

## Área 12 – Economia Social e Demografia Econômica

### FREQUÊNCIA AO ENSINO INFANTIL E FORMAÇÃO DE HABILIDADES COGNITIVAS E SOCIOEMOCIONAIS EM ESCOLAS DA REDE ESTADUAL DO RIO DE JANEIRO; UMA ANÁLISE DE MISTURAS FINITAS E *PROPENSITY SCORE MATCHING*

Walcir Soares da Silva Junior<sup>1</sup> e Flávio de Oliveira Gonçalves<sup>2</sup>  
RESUMO

O objetivo deste artigo é medir o impacto da frequência ao ensino infantil nas habilidades cognitivas e capacidades socioemocionais em uma amostra de alunos da rede estadual do Rio de Janeiro. Utiliza-se as bases Prova Brasil 2013 e SENNA. Sob a hipótese de seleção por observáveis utiliza-se a metodologia *Propensity Score Matching* (PSM). O método de Mínimos Quadrados Ponderados pelos escores de propensão (MQP) foi utilizado para controlar variáveis que influenciam o desempenho. Sob a hipótese de heterogeneidade não observada estima-se o modelo de misturas finitas (FMM). Foram encontrados efeitos positivos e significativos para a pré-escola em todos os anos avaliados nas dimensões cognitivas. Efeitos positivos e significativos da pré-escola foram encontrados sobre as dimensões locus de controle e estabilidade emocional dos alunos do quinto ano e sobre a dimensão abertura a novas experiências para os alunos do quinto e nono anos.

**Palavras-chave:** ensino infantil, *propensity score matching* (PSM), habilidades cognitivas, competências socioemocionais, misturas finitas (FMM).

#### ABSTRACT

This paper aims to measure the impact of the frequency childhood education in cognitive and social-emotional skills in a sample of students from Rio de Janeiro. It was used the data bases Prova Brasil 2013 and SENNA instrument. Under the assumption of selection by observable variables, it was used propensity score matching methodology. The Least Squares weighted by propensity scores methodology was used to control variables that influence performance. Under the assumption of unobserved heterogeneity were estimated results from a model of finite mixtures. We found significant positive effects for preschool in every year over cognitive dimensions. And we found significant positive effects of preschool over the locus of control and emotional stability of fifth graders and over openness to new experiences to the students of the fifth and ninth years.

**Keywords:** early childhood education, propensity score matching (PSM), cognitive skills, social-emotional skills, finite mixtures (FMM).

**Classificação JEL:** I21, I24, I25, I28

---

<sup>1</sup> Doutorando do Programa de Pós Graduação em Desenvolvimento Econômico UFPR.

<sup>2</sup> Doutor em Economia pela UNB e professor do Programa de Pós Graduação em Desenvolvimento Econômico UFPR.

## 1. Introdução

Há uma ampla literatura dentro da área de economia da educação que estabeleceu ou tenta estabelecer associações e relações causais entre habilidades e resultados educacionais e no mercado de trabalho. Habilidades aqui são consideradas aquelas competências que têm algum nível de maleabilidade durante a vida, e podem ser divididas em dois grupos: habilidades cognitivas (*hard skills*) e as habilidades ou competências socioemocionais (*soft skills*), também conhecidas como não-cognitivas. Habilidades cognitivas estão relacionadas à inteligência, como capacidade de raciocínio e lógica, por exemplo, e têm como medida mais próxima o QI. Capacidades socioemocionais ou habilidades não-cognitivas são traços de personalidade que estão relacionados à motivação, perseverança, criatividade e autoestima, por exemplo. Entender como essas habilidades se desenvolvem e afetam o desempenho dos alunos tem, portanto, uma grande relevância nesta literatura.

A primeira infância <sup>3</sup>tem tido grande destaque nos estudos sobre educação e sobre o desenvolvimento de habilidades tanto cognitivas quanto socioemocionais (Curi e Menezes-Filho, 2006). Heckman (2000), Cunha et. al. (2006) e Carneiro e Heckman (2003) reconhecem que tanto as habilidades cognitivas quanto as competências socioemocionais possuem um componente genético e outro componente influenciado pelo ambiente.

Segundo Heckman (2008), evidências da epigenética<sup>4</sup> sugerem que a distinção entre genética e ambiente é obsoleta, assim como a prática de discernir entre resultados “*nature*” e “*nurture*”, ou seja, o que é da natureza e o que é estimulado. Segundo Heckman existe uma extensa literatura mostrando que a interação entre os genes e o ambiente é central para explicar o desenvolvimento humano e animal<sup>5</sup>. Autores como Caspi, Williams, Kim-Cohen et al. (2007) mostram que o desenvolvimento intelectual das crianças é influenciado por fatores genéticos e ambientais. Turkheimer, Haley, Waldron et al. (2003) encontraram um importante papel para o ambiente na determinação de que o QI possa ser herdado.

O foco nos programas educacionais da primeira infância enquanto políticas públicas mais eficientes vem da evidência de que a primeira infância parece ser um período sensível e crítico na produção de algumas habilidades importantes, como por exemplo o QI, que se estabiliza aos dez anos de idade. (Cunha et al., 2006). Segundo os autores, determinadas habilidades possuem um período sensível, em que seu desenvolvimento é muito mais eficiente. Se a duração desse período é limitada, esse passa a ser chamado também de período crítico, a partir do qual a maleabilidade de aprendizado é muito pequena.

No entanto, para mensurar o impacto do ensino infantil, não seria possível simplesmente comparar as médias dos alunos que frequentaram ou não o ensino infantil. Os alunos apenas podem ser comparáveis se eles pertencem a grupos com características de seleção parecidas (a seleção aqui ocorre na escolha dos pais entre matricular ou não o filho

---

<sup>3</sup> Consideram-se aqui todas as fases que precedem o ensino fundamental: a creche (0 a 3 anos) e pré-escola (4 a 5 anos).

<sup>4</sup> Termo utilizado na biologia para se referir a características de organismos unicelulares e multicelulares que são estáveis ao longo de diversas divisões celulares, mas que não envolvem mudanças na sequência de DNA do organismo. Herança epigenética é a transmissão de experiências ocorridas com os pais para os filhos e que não ocorre através do DNA.

