Pré-escola – nota em matemática – 9º ano.	Inicio_pre	10.26	0.116	88.36	0.000	10.03	10.48
Estimativa 3 – Ensino Infantil – nota em português – 9º ano.	Inicio_prim_inf	6.47	0.111	57.88	0.000	6.26	6.69
Estimativa 3 – Ensino Infantil – nota em matemática – 9º ano.	Inicio_prim_inf	5.81	0.113	51.35	0.000	5.59	6.04

Fonte: Dados estimados a partir do Prova Brasil 2011 por meio do *software stata 11*. Elaboração própria.

Na tabela 6, todos os efeitos resultaram positivos e altamente significativos para as variáveis de interesse. A primeira estimativa referente às notas do 5º ano estimou o efeito da creche sobre a nota dos alunos. Os efeitos encontrados foram de 5,06 e 5,73 pontos em português e matemática, respectivamente, o que indica que em média, um aluno na mediana da distribuição, passa na frente de 5% dos alunos só por ter feito a creche. Para a segunda estimativa, da pré-escola, o efeito é ainda maior: 11,21 e 12,49 pontos para português e matemática, respectivamente. Isso significa que só por feito a pré-escola, em média, um aluno na mediana da distribuição, passa a frente de 10% dos alunos. Para a terceira estimativa, com o ensino infantil como um todo, esse impacto é de 7,17 e 8,10 pontos, em português e matemática, respectivamente.

A tabela 6 também apresenta as estimativas para os alunos do 9º ano. Para a primeira estimativa, o impacto de ter feito a creche resultou em 4,09 e 3,52 pontos em português e matemática, respectivamente. Para a pré-escola, o efeito encontrado foi de 11,01 e 10,26 pontos em português e matemática, respectivamente. Para o ensino infantil com um todo, o impacto nas notas é de 6,47 e 5,81 pontos, em português e matemática, respectivamente. Sendo esse um método duplamente robusto, a evidência é de que se verifica empiricamente um impacto positivo no desempenho dos alunos que entram na escola no ensino infantil.

6. RETORNO ECONÔMICO

Diversos estudos recentes sobre a avaliação de programas de educação infantil têm mostrado altos retornos, e efeitos importantes que se mantêm da infância até a vida adulta. (Bowman et al., 2000). Schweinhart et al. (2005) apresentam um estudo em que comparam 123 crianças de famílias pobres, todas entre 3 e 4 anos de idade. Essas crianças foram distribuídas em dois grupos, um com um programa de educação infantil de alta qualidade, e o outro em nenhum programa, e foram acompanhadas até os 40 anos de idade. O estudo apontou que o programa de educação infantil teve efeitos positivos e significativos sobre a conclusão do ensino médio, a renda e o emprego na vida adulta, entre outros. O retorno econômico deste programa para a sociedade foi de US\$17,07 por dólar investido, a preços de 2000.

Campbell et al. (2002) apresentam um estudo em que 111 crianças de 4 anos foram aleatoriamente distribuídas em dois grupos. O primeiro era formado por crianças que participaram de um programa especial, e o outro de um programa de educação infantil típico. O grupo do programa especial resultou melhor desempenho no ensino fundamental, menor índice de repetência e gravidez na adolescência, além desse grupo ter tido maior frequência na conclusão do ensino médio e do ensino superior. O programa rendeu para a sociedade US\$3,78 por dólar investido, a preços de 2000.

Reynolds et al. (2001) apresentaram outro estudo comparando 989 crianças pobres que tinham frequentado um programa oferecido a crianças de 3 e 4 anos de idade em tempo parcial com um grupo de 550 crianças que não frequentaram. O grupo tratamento apresentou melhor desempenho educacional e comportamento social, menor repetência, menor índice de marginalidade, além de maior taxa de conclusão escolar. O retorno para a sociedade foi de US\$7,10 por dólar investido, a preços de 2000.

Um ponto importante a ser ressaltado é que grande parte dos estudos mais recentes sobre educação infantil não utilizam dados para o Brasil, e são voltados para os países desenvolvidos. Neste sentido, esses estudos já não estão em um nível de demonstrar os efeitos do ensino infantil apenas, mas também relacionar a qualidade das metodologias de ensino utilizadas para obter maiores retornos. Como já citado anteriormente, o Brasil em primeira

instância carece de investimentos em educação infantil, para que depois a eficiência das diversas metodologias seja avaliada.

Os resultados apresentados na tabela 5 mostram um impacto positivo e significativo do ensino infantil sobre as notas de matemática e português no 5º e 9º anos. Para o cálculo do retorno econômico, será utilizado um valor médio do impacto estimado para o ensino infantil como um todo, de 8,46 pontos. Murnane et al (2000) mostra através de um estudo com dados em painel para os EUA que a relação positiva entre desempenho em testes padronizados e retorno monetário existe. Curi e Menezes-Filho (2007) apresentam essa evidência para o Brasil, afirmando que o desempenho em testes padronizados impacta os salários futuros com elasticidade de 0,3. Assim, essas duas estimativas serão levadas em consideração para o cálculo de retorno econômico. A seguir, algumas suposições:

- a) O impacto médio positivo de 14,90 pontos na nota média do aluno do 5º e 9º anos estimado para a pré-escola se mantém em valores absolutos para todos os alunos que iniciaram seus estudos em um ensino infantil de 3 anos de duração (4 a 6 anos de idade).
- b) Seguindo a estimativa de Curi e Menezes-Filho (2007), a elasticidade do desempenho nos salários é de 0,3.
- c) A nota média dos alunos em português e matemática, e para o 5º e 9º ano é de 221,22 pontos.
- d) Seguindo dados do IBGE (2010), o salário médio no Brasil está estimado em R\$1650,30.
- e) Seguindo dados da previdência, a suposição de tempo de trabalho será de 35 anos, que é o tempo mínimo de contribuição que um homem precisa para se aposentar. Assim, o cálculo levará em conta os ganhos obtidos dos 18 aos 52 anos de idade.
- f) O investimento por estudante, segundo dados do INEP/MEC para a educação infantil é de R\$5.885,00 (valores atualizados para 2010 pelo IPCA). Esse custo, que é distribuído dos 4 aos 6 anos de idade da criança, representam um gasto de R\$1961,66 anuais.
- g) A taxa anual de desconto utilizada será de 5%.