<sup>5</sup> Rutter (2006).

no ensino infantil, já que neste estágio, no Brasil, ele não é obrigatório). Desse modo, se faz necessário ter um contrafactual. Para controlar essa seleção por algumas variáveis observáveis, será utilizado um modelo de *Propensity Score Matching* que tem como objetivo definir um grupo de controle para comparação (contrafactual) modelando a probabilidade de participar do tratamento dadas algumas características observáveis.

Levando em consideração que algumas variáveis relacionadas ao aluno – como raça, condição socioeconômica, ter sempre estudado em escola pública, trabalhar fora, já ter sido reprovado – influenciam no seu desempenho, será utilizado o método dos Mínimos Quadrados Ponderados pelos Escores de Propensão. Nesse método também será considerada a hipótese de correlação entre notas dentro de uma mesma escola.

Por fim, para controlar determinantes não observáveis será utilizado um modelo de misturas finitas (FMM) ponderado também pelos escores de propensão. Esse modelo utiliza uma estimação por máxima verossimilhança em que a densidade de probabilidades pode variar.

A principal contribuição deste trabalho é apresentar uma medida para aquelas variáveis reconhecidamente consideradas na literatura como não observáveis, como uma variável de interesse dos pais construída através de análise componentes principais policórica<sup>6</sup> e os seis construtos para capacidades socioemocionais mensuradas pelo instrumento SENNA.

Esse artigo está dividido em seis seções, a contar com esta introdução. A seção 2 fará uma breve apresentação das evidências a respeito das habilidades cognitivas e competências socioemocionais, e apresentará um modelo teórico explorando a escolha dos pais entre colocar ou não os filhos no ensino infantil. A seção 3 além de apresentar brevemente as metodologias utilizadas neste artigo e apresenta também algumas evidências a partir das descritivas que lançam luz ao problema da seleção, e estima um modelo de escolha binária. A seção 4 apresenta os resultados das estimações. As considerações finais são apresentadas na seção 5.

## **2. Habilidades cognitivas e capacidades socioemocionais e sua importância na produção de resultados educacionais**

Diversos estudos associam as habilidades cognitivas, relativamente fáceis de serem medidas através de testes padronizados, com resultados educacionais e no mercado de trabalho. No entanto, a discussão sobre as competências socioemocionais<sup>7</sup> ainda é bastante controversa, seja pela escolha do modelo mais adequado para classificar e explicar que atributos socioemocionais são mais importantes, seja mesmo pelo consenso de se “competências socioemocionais” são o meio adequado de se descrever o conjunto de questões sob discussão.

Outro problema é a escassez de trabalhos experimentais, principalmente com dados brasileiros, para realmente dar robustez às evidências de causalidade entre competências socioemocionais e resultados educacionais e no mercado de trabalho. Trabalhos como

---

<sup>6</sup> A seção 4 detalha a construção deste índice.

<sup>7</sup> Neste trabalho, é definido o termo habilidades ou competências “socioemocionais” em detrimento do termo habilidades “não-cognitivas”. De acordo com Borghans et al. (2008), poucos aspectos do comportamento humano podem ser considerados como separados da cognição e, portanto, chamar competências socioemocionais de “não-cognitivas” é no mínimo inadequado.

Bowles e Gintis (2002), Farkas (2003), Heckman, et al. (2006), Jencks (1979); Lleras (2008) mostram que existem muitas evidências de que as competências socioemocionais estão associadas com resultados positivos, mas a evidência robusta de relação causal ainda é bastante limitada. Portanto, justifica-se também a produção de evidências empíricas controlando medidas de competências socioemocionais a partir de métodos experimentais ou quase-experimentais.

Para tratar teoricamente das competências socioemocionais, dentre os diversos modelos existentes, o modelo dos *Big Five* tem sido uma forma comumente utilizada pelos economistas na hora de classificar e mensurar essas competências. Esse modelo tenta explicar a maior parte da variabilidade da personalidade humana em cinco fatores: conscienciosidade (propensão a ser organizado, responsável e esforçado), extroversão (direcionamento de interesses e energia em direção ao mundo externo de pessoas e coisas, ao invés do mundo interno de experiências subjetivas), neuroticismo/estabilidade emocional (propensão ao desequilíbrio psicológico e consistência de reações emocionais), amabilidade/cooperatividade (propensão a agir de modo cooperativo e não egoísta) e abertura a novas experiências (propensão a aceitar novas experiências estéticas, culturais ou intelectuais).<sup>8</sup>

Recentemente no Brasil, a parceria entre as instituições Organização para a Cooperação e Desenvolvimento Econômico (OCDE), o Centro para Pesquisa e Inovação Educacional (CERI) e o Instituto Ayrton Senna (IAS) resultou na construção de um instrumento pioneiro para medir competências socioemocionais batizado de SENNA. De acordo com Santos (2014) o instrumento SENNA apresenta além da medição dos cinco atributos, um atributo adicional chamado Locus de Controle. Locus de Controle é uma variável de autoestima em que indivíduos com alto Locus de Controle (ou locus interno) tendem a acreditar que a sua vida e realizações são resultados de suas próprias escolhas no passado, enquanto indivíduos com baixo Locus de Controle (locus externo) acreditam que o que acontece em suas vidas é resultado do acaso, sorte ou da ação de terceiros. A base SENNA contém além dos seis construtos mensurados pelo instrumento, informações a respeito das características individuais do aluno, do ambiente familiar em que vivem, e do comportamento dos pais para uma amostra de alunos da rede estadual do Rio de Janeiro.

Neste artigo, como *proxy* para cognição utilizou-se as notas dos alunos em português e matemática no 5º e 9º anos do ensino fundamental das escolas públicas coletados do Prova Brasil 2013. Como medida para as competências socioemocionais, foram utilizados os seis construtos mensurados pelo instrumento SENNA. Como a análise é feita através da união de duas bases de dados ao nível do aluno, a amostra é apenas representativa para escolas da rede estadual do Rio de Janeiro.

## 2.1 O processo de escolha dos pais no período do ensino infantil

Essa seção estende o modelo de Cunha, Heckman e Schenach (2010) de formação de habilidades, enfatizando o período do ensino infantil e a escolha dos pais entre colocar seus filhos na creche ou pré-escola, ou diretamente no primário. O novo modelo assume  $T$  estágios de desenvolvimento  $I$  e  $F$ , em que  $I+F \leq T$ . O período  $I$  representa o ensino infantil, que acontece na primeira infância e o período  $F$ , o ensino fundamental, que no Brasil inicia-se no primário. Os resultados dos alunos são produzidos pelas habilidades cognitivas,  $\theta_{F+I}^C$ ,

---

<sup>8</sup> Para mais detalhes, Goldberg (1980).

e competências socioemocionais  $\theta_{F+1}^N$  no final do segundo estágio  $F$ .

Desse modo, as crianças que iniciam os estudos diretamente no primário, estudam apenas no segundo estágio de desenvolvimento  $F$ , e no estágio  $I$  ficam em casa, e são educadas diretamente pelos seus pais. Os investimentos dos pais no período  $I$  na habilidade  $k$  são denotados por  $I_F^k$   $k \in \{C, N\}$ , onde  $C$  representa as habilidades cognitivas e  $N$  as habilidades não-cognitivas ou competências socioemocionais. Neste modelo, os investimentos dos pais no período  $F$  (ensino fundamental) na habilidade  $k$ , que são denotados por  $I_F^k$   $k \in \{C, N\}$ , são fixos, considerando que o ensino primário no Brasil é obrigatório e na maior parte público, sendo assim os pais não fazem uma escolha entre colocar ou não os seus filhos no ensino fundamental, e essa escolha, que é o ponto principal deste modelo, se dá apenas no ensino infantil.

Assim, cada agente nasce com dotações de habilidades iniciais  $\theta_d = (\theta_d^C, \theta_d^N)$ . Os fatores do ambiente familiar e da genética também influenciam essas dotações iniciais. Suponha que  $\theta_I = (\theta_I^C, \theta_I^N)$  denota o vetor de habilidades estocadas no período  $I$ .  $\theta_F = (\theta_F^C, \theta_F^N)$  denota o vetor de habilidades estocadas no período  $F$ .  $\eta_I^k = (\eta_I^C, \eta_I^N)$  denota os choques e/ou insumos não observados que afetam a acumulação de habilidades cognitivas e socioemocionais, respectivamente, no período  $I$  e  $\eta_F^k = (\eta_F^C, \eta_F^N)$  os choques não observados que afetam a acumulação de habilidades cognitivas e socioemocionais, respectivamente, no período  $F$ . Análogo ao investimento,  $k \in \{C, N\}$ , onde  $C$  representa as habilidades cognitivas e  $N$  as habilidades não-cognitivas ou competências socioemocionais.

Desse modo, a função de produção no estágio  $F$  é:

$$\theta_F^k = f_F^k(\theta_F, I_F^k, \eta_F^k) \quad (1)$$

Apesar da suposição de que no estágio  $F$  o componente  $I_F^k$  de investimento dos pais é fixo, no estágio  $I$ , esse componente depende de outras variáveis. No estágio  $I$  a função de produção é:

$$\theta_I^k = g_I^k(\theta_I, I_I^k, \eta_I^k) \quad (2)$$

$$I_I^k = h_I^k(\theta^p, \lambda^p, \gamma^k) \quad (3)$$

onde  $\theta^p = (\theta_P^C, \theta_P^N)$  são as habilidades cognitivas e socioemocionais dos pais, respectivamente;  $\lambda^p$  é uma medida de altruísmo que denota o interesse dos pais pela educação de seus filhos; e  $\gamma$  é o preço do investimento em  $k$  no período  $I$ .

Assume-se que  $f_I^k$  e  $f_F^k$  são monotônicas e crescentes nos seus argumentos, duas vezes continuamente diferenciáveis e côncavas em  $I_I^k$ . Desse modo, os investimentos em  $k$  são produtivos, porém, a taxas decrescentes. No entanto, os estoques de habilidades nos períodos correntes produzem o próximo período de habilidades e afetam a produtividade dos investimentos no período corrente tal qual a evidência da auto produtividade. Isso acontece porque  $\theta_F$  e  $\theta_I$  são argumentos das respectivas funções de produção. Assim, as crianças que tiveram os respectivos investimentos na habilidade  $k$  no período  $I$  produzem a habilidade  $k$  no período  $F$  de maneira mais eficiente.

A evidência da complementaridade pode ser mostrada na equação 4. A complementaridade entre o estoque de habilidade  $I$  e a produtividade do investimento  $I_F^k$  na produção da habilidade  $k$  no período  $F$  aumenta se:

$$\frac{\partial^2 f_F^k(.)}{\partial I^k_t \partial \theta^l_F} > 0 \quad (4)$$

$$l, k \in \{C, N\}$$

Desse modo, no período  $F$  os estoques de habilidades e competências promovem a aquisição de habilidades através da realização de investimentos mais produtivos. Estudantes com grandes habilidades cognitivas e socioemocionais iniciais são mais eficientes no aprendizado futuro de ambas habilidades, o que é compatível com o conceito de complementaridade. O resultado do aluno  $j$ ,  $Q_j$  ao final do período  $F$ , é produzido pela combinação de diferentes habilidades estocadas no final desse período.

$$Q_j = m_j(\theta^C_F, \theta^N_F), \quad j \in \{1, \dots, J\} \quad (5)$$

Desse modo, os resultados do modelo de Cunha, Heckman e Schenach (2010) permanecem: o resultado captura o conceito dual de que ambas habilidades, cognitivas e socioemocionais, importam na predição de resultados importantes. Resultados nessa extensão se traduzem nas medições de estoque de habilidades cognitivas e capacidades socioemocionais feitas no meio do período escolar, neste caso, especificamente no 5º e 9º anos do ensino fundamental, que serão importantes para outros inúmeros resultados mais tarde, como salários e nível de escolaridade, por exemplo.

Retomando a função de produção do ensino infantil (2), tem-se que ela é uma função do estoque de habilidades  $\theta_I = (\theta^C_I, \theta^N_I)$  estocadas no período  $I$  (que nesse período são basicamente as dotações iniciais provenientes do ambiente e da genética), dos investimentos dos pais na habilidade  $k$  no período  $I$ ,  $I^k_I$ , (os pais colocarem as crianças no ensino infantil) e dos choques e/ou insumos  $\eta^k_{I=}$  ( $\eta^C_I, \eta^N_I$ ) não observados que afetam a acumulação de habilidades cognitivas e socioemocionais, respectivamente, no período  $I$ .