Assim, calculando a diferença salarial oriunda desse aumento na nota, tem-se que o salário médio anual seria acrescido de R\$433,55⁷. Trazendo os valores referentes ao tempo de serviço dos 18 aos 52 anos (35 anos de trabalho), tem-se um valor de R\$7.099,03 de receita líquida por aluno. Trazendo os valores referentes aos custos de 3 anos do ensino infantil para o valor presente, tem-se um valor de R\$5.342,09. Ou seja, a cada real investido no ensino infantil, o retorno é de R\$1,32. A TIR (Taxa Interna de Retorno) calculada é de 6% ao ano.

Além dos efeitos sobre a renda e o crescimento econômico, todos sustentados por evidências empíricas, em sua maioria, robustas, a educação possui muitos outros efeitos positivos sociais. O aumento da educação de um grupo que interage com determinado indivíduo, por exemplo, pode aumentar também a sua produtividade. E ainda assim, há outros motivos que justificam os investimentos em educação, alguns deles não especificamente voltados ao mercado de trabalho, mas para outras atividades que o indivíduo possa exercer. Deste ponto de vista, os retornos estão subestimados, pois não levam em conta os retornos sociais da educação, que provavelmente ultrapassam os retornos privados. (GROSSMAN, 2006, in BARBOSA FILHO e PESSOA, 2009).

⁷ Referentes a 13 salários anuais.

7. CONCLUSÃO

O ensino na primeira infância no Brasil, compreendido de duas fases, a creche, de 0 a 3 anos, e a pré-escola, de 4 a 5 anos, tem recebido investimentos muito aquém do que seria esperado, dada a proporção de evidências nesse campo, que mostram os altos retornos que essa fase produz. Dados do IBGE mostram que pouco mais de 23% das crianças de 0 a 3 anos, frequentavam a creche em 2010, e 80% das crianças de 4 e 5 anos frequentavam a pré-escola. Já a cobertura para o ensino básico no Brasil era de 97% das crianças entre 6 e 14 anos em 2010. Mesmo que crescentes, os investimentos em educação infantil ainda são muito baixos. Investimentos na primeira infância podem ser o canal pelo qual a qualidade da educação, medida tradicionalmente pelo desempenho em testes padronizados, aumentaria.

Este trabalho estimou o impacto que a primeira infância produz no desempenho dos alunos no ensino básico de 5^o e 9^o anos, em português e matemática, levando em consideração um possível problema de seleção por observáveis que pode ocorrer na escolha dos pais colocarem seus filhos no ensino infantil, ou diretamente no ensino fundamental. A primeira metodologia, que diz respeito à consideração deste problema, e que, portanto, foi utilizada nesse trabalho, foi o *Propensity Score Matching (PSM)*, um método de pareamento que cria grupos comparáveis entre si. Assim, os grupos criados teriam alunos iguais em relação às suas características familiares, condição socioeconômica e escolaridade dos pais (variáveis de controle), mas fariam distinção quanto à participação no tratamento ou não, neste caso, ter frequentado o ensino infantil ou não.

Para controlar também para as variáveis que influenciam no desempenho do aluno, uma segunda metodologia foi utilizada, a saber, Mínimos Quadrados Ponderados pelo *Propensity Score (MQP-PS)*. Esse método estima uma regressão em que a variável dependente é o desempenho do aluno, e as independentes são, além da variável de interesse, ter frequentado a educação infantil, outras variáveis que podem influenciar no desempenho do aluno, como o fato de trabalhar fora, ter reprovado, ser negro, ou sempre ter estudado em escola pública. Esse método é duplamente

robusto, pois, além disso, pondera a regressão pelos escores de propensão, que são a probabilidade de o aluno ter iniciado seus estudos no ensino infantil, dadas variáveis que controlam a seleção, a saber, a condição socioeconômica dos pais e a sua escolaridade.

Os resultados estimados mostram que o impacto de ter frequentado o ensino infantil na nota do aluno do 5º ano, encontrou um efeito de 8,2 pontos na nota de português, e um efeito de 8,6 na nota de matemática para a estimativa do PSM, e 7,2 e 8,1, na nota de português e matemática, respectivamente para a estimativa do MQP. Para a pré-escola os efeitos são ainda maiores, 15,4 e 16,5 pontos para a estimativa do PSM e 11,23 e 12,5 para estimativa do MQP, respectivamente. O impacto para dados do 9º ano também resultaram positivos e altamente significativos para a pré-escola: 14,4 e 13,4 pontos para português e matemática, respectivamente, sob a metodologia do PSM, e 11,0 e 10,2, respectivamente, para a metodologia do MQP. Para o ensino infantil como um todo, o impacto é menor: 9 e 8,2 pontos, respectivamente, para a estimativa do PSM, e 6,5 e 5,9, respectivamente, para a estimativa do MQP. Um efeito médio de 10,4 pontos na nota na escala do SAEB significa que um indivíduo na mediana, ultrapassaria 9% dos alunos na distribuição das notas apenas por ter frequentado o ensino infantil.

As estimativas de Felício e Vasconcellos (2007) mostram um aumento de 8,6% nota para os alunos que frequentaram a creche, e 6,3% para os que frequentaram a pré-escola. Para o ensino infantil como um todo, eles estimaram um efeito de 6,6% na nota média. Para efeito de comparação, em termos percentuais, o efeito estimado na nota do 5º ano, em matemática, para o Prova Brasil 2011 foi de 2,33% para creche, 7,86% para o pré e 4,05% para o ensino infantil. A estimativa atualizada apresentou efeitos menores, provavelmente devido à melhor especificação em relação ao trabalho anterior, já que Felício e Vasconcellos (2007) incluíram no pareamento, variáveis que acreditavam explicar tanto a frequência quanto a proficiência dos alunos. Neste trabalho, o pareamento foi realizado apenas com variáveis que explicam a seleção, dando atenção além da condição socioeconômica, a escolaridade dos pais, e apenas na metodologia duplamente robusta foram incluídos controles que impactam no desempenho dos alunos. Apesar das estimativas do Prova Brasil 2011 apresentarem resultados menores, os efeitos ainda assim altos e

positivos no desempenho para aqueles alunos que fizeram o ensino infantil, atualizando a evidência disponível na literatura.

No entanto, ainda que a estimativa tenha mostrado efeitos altamente significativos, um dos limites deste trabalho, e que pode ser sugerido para pesquisas futuras, é que os métodos usados não levam em consideração uma possível seleção por não-observáveis, ou seja, variáveis que apresentam maior dificuldade de ser medidas, já que não podem ser observadas, mas que também podem ter alguma influência na escolha de o pai colocar ou não seu filho no ensino infantil, como por exemplo, a proximidade ou ausência de creches perto das famílias.

Por último, uma análise de retorno econômico mostrou que, mesmo sendo uma projeção subestimada, já que os retornos sociais da educação são muito altos, cada real investido em educação resulta em R\$1,32 (a preços de 2010). Em se tratando de investimentos educacionais, medir o retorno da educação pode ser problemático, já que além dos retornos privados e econômicos, inúmeras externalidades positivas podem ser geradas em investimentos em educação. No entanto, existem diversas evidências dos altos retornos que a educação infantil pode proporcionar à sociedade, e as políticas públicas necessitam de uma reorientação em direção à maior eficiência de seus investimentos em educação.

REFERÊNCIAS

Barbosa Filho, F. H e Pessoa, S. **Educação, Crescimento e Distribuição de Renda: A Experiência Brasileira em Perspectiva Histórica**. In: VELOSO, Fernando; PESSOA, Samuel; HENRIQUES, Ricardo e GIAMBIAGI, Fábio. (Orgs.). **Educação Básica no Brasil**. São Paulo: Elsevier Editora Ltda, 2009.

Becker, G.S., Tomes, N. **Human capital and the rise and fall of families**. Journal of Labor Economics 4 (3, Part 2), S1–S39, 1986.

Becker S.O., Ichino, A. **Estimation of average treatment effects based on propensity score**. Stata Journal, v.2, n.4, p.358-377, 2002.

Behrman, Jere R., **Measuring the effectiveness of schooling policies in developing countries: revisiting issues of methodology**. Economics of Education Review, v. 15, n. 4, p. 345-364, 1996.

Berlinski, S., Galiani, S., Gertler, P., **The effect of pre-primary education on primary school performance**, 2006.

Bowman BT, Donovan MS, Burns MS, **Eager to learn: Educating our preschoolers**. **Committee on Early Childhood Pedagogy**, Commission on Behavioural and Social Sciences and Education, National Research Council, National Academy Press, 2000.

Brito, Ariana M., **Salários de professores e qualidade da educação no Brasil** Dissertação de mestrado UFF, 2012

Cameron, A. C., Trivedi, P. K. **Microeconometrics: methods and applications**, Cambridge University Press, 2005.

Campbell FA, Ramey CT, Pungello E, Sparling J, Miller-Johnson S. **Early childhood education: Young adult outcomes from the Abecedarian Project**. Applied Developmental Science, 2002.

Campos, M.M. **Educação infantil: o debate e a pesquisa**. Cadernos de Pesquisa, São Paulo, n. 101, p. 113-127, jul. 1997.

Carneiro, P., Heckman, J.J. **Human capital policy**. In: Heckman, J.J., Krueger, A.B., Friedman, 2003.

Carneiro, P., Heckman, J.J. **The evidence on credit constraints in post-secondary schooling**. Economic Journal 112 (482), 705–734, 2002.

Cerqueira, D. R. C.; Carvalho, A.; Lobão, W. J. A.; RODRIGUES, Rute I. **Análise dos custos e conseqüências da violência no Brasil**. Texto para Discussão no. 1284. Brasília: IPEA, 2007

Cunha, F., Heckmanm J., Lochner, L., Masterov, D., Chapter 12 **Interpreting the evidence on life cycle skill formation** Handbook of the Economics of Education. (pp. 695 a 812). Elsevier, 2006.

Curi, A., Menezes-Filho, N., **Os efeitos da pré-escola sobre os salários, a escolaridade e a proficiência escolar**. Anais da ANPEC, 2006.

Curi, A. Z., Menezes-Filho, N. **The relationship between school performance and future wages in Brazil**. Universidade de São Paulo, 2007.

Feinstein, L. Robertson, D. e Symons, J. **Pre-school Education and Attainment in the NCDS and BCS**. CEP Discussion Paper, 1998.

Felício, F., Vasconcellos, L. **O Efeito da Educação Infantil sobre o Desempenho Escolar Medido em Exames Padronizados**. Anais do XXXV Encontro da Anpec, Recife, Pernambuco, 2007.

Galor, O. and Zeira, J. **Income Distribution and Macroeconomics**, Review of Economic Studies, 1993.

Galor, O. and Zeira, J. **Income Distribution and Investment in Human Capital: Macroeconomics Implication**. Working Paper No. 197, Department of Economics, Hebrew University, 1998.

Gertler, P. J. and Fernald, L.C. **The Medium Term Impact of Oportunidades on Child Development in Rural Areas**. Unpublished manuscript, 2004

Graces, E., Thomas, D. and Currie, J. **Longer-Term Effects of Head Start**. The American Economic Review 92(4): 999 – 1012, 2002.

Guimarães, J. L.; Pinto J. M. R. **A Demanda pela educação infantil e os recursos disponíveis para seu financiamento**. Em aberto, v. 18, n. 74, p. 12-105, jul, 2001.

Heckman, J.J. **Policies to foster human capital**. Research in Economics 54 (1), 3–56. With discussion, 2000.

Heckman, J., Navarro-Lozano. S. **Using matching, instrumental variables and control functions to estimate economic choice models**. NBER Working Paper 9497. Fev 2003.