O foco dessa extensão é justamente o  $I^k_I$ , que representa o investimento dos pais no período  $I$ . Esse investimento é uma função  $h^k_I$  de três componentes,  $\theta^P = (\theta^C_P, \theta^N_P)$  que são as habilidades cognitivas e socioemocionais dos pais, respectivamente,  $\lambda^P$  que é uma medida de altruísmo que denota o interesse dos pais pela educação de seus filhos e  $\gamma^k_I$  que representa o preço do investimento em  $k$  no período  $I$ . As derivadas parciais da função com relação a essas variáveis são:

$$\frac{\partial h^k_I(.)}{\partial \theta^P} > 0 \quad \frac{\partial^2 h^k_I(.)}{\partial \theta^P} \leq 0 \quad (6)$$

$$\frac{\partial h^k_I(.)}{\partial \lambda^P} > 0 \quad \frac{\partial^2 h^k_I(.)}{\partial \lambda^P} > 0 \quad (7)$$

$$\frac{\partial h^k_I(.)}{\partial \gamma^k_I} < 0 \quad \frac{\partial^2 h^k_I(.)}{\partial \gamma^k_I} \leq 0 \quad (8)$$

A equação 6 mostra que o investimento no período do ensino infantil é uma função crescente nas habilidades cognitivas e socioemocionais dos pais, no entanto, a taxas decrescentes. Dessa forma, quanto maior a escolaridade dos pais, maior o custo de oportunidade entre tomar conta dos filhos em casa ou colocá-los na escola. Mas para níveis altos de escolaridade, os pais podem preferir educar os filhos em casa ou mesmo leva-los

ao trabalho.

Já a equação 7 mostra que o investimento no período do ensino infantil é uma função crescente no altruísmo dos pais, ou interesse deles pela educação de seus filhos. Como um exemplo, pode-se considerar um primeiro nível em que o interesse dos pais reflete apenas no discurso destes, que valorizam a educação de seus filhos enquanto essa valorização depende apenas de incentivá-los a estudar. Já num segundo nível, o interesse dos pais reflete no exemplo destes, que valorizam a educação de seus filhos com ações específicas, como colocá-los numa boa escola e cumprirem deveres como ler para os filhos e frequentar reuniões escolares.

Já a equação 8 mostra que o investimento no período do ensino infantil é uma função decrescente no preço do ensino infantil. Dada a escolaridade e o respectivo custo de oportunidade entre o mercado de trabalho e tomar conta dos filhos, os pais podem considerar o preço entre assumir um custo de alguém cuidando de seus filhos em casa ou colocarem seus filhos na escola, que pode ser privada, mas também pode ser pública e ter relativamente altos custos relacionados com transporte ou materiais. Dentro do preço da escola estão as variáveis qualidade da educação e disponibilidade de creches e escolas perto do domicílio.

No Brasil, em 2014, de acordo com dados do Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira (INEP), o ensino infantil era composto de 63,3% das matrículas em creches públicas. Já para a pré-escola, 74,6% das matrículas em 2014 eram públicas. Desse modo, diretamente na seleção, a renda parece não ser um fator relevante, hipótese atestada através de um teste de médias que mostrou que não existe diferença estatisticamente significativa de renda entre os grupos de alunos que iniciaram na creche ou na pré-escola para aquele grupo que iniciou diretamente no primário.

No entanto, para as habilidades dos pais  $\theta^p$ , cuja escolaridade dos pais foi utilizada como *proxy*, pais com maior escolaridade possuem um maior custo de oportunidade com relação ao mercado de trabalho. Isso se dá principalmente com relação às mães, o que faz com que mães com mais educação estejam mais propensas a escolher trabalhar e colocarem seus filhos na escola mais cedo, e acima de certo ponto da renda, tomarem conta elas próprias da educação infantil de seus filhos. Desse modo, a relação da renda com a escolha entre colocar ou não a criança no ensino infantil se dá através da escolaridade e o seu custo de oportunidade no mercado de trabalho. Menezes Filho (2007) e Soares e Collares (2006) apresentam evidências que sustentam essa hipótese de que escolaridade importa nessa escolha.

Desse modo, existem quatro principais fatores que parecem afetar essa escolha: a) escolaridade dos pais (especialmente da mãe); b) disponibilidade de creche ou escola nas proximidades onde a criança mora; c) a criança morar na mesma casa em que a avó ou parente que possa tomar conta dela enquanto a mãe trabalha; d) o interesse dos pais pela educação dos seus filhos.

### 3. Metodologias e evidências a partir das descritivas

Esta seção apresenta uma breve abordagem das metodologias utilizadas neste artigo. Para auxiliar na especificação do modelo de escolha, essa seção apresentará também algumas estatísticas descritivas de variáveis que são consideradas importantes no momento da seleção do aluno entre creche e pré-escola (ensino infantil) ou iniciar diretamente no primário. Retomando o modelo teórico apresentado da seção 2.1, tem-se que o investimento dos pais no ensino infantil é uma função  $h^k_I$  de três componentes:  $\theta^p = (\theta^{C_p}, \theta^{N_p})$  que são as

habilidades cognitivas e socioemocionais dos pais, respectivamente;  $\lambda^p$  que representa uma medida de altruísmo denotando o interesse dos pais pela educação de seus filhos; e  $\gamma^k_I$  que representa o preço do investimento em  $k$  no período  $I$ . Desse modo, será preciso selecionar algumas *proxies* que representem essas variáveis no modelo empírico.

Antes de discutir essas *proxies* para o modelo empírico, as três metodologias utilizadas são apresentadas brevemente na próxima seção.

### 3.1 *Propensity Score Matching*, Mínimos Quadrados Ponderados pelos Escores de Propensão (MQP) e Modelo de Misturas Finitas (FMM).

O método de *propensity score* considera a hipótese de haver problemas de seleção por variáveis observáveis na frequência da educação infantil. Desse modo, o objetivo é definir um grupo de controle para comparação modelando a probabilidade de participar do tratamento conforme uma função de características observadas  $X$  que contém toda a informação desse vetor, o qual é chamado de *propensity scores* ou escores de propensão.