Heckman, J.J., Stixrud, J., Urzua, S. **The effects of cognitive and noncognitive abilities on labor market outcomes and social behavior**. Journal of Labor Economics 24 (3), p. 411–482, 2006.

Heckman, James. **Skill Formation and the Economics of Investing in Disadvantaged Children**. Science 30, 2006b.

Heckman, James. Masterov, Dimitriy V. **Argument for Investing in Young Children**, National Bureau of Economic Research, 2007.

Imbens, G. **Nonparametric estimation of average treatment effects under exogeneity: a review**. The Review of Economics and Statistics, Fevereiro, 86(1), 2004.

Imbens, G. M.; Wooldridge, J. M. **Recent Development in the Econometrics Program Evaluation**. NBER Technical Working Paper, n. 14251, 2008.

IPEA, **Apresentação da Prova Brasil**. Acesso em 15 de Novembro de 2012: http://portal.mec.gov.br/index.php?Itemid=324&id=210&option=com_content&view=article

Knudsen, E.I. **Sensitive periods in the development of the brain and behavior** Journal of Cognitive Neuroscience 16 (1), 1412–1425, 2004.

Menezes-Filho, N. A. **A evolução da educação no Brasil e seu impacto no mercado de trabalho**. Departamento de Economia. Universidade de São Paulo, 2001.

Murnane, R. J.; Willett, J. B.; Duhaldeborde, Y.; Tyler, J. H. **How important are the cognitive skills of teenagers in predicting subsequent earnings?** Journal of Policy Analysis and Management, vol. 19, n. 4, 2000.

Newport, E.L. **Critical periods in language development**. In: Nadel, L. (Ed.), Encyclopedia of Cognitive Science. Macmillan/Nature Publishing Group, London, pp. 737–740, 2002.

Pazello, E.T., e Almeida, R.B. **O efeito da pré-escola sobre o desempenho escolar futuro dos indivíduos**. ANPEC. Salvador, 2010.

Perotti, R. **Growth, Income Distribution, and democracy: what the Data say**. Journal of Economic Growth, 1996.

Reynolds AJ, Temple JA, Robertson DL, Mann EA. **Long-term effects of an early childhood intervention on educational achievement and juvenile arrest: A 15-year follow-up of low-income children in public schools**. JAMA - Journal of the American Medical Association, 2001.

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

Schweinhart LJ, Montie J, Xiang Z, Barnett WS, Belfield CR, Nores M. **Lifetime effects: The High/Scope Perry Preschool Study through age 40.** Ypsilanti, Mich: High/Scope Press, 2005.

Ugá, M. A. D. **Instrumentos de Avaliação Econômica dos Serviços de Saúde: Alcances e Limitações.** In: Piola SF, Viana SF, organizadores. *Economia da Saúde. Conceito e Contribuição para a Gestão da Saúde.* Rio de Janeiro: IPEA; 1995.

Veloso, F. **A Evolução Recente e Propostas para a Melhoria da Educação no Brasil,** in E. Bacha e S. Schwartzman (orgs.), *Brasil: A Nova Agenda Social.* Rio de Janeiro, LTC, pp. 215-253, 2011.

APÊNDICES

APÊNDICE 1. Modelo Probit. Estimativa 1: creche. 5º ano.

Variáveis	Coeficiente	Erro padrão	Z	P> z	Intervalo de confiança de 95%	
mae_superior_completo	0.297	0.005	56.85	0.000	0.287	0.307
Empregada	0.163	0.006	26.01	0.000	0.151	0.175
Computador	0.213	0.0039	54.58	0.000	0.206	0.221
tv_em_casa	0.233	0.0095	24.52	0.000	0.214	0.252
pai_sabe_ler	0.125	0.007	17.56	0.000	0.111	0.139
freezer_em_casa	-0.081	0.004	-18.31	0.000	-0.089	-0.072
_cons	-0.005	0.0111	0.48	0.631	-0.016	0.027

Fonte: Dados estimados a partir do Prova Brasil 2011 por meio do *software stata 11*.

Elaboração própria.

APÊNDICE 2. Limites inferiores, números de tratamentos e controles por bloco. Estimativa 1: creche. 5º ano.

Limite inferior do propensity score (com suporte mínimo)	inicio_maternal		Total
	D = 0 Controle	D=1 Tratamento	
.4	105	142	247
.5	1.535	1.738	3.273
.55	3.475	4.049	7.524
.5625	1.646	2.225	3.871
.575	8.028	11.666	19.694
.6	12.977	19.823	32.800
.625	49.010	88.230	137.240
.65	2.951	5.935	8.886
.6750001	12.638	27.848	40.486
.7	2.620	6.000	8.620
.7125	35.859	92.698	128.557
.725	160	366	526
.7375	5.899	17.889	23.788
.75	3.522	12.031	15.553
.775	3.165	12.210	15.375
.8	7.673	31.225	38.898
.825	1.244	5.934	7.178
.85	1.576	9.072	10.648

Total	154.083	349.081	503.164
--------------	----------------	----------------	----------------

Fonte: Prova Brasil 2011/Elaboração própria.

APÊNDICE 3. Modelo Probit. Estimativa 2: pré-escola. 5º ano.

Variáveis	Coeficiente	Erro padrão	z	P> z	Intervalo de confiança de 95%	
Maquina_de_lavar_em_casa	0.015	0.0038	4.11	0.000	0.008	0.023
Geladeira_em_casa	0.227	0.009	23.31	0.000	0.208	0.246
Pai_sabe_ler	0.237	0.006	36.84	0.000	0.224	0.249
Radio_em_casa	0.064	0.004	13.83	0.000	0.055	0.073
Mae_superior_completo	0.077	0.005	15.78	0.000	0.067	0.087
Tv_em_casa	0.205	0.008	24.45	0.000	0.189	0.222
_cons	-0.330	0.0127	-26.01	0.000	-0.355	-0.305

Fonte: Dados estimados a partir do Prova Brasil 2011 por meio do *software stata 11*.