Assim, nesse método, o pareamento entre o grupo tratado e o controle é baseado nessa probabilidade, que pode ser calculada através de um modelo *probit* (Rosenbaum e Rubin, 1983). A validade deste método depende de duas hipóteses. A primeira diz que o vetor de variáveis observáveis  $X$  deve conter todas as informações sobre o resultado potencial na ausência do tratamento ( $Y(0)$ ) que o indivíduo possui ao tomar a decisão de participar ou não do tratamento. A segunda condição garante que cada indivíduo no grupo tratamento tenha um par próximo de comparação no grupo de controle na distribuição do escore de propensão e no qual a variável resultado corresponderia à situação deste indivíduo na ausência do tratamento. (Becker e Ichino, 2002).

Já o método dos Mínimos Quadrados Ponderados pelos Escores de Propensão (MQP) contorna o problema de má especificação, por ser uma combinação do *propensity score* com a regressão, tendo o benefício de ser duplamente robusto. 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 (Imbens e Wooldridge, 2008).<sup>9</sup>

Por fim, o modelo de misturas finitas é uma combinação convexa semiparamétrica de diferentes densidades de probabilidade, em que cada densidade possui um peso  $p_i > 0$ . Desse modo, o modelo propõe uma mistura de distribuições de probabilidade com o objetivo de conseguir modelar heterogeneidade não observada nos efeitos para diferentes classes<sup>10</sup>. Assim, identifica-se os componentes da mistura (estimando os seus parâmetros através de estimação por máxima verossimilhança) e estimando as probabilidades de uma observação pertencer a esse componente. O método também pode ser usado para análise de *clusters*, sendo mais flexível e baseado em um modelo estatístico.

Explorando as estatísticas descritivas dos dados, foi estimada uma função de Densidade de Kernel<sup>11</sup> para as notas dos alunos em Português e Matemática para o 5º e 9º anos. Nos gráficos 1 e 2 é observa-se que para os dados do 5º ano, nos grupos que

---

<sup>9</sup> Para mais detalhes, ver Becker e Ichino (2002) e Pinto (2012).

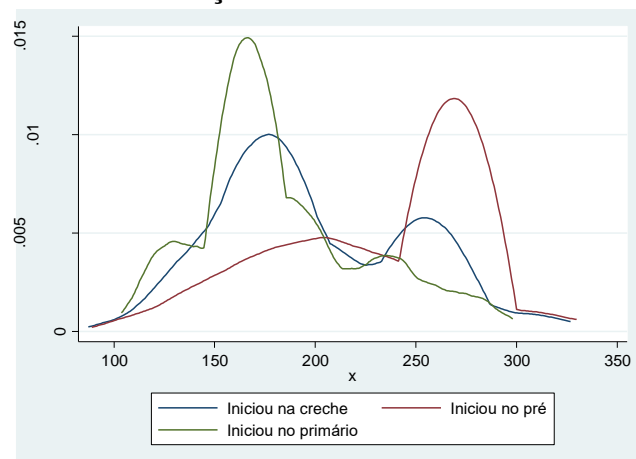
<sup>10</sup> McLachlan e Peel (2000).

<sup>11</sup> Rosenblatt (1956).



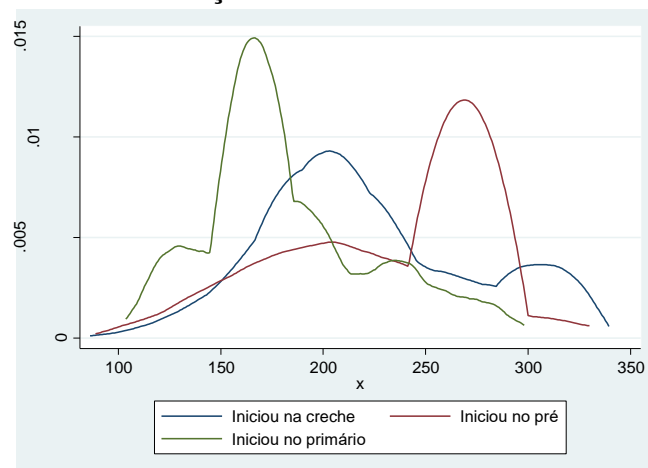
frequentaram o ensino infantil, as distribuições da variável de resultado<sup>12</sup> têm distribuição bimodal.

**Gráfico 1. Função de Densidade de Kernel – Português – 5º Ano.**



Fonte: elaboração própria a partir dos dados do Prova Brasil 2013 e SENNA.

**Gráfico 2. Função de Densidade de Kernel – Matemática – 5º Ano.**



Fonte: elaboração própria a partir dos dados do Prova Brasil 2013 e SENNA.

Assim, se existe essa suposição de bimodalidade na distribuição das notas dos alunos que frequentaram o ensino infantil, o modelo de misturas finitas permite misturas diferentes distribuições de probabilidade. No entanto, o modelo de regressão linear não é apropriado, já que isso viola a suposição de normalidade.

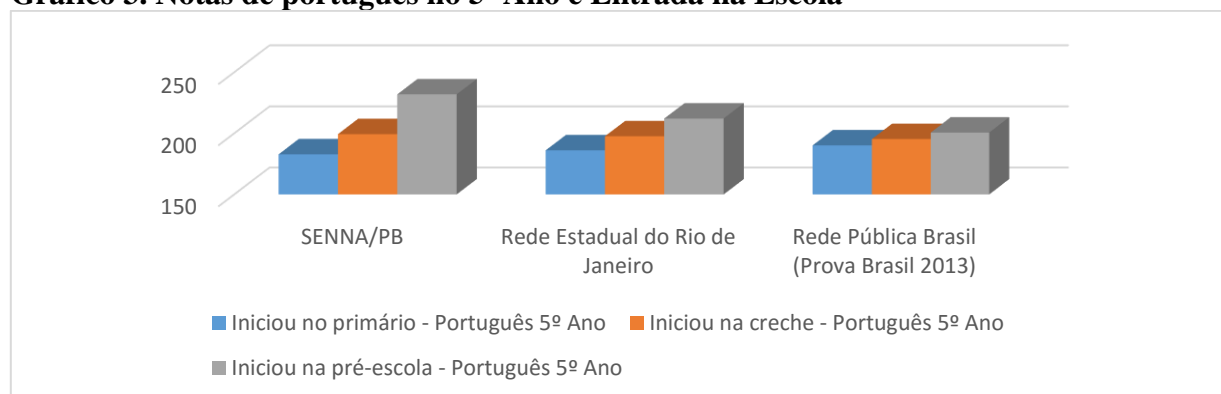
### 3.2 Características dos alunos da amostra SENNA/PB

Para o exercício em questão as bases SENNA e Prova Brasil 2013 foram unidas, originando a nova base de dados SENNA/PB. É preciso salientar que para cada estimação o número de observações variou, devido à quantidade de valores faltantes em ambas amostras. A base de dados SENNA é uma amostra representativa da rede estadual do Rio de Janeiro.