Elaboração própria.

APÊNDICE 4. Limites inferiores, números de tratamentos e controles por bloco. Estimativa 2: pré-escola. 5º ano.

Limite inferior do propensity score (com suporte mínimo)	inicio_maternal		Total
	D = 0 Controle	D=1 Tratamento	
.2	530	405	935
.4	4.028	3.976	8.004
.5	3.777	4.453	8.230
.55	23.784	30.934	54.718
.6	1.390	1.915	3.305
.625	23.031	39.948	62.979
.65	34.074	64.722	98.796
.6625	91.189	180.922	272.111
.6750001	20.859	46.511	67.370
Total	202.662	373.786	576.448

Fonte: Prova Brasil 2011/Elaboração própria.

APÊNDICE 5. Modelo Probit. Estimativa 3: ensino infantil. 5º ano.

Variáveis	Coeficiente	Erro padrão	Z	P> z	Intervalo de confiança de 95%	
Empregada	0.092	0.0054	16.97	0.000	0.081	0.102
Pai_sabe_ler	0.115	0.0060	19.00	0.000	0.103	0.127
Mae_le	0.152	0.0054	28.24	0.000	0.142	0.163
Mae_superior_completo	0.191	0.0045	42.17	0.000	0.182	0.200
Carro	0.088	0.0032	26.93	0.000	0.082	0.095
Geladeira_em_casa	0.154	0.0091	16.95	0.000	0.136	0.172
Tv_em_casa	0.178	0.0078	22.65	0.000	0.163	0.194
Radio_em_casa	0.134	0.0043	3.11	0.002	0.004	0.021
Mora_ate_5	0.156	0.0037	42.09	0.000	0.149	0.164
_cons	0.138	0.0125	11.04	0.000	0.114	0.163

Fonte: Dados estimados a partir do Prova Brasil 2011 por meio do *software stata 11*.
Elaboração própria.

APÊNDICE 6. Limites inferiores, números de tratamentos e controles por bloco. Estimativa 3: ensino infantil. 5º ano.

Limite inferior do propensity score (com suporte mínimo)	inicio_maternal		Total
	D = 0 Controle	D=1 Tratamento	
.4	80	139	219
.6	572	1.037	1.609
.65	2.837	6.321	9.158
.7	2.677	6.837	9.514
.725	3.819	10.163	13.982
.7375	3.791	11.161	14.952
.75	1.367	4.047	5.414
.7625	18.827	63.287	82.114
.775	9.279	32.324	41.603
.7875	3.104	12.134	15.238
.8	12.569	50.005	62.574
.8125	38.847	176.922	215.769
.825	3.056	14.190	17.246
.8375	37.982	200.194	238.176
.85	9.308	57.958	67.266
.875	7.058	53.153	60.211
.8875	221	2.203	2.424

.9	1.716	17.143	18.859
Total	157.110	719.218	876.328

Fonte: Prova Brasil 2011/Elaboração própria.

APÊNDICE 7. Modelo Probit. Estimativa 1: creche. 9º ano.

Variáveis	Coeficiente	Erro padrão	z	P> z	Intervalo de confiança de 95%	
Mora_ate_5	0.192	0.0037	51.63	0.000	0.185	0.200
Mae_sabe_ler	0.374	0.0057	64.74	0.000	0.363	0.385
Carro	0.047	0.0029	15.94	0.000	0.041	0.052
Empregada	0.205	0.0049	41.59	0.000	0.195	0.215
Mae_superior_completo	0.499	0.0054	90.87	0.000	0.488	0.510
_cons	-0.302	0.0061	-49.35	0.000	-0.314	-0.290

Fonte: Dados estimados a partir do Prova Brasil 2011 por meio do *software stata 11*.
Elaboração própria.

APÊNDICE 8. Limites inferiores, números de tratamentos e controles por bloco. Estimativa 1: creche. 9º ano.

Limite inferior do propensity score (com suporte mínimo)	inicio_maternal		Total
	D = 0 Controle	D=1 Tratamento	
.2	8.503	5.668	14.171
.4	19.244	16.048	35.292
.525	34.287	36.805	71.092
.5375	16.829	21.602	38.431
.55	329	376	705
.6	115.834	179.661	295.495
.6125	85.999	138.299	224.298
.625	1.669	3.169	4.838
.65	69	91	160
.6750001	6.396	12.644	19.040
.6875001	8.789	22.080	30.869
.7	836	1.892	2.728
.725	879	2.511	3.390
.775	4.145	14.489	18.634
.7875	6.603	25.985	32.588
.8	2.960	15.089	18.049

Total	313.371	496.409	809.780
--------------	----------------	----------------	----------------

Fonte: Prova Brasil 2011/Elaboração própria.

APÊNDICE 9. Modelo Probit. Estimativa 2: pré-escola. 9º ano.

Variáveis	Coeficiente	Erro padrão	z	P> z	Intervalo de confiança de 95%	
Radio_em_casa	0.102	0.0043	23.36	0.000	0.093	0.110
Freezer_em_casa	-0.099	0.0034	-28.76	0.000	-0.106	-0.092
Tv_em_casa	0.352	0.0092	37.94	0.000	0.334	0.370
Mae_superior_completo	0.339	0.0058	57.84	0.000	0.327	0.350
Pai_sabe_ler	0.338	0.0047	70.66	0.000	0.328	0.347
_cons	-0.227	0.0104	-21.79	0.000	-0.247	-0.206

Fonte: Dados estimados a partir do Prova Brasil 2011 por meio do *software stata 11*.

Elaboração própria.

APÊNDICE 10. Limites inferiores, números de tratamentos e controles por bloco. Estimativa 2: pré-escola. 9º ano.

Limite inferior do propensity score (com suporte mínimo)	inicio_maternal		Total
	D = 0 Controle	D=1 Tratamento	
.2	54	34	88
.4	1.845	1.554	3.399
.5	7.445	9.388	16.833
.55	4.539	5.613	10.152
.575	4.065	5.293	9.358
.5875	21.313	30.119	51.432
.6	4.309	7.465	11.774
.65	67.129	141.594	208.723
.7	136.415	341.771	478.186
.75	5.066	18.538	23.604
.8	6.940	31.290	38.230
Total	259.120	592.659	851.779

Fonte: Prova Brasil 2011/Elaboração própria.

APÊNDICE 11. Modelo Probit. Estimativa 3: Ensino infantil. 9º ano.

Variáveis	Coeficiente	Erro padrão	Z	P> z	Intervalo de confiança de 95%	
Pai_superior_completo	0.325	0.0056	57.16	0.000	0.314	0.336

Empregada	0.170	0.0047	36.12	0.000	0.160	0.179
Tv_em_casa	0.146	0.0090	16.20	0.000	0.128	0.164
K7_dvd_em_casa	0.264	0.0044	59.63	0.000	0.256	0.273
Geladeira_em_casa	0.332	0.0084	39.43	0.000	0.315	0.348
_cons	0.124	0.0104	11.95	0.000	0.104	0.145

Fonte: Dados estimados a partir do Prova Brasil 2011 por meio do *software stata 11*.
Elaboração própria.

APÊNDICE 12. Limites inferiores, números de tratamentos e controles por bloco. Estimativa 2: Ensino infantil. 9º ano.

Limite inferior do propensity score (com suporte mínimo)	inicio_maternal		Total
	D = 0 Controle	D=1 Tratamento	
.4	2.108	2.816	4.924
.6	2.540	4.410	6.950
.65	3.762	8.224	11.986
.7	3.906	8.302	12.208
.725	23.026	59.800	82.826
.75	1.737	4.676	6.413
.775	1.225	4.155	5.380
.8	171.842	723.571	895.413
.8125	612	3.020	3.632
.825	14	100	114
.85	13.757	78.643	92.400
.875	7.559	57.148	64.707
.9	1.897	19.948	21.845
Total	233.985	974.813	1.208.798

Fonte: Prova Brasil 2011/Elaboração própria.

APÊNDICE 13. Regressão Mínimos Quadrados Ponderados pelo Propensity Score. Estimativa 1: Creche - 5º ano – Português.

Variáveis	Coefficiente	Erro padrão	t	P> t	Intervalo de confiança de 95%	
Inicio_creche	5.06	0.148	34.04	0.000	4.77	5.35
Trab_fora	-18.76	0.203	-92.34	0.000	-19.15	-18.36
Sempre_escola_publica	2.21	0.166	13.30	0.000	1.88	2.53
Abandonou_escola	-6.99	0.304	-22.97	0.000	-7.59	-6.40

Reprovou	-26.39	0.151	-174.67	0.000	-26.68	-26.09
Negro	-9.03	0.224	-40.21	0.000	-9.47	-8.59
Freq._biblioteca	-7.30	0.158	-46.11	0.000	-7.61	-6.99
Le_livros_sempre	3.07	0.142	21.53	0.000	2.79	3.35
Faz_dever_port_sempre	11.71	0.187	62.41	0.000	11.34	12.08
Faz_dever_mat_sempre	4.46	0.195	22.81	0.000	4.08	4.84
_cons	190.76	0.252	755.99	0.000	190.26	191.25

Fonte: Dados estimados a partir do Prova Brasil 2011 por meio do *software stata 11*.
Elaboração própria.

APÊNDICE 14. Regressão Mínimos Quadrados Ponderados pelo *Propensity Score*. Estimativa 1: Creche - 5º ano – Matemática.

Variáveis	Coefficiente	Erro padrão	T	P> t	Intervalo de confiança de 95%	
Inicio_creche	5.73	0.155	36.99	0.000	5.43	6.03
Trab_fora	-12;40	0.217	-56.92	0.000	-12.82	-11.97
Sempre_escola_publica	3.86	0.171	22.60	0.000	3.53	4.20
Abandonou_escola	-7.22	0.312	-23.14	0.000	-7.83	-6.61
Reprovou	-26.16	0.157	-165.96	0.000	-26.47	-25.85
Negro	-10.57	0.230	-45.94	0.000	-11.03	-10.12
Freq._biblioteca	-7.50	0.164	-45.64	0.000	-7.83	-7.18
Le_livros_sempre	-0.62	0.148	-4.22	0.000	-0.91	-0.33
Faz_dever_port_sempre	4.90	0.198	24.67	0.000	4.51	5.29
Faz_dever_mat_sempre	12.95	0.201	64.36	0.000	12.56	13.35
_cons	208.37	0.260	799.09	0.000	207.86	208.88

Fonte: Dados estimados a partir do Prova Brasil 2011 por meio do *software stata 11*.
Elaboração própria.

APÊNDICE 15. Regressão Mínimos Quadrados Ponderados pelo *Propensity Score*. Estimativa 2: Pré – 5º ano. Português.

Variáveis	Coefficiente	Erro padrão	T	P> t	Intervalo de confiança de 95%	
Inicio_pre	11.21	0.134	83.33	0.000	10.950	11.477
Trab_fora	-18.20	0.193	-93.97	0.000	-18.589	-17.829
Sempre_escola_publica	6.99	0.163	42.66	0.000	6.669	7.311
Abandonou_escola	-5.15	0.285	-18.06	0.000	-5.712	-4.594
Reprovou	-27.20	0.143	-190.18	0.000	-27.481	-26.920

Negro	-9.79	0.217	-45.11	0.000	-10.224	-9.373
Freq._biblioteca	-6.40	0.149	-42.94	0.000	-6.697	-6.112
Le_livros_sempre	2.51	0.132	19.06	0.000	2.260	2.778
Faz_dever_port_sempre	12.43	0.175	70.71	0.000	12.094	12.784
Faz_dever_mat_sempre	5.21	0.184	28.34	0.000	4.855	5.576
_cons	183.54	0.233	785.15	0.000	183.087	184.003

Fonte: Dados estimados a partir do Prova Brasil 2011 por meio do *software stata 11*.
Elaboração própria.