<sup>12</sup> Notas dos alunos em Português e matemática na escala SAEB.

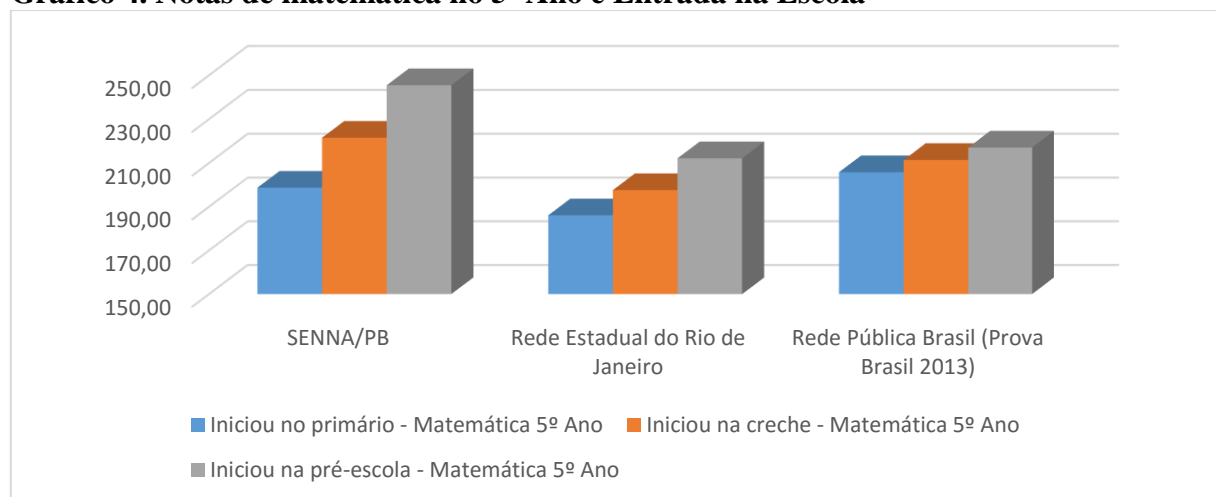
Os gráficos 3 e 4 apresentam as médias das notas dos alunos para português e matemática, respectivamente, no quinto ano, distribuídos pelo ano de entrada na escola. Os dados são referentes às diferentes amostras, SENNA/PB, rede estadual do Rio de Janeiro, e rede pública nacional. Pode-se observar que para ambas as disciplinas, os alunos que iniciaram no ensino infantil possuem uma nota superior aos alunos que iniciaram diretamente no primário. Isso é ainda mais evidente nos alunos que iniciaram na pré-escola.

**Gráfico 3. Notas de português no 5º Ano e Entrada na Escola**



Fonte: elaboração própria a partir dos dados do Prova Brasil 2013 e SENNA.

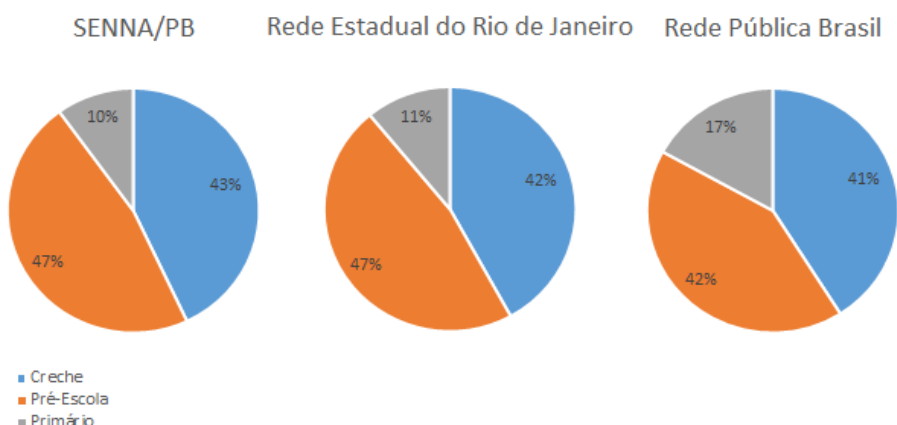
**Gráfico 4. Notas de matemática no 5º Ano e Entrada na Escola**



Fonte: elaboração própria a partir dos dados do Prova Brasil 2013 e SENNA.

O gráfico 5 apresenta os dados de frequência no ensino infantil para as diferentes amostras. Para as amostras SENNA-PB e os dados censitários da rede estadual do Rio de Janeiro, 43% e 42% dos alunos, respectivamente, iniciaram os estudos diretamente na creche, 47% iniciaram diretamente na pré-escola nas duas amostras e 10% e 11%, respectivamente, diretamente no primário. Já para os dados comparativos para toda a rede pública brasileira, esses números são de 41%, 42% e 17% para creche, pré-escola e primário, respectivamente. Desse modo, a frequência na creche está ligeiramente super-representada na amostra SENNA-PB, e a frequência no primário, ligeiramente sub-representada.

**Gráfico 5. Frequência no Ensino Infantil – Diferentes amostras.**



Fonte: elaboração própria a partir dos dados do Prova Brasil 2013 e SENNA.

Um teste de médias entre essas duas amostras resultou que para a frequência na pré-escola, a diferença não é estatisticamente significativa. Já para a creche e primário, as diferenças das duas amostras são estatisticamente significantes a 1%, ainda que pequenas.