APÊNDICE 16. Regressão Mínimos Quadrados Ponderados pelo Propensity Score. Estimativa 2: Pré – 5º ano. Matemática.

Variáveis	Coeficiente	Erro padrão	T	P> t	Intervalo de confiança de 95%	
Inicio_pre	12.49	0.140	88.88	0.000	12.22	12.77
Trab_fora	-11.17	0.208	-53.55	0.000	-11.58	-10.76
Sempre_escola_publica	8.61	0.168	51.00	0.000	8.28	8.94
Abandonou_escola	-5.45	0.295	-18.47	0.000	-6.03	-4.87
Reprovou	-26.95	0.149	-180.29	0.000	-27.24	-26.66
Negro	-11.75	0.223	-52.57	0.000	-12.19	-11.31
Freq._biblioteca	-6.24	0.155	-40.30	0.000	-6.55	-5.94
Le_livros_sempre	-0.94	0.137	-6.91	0.000	-1.21	-0.679
Faz_dever_port_sempre	5.94	0.186	31.87	0.000	5.57	6.30
Faz_dever_mat_sempre	13.25	0.189	69.99	0.000	12.88	13.62
_cons	200.99	0.242	828.36	0.000	200.51	201.47

Fonte: Dados estimados a partir do Prova Brasil 2011 por meio do *software stata 11*.
Elaboração própria.

APÊNDICE 17. Regressão Mínimos Quadrados Ponderados pelo Propensity Score. Estimativa 3: Ensino Infantil – 5º ano. Português.

Variáveis	Coeficiente	Erro padrão	T	P> t	Intervalo de confiança de 95%	
Inicio_prim_inf	7.17	0.134	53.51	0.000	6.91	7.43
Trab_fora	-19.22	0.159	-120.88	0.000	-19.53	-18.91
Sempre_escola_publica	4.14	0.128	32.21	0.000	3.89	4.39
Abandonou_escola	-6.83	0.240	-28.47	0.000	-7.31	-6.36
Reprovou	-28.34	0.117	-242.18	0.000	-28.56	-28.11

Negro	-9.97	0.175	-56.72	0.000	-10.31	-9.63
Freq._biblioteca	-6.83	0.120	-56.56	0.000	-7.06	-6.59
Le_livros_sempre	2.86	0.107	26.64	0.000	2.65	3.07
Faz_dever_port_sempre	12.43	0.144	86.03	0.000	12.15	12.71
Faz_dever_mat_sempre	5.06	0.151	33.53	0.000	4.76	5.36
_cons	188.44	0.208	904.87	0.000	188.03	188.85

Fonte: Dados estimados a partir do Prova Brasil 2011 por meio do *software stata 11*.
Elaboração própria.

APÊNDICE 18. Regressão Mínimos Quadrados Ponderados pelo Propensity Score. Estimativa 3: Ensino Infantil – 5º ano. Matemática.

Variáveis	Coefficiente	Erro padrão	T	P> t	Intervalo de confiança de 95%	
Inicio_prim_inf	8.10	0.139	57.90	0.000	7.82	8.37
Trab_fora	-12.40	0.170	-72.72	0.000	-12.73	-12.06
Sempre_escola_publica	6.02	0.131	45.63	0.000	5.76	6.28
Abandonou_escola	-7.17	0.247	-28.99	0.000	-7.66	-6.69
Reprovou	-28.01	0.122	-229.37	0.000	-28.24	-27.77
Negro	-11.71	0.180	-64.99	0.000	-12.06	-11.35
Freq._biblioteca	-6.87	0.125	-54.81	0.000	-7.11	-6.62
Le_livros_sempre	-0.77	0.111	-6.90	0.000	-0.989	-0.55
Faz_dever_port_sempre	5.64	0.152	36.92	0.000	5.34	5.94
Faz_dever_mat_sempre	13.52	0.155	87.20	0.000	13.22	13.83
_cons	205.76	0.215	954.95	0.000	205.34	206.18

Fonte: Dados estimados a partir do Prova Brasil 2011 por meio do *software stata 11*.
Elaboração própria.

APÊNDICE 19. Regressão Mínimos Quadrados Ponderados pelo Propensity Score. Estimativa 1: Creche – 9º ano. Português.

Variáveis	Coefficiente	Erro padrão	T	P> t	Intervalo de confiança de 95%	
Inicio_creche	4.09	0.114	35.84	0.000	3.87	4.31
Trab_fora	-11.61	0.143	-81.12	0.000	-11.89	-11.33
Sempre_escola_publica	0.35	0.184	1.91	0.057	-0.01	0.71
Abandonou_escola	-1.97	0.252	-7.85	0.000	-2.47	-1.48
Reprovou	-22.35	0.119	-186.57	0.000	-22.58	-22.11

Negro	-6.65	0.174	-38.25	0.000	-6.99	-6.31
Freq._biblioteca	-5.61	0.173	-32.28	0.000	-5.95	-5.27
Le_livros_sempre	8.53	0.129	65.86	0.000	8.27	8.78
Faz_dever_port_sempre	4.95	0.126	39.25	0.000	4.70	5.20
Faz_dever_mat_sempre	1.69	0.123	13.69	0.000	1.44	1.93
_cons	243.24	0.222	1094.80	0.000	242.81	243.68

Fonte: Dados estimados a partir do Prova Brasil 2011 por meio do *software stata 11*.
Elaboração própria.

APÊNDICE 20. Regressão Mínimos Quadrados Ponderados pelo *Propensity Score*. Estimativa 1: Creche – 9º ano. Matemática.