### 3.3 Características dos alunos que frequentaram o ensino infantil

A habilidade dos pais  $\theta^P$ , levando em conta o maior custo de oportunidade para mães com alta escolaridade entre o mercado de trabalho e cuidar dos seus filhos em casa, será representada por uma variável binária em que 1 = mãe possui ensino superior completo e 0 = mãe não possui ensino superior completo.  $\lambda^P$  será representada por um índice de interesse dos pais construído através de análise de componentes principais policórica, específica para variáveis discretas<sup>13</sup>, incluindo variáveis como a mãe ler para o filho, os pais frequentarem reuniões da escola, incentivarem seus filhos a ler, estudar, fazer o dever de casa, comparecer às aulas e os pais conversarem com seus filhos sobre a escola.

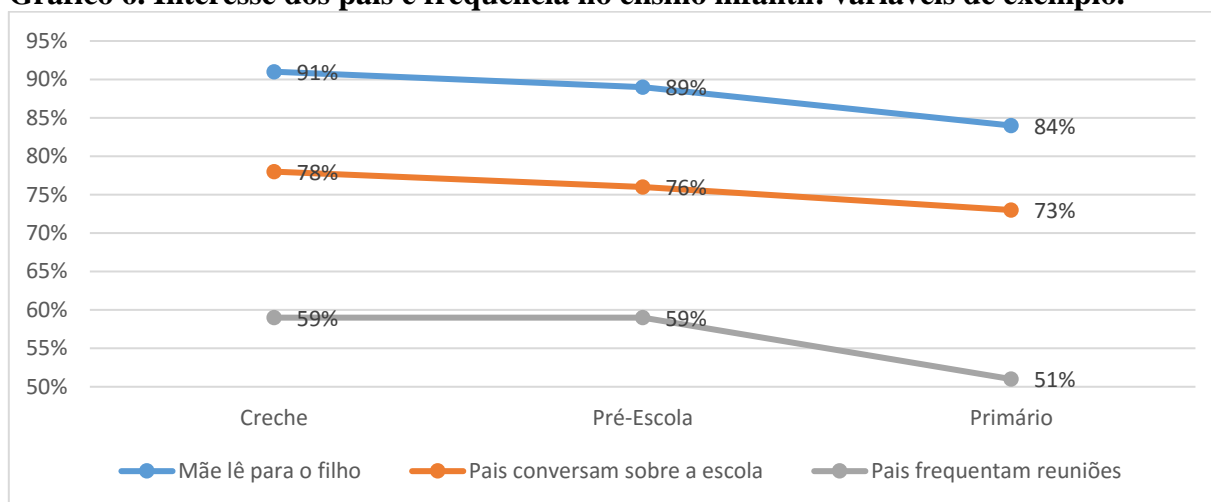
Já a variável  $\gamma^k_I$  que representa o preço do investimento em  $k$  no período  $I$ , empiricamente é representada por dois tipos de custo associado ao ensino infantil, que no Brasil é em sua maioria público: a) disponibilidade de creche ou escola nas proximidades onde a criança mora; e b) a criança morar na mesma casa em que a avó ou parente que possa tomar conta dela enquanto a mãe trabalha. Pelo menos um problema de viés por não observáveis pode ser apontado como um dos limites deste trabalho, que é a disponibilidade de escolas ou creches perto do lugar onde o aluno mora. Como *proxy* para ter à disposição alguém que possa tomar conta do filho em casa, será utilizada a variável binária ter empregada, em que 1 representa ter empregada e 0 representa não ter.

O gráfico 6 apresenta as variáveis de interesse que representam ações específicas dos pais como mãe ler para o filho sempre ou quase sempre, pais conversarem sobre o que

<sup>13</sup> Segundo Kolenikov (2004), variáveis discretas violam a importante suposição de normalidade na análise de componentes principais. Outra suposição violada é que no caso das variáveis discretas as covariâncias ou correlações estimadas entre as variáveis não são iguais às verdadeiras e não observadas covariâncias ou correlações, sendo na maior parte das vezes viesados em direção à 0, dentre outros problemas.

acontece na escola sempre ou quase sempre e pais frequentarem as reuniões escolares sempre ou quase sempre nos diferentes grupos de entrada na escola

**Gráfico 6. Interesse dos pais e frequência no ensino infantil: variáveis de exemplo.**

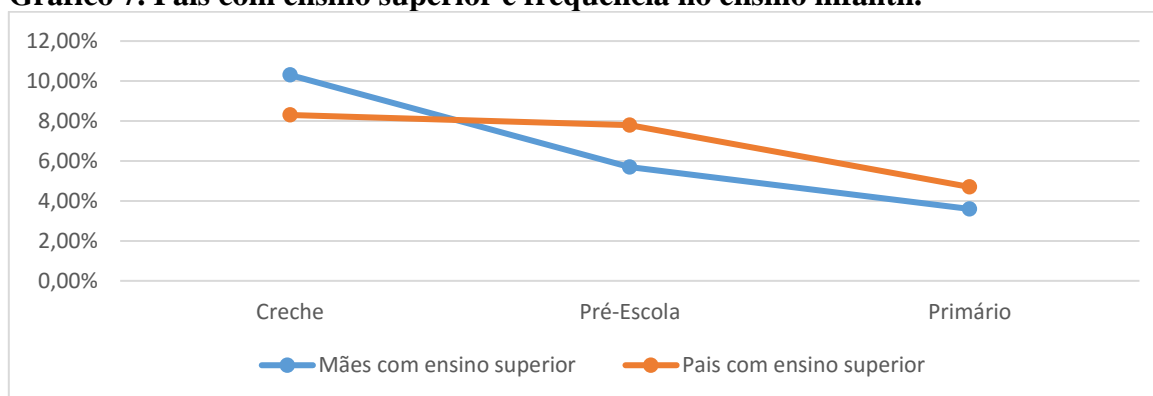


Fonte: elaboração própria a partir dos dados do Prova Brasil 2013 e SENNA.