Variáveis	Coefficiente	Erro padrão	T	P> t	Intervalo de confiança de 95%	
Inicio_creche	3.52	0.115	30.60	0.000	3.29	3.74
Trab_fora	-3.76	0.143	-26.23	0.000	-4.04	-3.48
Sempre_escola_publica	0.30	0.187	1.62	0.106	-0.06	0.67
Abandonou_escola	-4.68	0.248	-18.84	0.000	-5.17	-4.20
Reprovou	-21.71	0.119	-181.26	0.000	-21.94	-21.47
Negro	-7.01	0.173	-40.49	0.000	-7.35	-6.67
Freq._biblioteca	-6.77	0.175	-38.68	0.000	-7.11	-6.43
Le_livros_sempre	1.44	0.129	11.11	0.000	1.18	1.69
Faz_dever_port_sempre	-3.08	0.125	-24.54	0.000	-3.33	-2.84
Faz_dever_mat_sempre	14.39	0.122	117.19	0.000	14.15	14.63
_cons	248.52	0.224	1106.28	0.000	248.08	248.96

Fonte: Dados estimados a partir do Prova Brasil 2011 por meio do *software stata 11*.
Elaboração própria.

APÊNDICE 21. Regressão Mínimos Quadrados Ponderados pelo *Propensity Score*. Estimativa 2: Pré – 9º ano. Português.

Variáveis	Coefficiente	Erro padrão	t	P> t	Intervalo de confiança de 95%	
Inicio_pre	11.01	0.114	96.00	0.000	10.78	11.23
Trab_fora	-10.60	0.139	-76.04	0.000	-10.87	-10.33
Sempre_escola_publica	2.63	0.199	13.24	0.000	2.24	3.02
Abandonou_escola	-1.56	0.259	-6.03	0.000	-2.07	-1.05
Reprovou	-24.87	0.117	-210.87	0.000	-25.10	-24.63
Negro	-8.02	0.181	-44.31	0.000	-8.38	-7.67

Freq._biblioteca	-4.49	0.165	-27.18	0.000	-4.81	-4.16
Le_livros_sempre	7.93	0.121	65.42	0.000	7.69	8.16
Faz_dever_port_sempre	5.52	0.120	45.96	0.000	5.28	5.76
Faz_dever_mat_sempre	2.70	0.117	22.99	0.000	2.47	2.93
_cons	240.41	0.234	1025.70	0.000	239.95	240.87

Fonte: Dados estimados a partir do Prova Brasil 2011 por meio do *software stata 11*.
Elaboração própria.

APÊNDICE 22. Regressão Mínimos Quadrados Ponderados pelo Propensity Score. Estimativa 2: Pré – 9º ano. Matemática.

Variáveis	Coefficiente	Erro padrão	T	P> t	Intervalo de confiança de 95%	
Inicio_pre	10.26	0.116	88.36	0.000	10.03	10.48
Trab_fora	-1.80	0.139	-12.93	0.000	-2.08	-1.53
Sempre_escola_publica	3.20	0.202	15.87	0.000	2.81	3.60
Abandonou_escola	-4.96	0.256	-19.36	0.000	-5.46	-4.46
Reprovou	-24.09	0.118	-203.85	0.000	-24.32	-23.86
Negro	-8.48	0.179	-47.17	0.000	-8.84	-8.13
Freq._biblioteca	-4.95	0.167	-29.56	0.000	-5.28	-4.62
Le_livros_sempre	0.68	0.122	5.62	0.000	0.44	0.92
Faz_dever_port_sempre	-2.31	0.120	-19.25	0.000	-2.55	-2.07
Faz_dever_mat_sempre	15.20	0.117	129.64	0.000	14.97	15.43
_cons	244.93	0.237	1031.81	0.000	244.47	245.39

Fonte: Dados estimados a partir do Prova Brasil 2011 por meio do *software stata 11*.
Elaboração própria.

APÊNDICE 23. Regressão Mínimos Quadrados Ponderados pelo Propensity Score. Estimativa 3: Ensino Infantil – 9º ano. Português.

Variáveis	Coefficiente	Erro padrão	t	P> t	Intervalo de confiança de 95%	
Inicio_prim_inf	6.47	0.111	57.88	0.000	6.26	6.69
Trab_fora	-11.60	0.117	-98.45	0.000	-11.84	-11.37
Sempre_escola_publica	1.13	0.152	7.42	0.000	0.83	1.43
Abandonou_escola	-2.47	0.221	-11.15	0.000	-2.90	-2.03
Reprovou	-25.24	0.099	-253.69	0.000	-25.43	-25.04
Negro	-8.17	0.148	-54.92	0.000	-8.46	-7.88
Freq._biblioteca	-5.27	0.140	-37.55	0.000	-5.55	-5.00

Le_livros_sempre	8.30	0.103	80.30	0.000	8.10	8.50
Faz_dever_port_sempre	5.53	0.102	54.19	0.000	5.33	5.73
Faz_dever_mat_sempre	2.53	0.100	25.37	0.000	2.34	2.73
_cons	243.23	0.196	1238.94	0.000	242.84	243.62

Fonte: Dados estimados a partir do Prova Brasil 2011 por meio do *software stata 11*.
Elaboração própria.

APÊNDICE 24. Regressão Mínimos Quadrados Ponderados pelo Propensity Score. Estimativa 3: Ensino Infantil – 9º ano. Matemática.

Variáveis	Coeficiente	Erro padrão	t	P> t	Intervalo de confiança de 95%	
Inicio_prim_inf	5.81	0.113	51.35	0.000	5.59	6.04
Trab_fora	-3.19	0.118	-26.99	0.000	-3.42	-2.96
Sempre_escola_publica	1.43	0.155	9.28	0.000	1.13	1.74
Abandonou_escola	-5.47	0.220	-24.88	0.000	-5.90	-5.04
Reprovou	-24.34	0.099	-243.96	0.000	-24.54	-24.14
Negro	-8.41	0.148	-56.81	0.000	-8.70	-8.12
Freq._biblioteca	-6.22	0.142	-43.74	0.000	-6.50	-5.94
Le_livros_sempre	1.16	0.104	11.14	0.000	0.95	1.36
Faz_dever_port_sempre	-2.40	0.101	-23.63	0.000	-2.60	-2.20
Faz_dever_mat_sempre	15.07	0.099	151.54	0.000	14.87	15.26
_cons	248.40	0.199	1247.80	0.000	248.01	248.79

Fonte: Dados estimados a partir do Prova Brasil 2011 por meio do *software stata 11*.
Elaboração própria.