No gráfico 6 é possível observar que os pais com maiores taxas de interesse pela educação de seus filhos e o demonstram em ações específicas como comparecer às reuniões parecem ter colocado seus filhos na escola mais cedo. As taxas de interesse para variáveis de incentivo, como incentivar os filhos a ler e estudar, por exemplo, não disponível no gráfico, são bastante semelhantes entre os grupos que iniciaram no ensino infantil ou diretamente no primário. Essa separação entre características de interesse de discurso (incentivar os filhos) e de exemplo (ações específicas) ficou ainda mais evidente após a estimação do índice através de análise de componentes principais policórica. As variáveis de incentivo ou discurso, ficaram todas reunidas no primeiro fator, enquanto o segundo resultou cargas fatoriais maiores para as variáveis de exemplo ou ação dos pais. No entanto apenas o primeiro fator resultou um autovalor maior que 1, de 3,47, enquanto o segundo, resultou um autovalor de 0,99.

O gráfico 7 apresenta o percentual de mães e pais com ensino superior segundo a seleção no ensino infantil.

**Gráfico 7. Pais com ensino superior e frequência no ensino infantil.**



Fonte: elaboração própria a partir dos dados do Prova Brasil 2013 e SENNA.

De acordo com o gráfico 7 existe uma diferença visível de pais com ensino superior completo entre os grupos que escolheram colocar seus filhos no ensino infantil ou diretamente no primário. Um teste de médias indicou que essa diferença entre os grupos é estatisticamente significativa a 1%. Desse modo, a escolaridade dos pais, tendo como *proxy* a mãe ter ensino superior completo, foi incluída no modelo de escolha.

A tabela 1 apresenta os resultados da estimação de um modelo de escolha binária *Logit* para a creche e pré-escola. Esse modelo estima a probabilidade de um determinado indivíduo ser selecionado para o tratamento (ensino infantil) segundo variáveis observáveis. Assim, os grupos tratamento serão dois, ter feito creche e ter feito pré-escola. Para as estimações da creche, terão dois grupos de controle: ter iniciado na pré-escola e ter iniciado no primário. A tabela 1 apresenta os resultados levando em consideração os três diferentes grupos de controle.

**Tabela 1: Modelo de escolha binária *Logit* para Creche e Pré-Escola.**

Grupos de controle/ Variáveis de Seleção	Creche		Pré-Escola
	Primário	Pré-escola	Primário
Índice de interesse dos pais – 1	0.28* (9.41)	0.19* (8.87)	0.08* (3.09)
Mãe Ensino Superior	0.95* (7.62)	0.63* (9.33)	0.28* (2.18)
Empregada	-0.13*** (-1.76)	-0.10** (-2.12)	0.03 (0.41)
Número de observações	7.767	12.815	8226

Fonte: STATA 13. Elaboração própria. \* significativa a 1%. \*\* significativa a 5%. \*\*\*significante a 10%

Os resultados estimados não rejeitam as hipóteses levantadas no modelo teórico. As habilidades dos pais  $\theta^P$  representadas pela variável binária mãe possui ensino superior completo, apresentou uma probabilidade positiva e estatisticamente significativa, de 95% na decisão dos pais entre colocar a criança na creche (e não diretamente no primário). Essa probabilidade é de 63% quando a escolha é entre a creche e a pré-escola, e de 28% quando a escolha é entre a pré-escola e o primário. Todas positivas e estatisticamente significantes a 1%.

A variável  $\lambda^P$  representada pelo índice de interesse dos pais resultou positiva e significativa a 1% na probabilidade de influenciar a escolha do ensino infantil, especialmente para a creche em comparação com o primário, com probabilidade de 28%. Já a variável  $\gamma^k_I$  representando o preço do investimento em  $k$  no período  $I$ , empiricamente representada pela variável binária ter empregada, é negativa e significativa apenas para a creche. Famílias que possuem empregada têm uma probabilidade de 13% de não colocar seus filhos na creche (negativo e significativo a 10%) com relação ao primário e 10% com relação ao pré (negativo e significativo a 5%).

#### 4. Resultados das estimações dos modelos empíricos e análise dos resultados

Segundo Crump, Hotz, Imbens e Mitinik (2009, 2008) o estimador de pareamento é muito sensível à escolha das variáveis  $X$ . Desse modo, um número muito grande de variáveis explicativas pode invalidar a hipótese de sobreposição, e um número muito pequeno pode violar a hipótese de ignorabilidade. Seguindo Rosenbaum e Rubin (1983) será feita

comparação de teste de médias de cada uma das variáveis observáveis no grupo de tratamento e no grupo de seus pares. Se a diferença nessas médias for grande, o teste de diferença de médias irá rejeitar a hipótese de que os grupos de tratamento e o grupo de pares são balanceados em relação a estas variáveis observáveis.

A tabela 2 apresenta os resultados da especificação do modelo *probit* que conseguiu satisfazer a hipótese de balanceamento para todas as estimativas, creche e pré-escola e seus respectivos grupos de tratamento e controle. A tabela também apresenta o número de controles e tratados para cada estimação. Todas as estimativas satisfizeram a hipótese de que em média, os *propensity scores* não são diferentes para tratados e controles, através do número de blocos utilizados.

**Tabela 2: Modelo de escolha binária *Probit* para Creche e Pré-Escola após a condição de balanceamento.**

Grupos de controle/ Variáveis de Seleção	Creche		Pré-Escola
	Primário	Pré-escola	Primário
Índice de interesse dos pais – 1	0.17* (9.31)	0.11* (7.70)	-
Índice de interesse dos pais - 2	-	-	0.12* (4.99)
Mãe lê para o filho	-	0.04 (0.97)	0.09*** (1.89)
Mãe Ensino Superior	0.52* (8.01)	0.39* (9.34)	0.15** (2.03)
Empregada	-0.07*** (-1.68)	-0.07** (-2.16)	-
Número de Tratados	5.944	5.944	6.439
Número de Controles	1.823	6.871	1.838
Total:	7.767	12.815	8.277

Fonte: STATA 13. Elaboração própria. \* significante a 1%. \*\* significante a 5%. \*\*\*significante a 10%

No segundo modelo, Mínimos Quadrados Ponderados pelos Escores de Propensão, entram as variáveis controlando desempenho dos alunos e as competências socioemocionais. Segundo Soares e Alves (2003) existe um grande hiato entre alunos brancos e negros em relação ao desempenho escolar, e de acordo com Soares e Collares (2006) fatores culturais como o envolvimento dos pais com a educação de seus filhos estão também fortemente associados ao desempenho dos mesmos. Segundo Menezes Filho (2007) alunos que estão atrasados (repetiram ano), alunos de escola pública, que trabalham fora ou não moram com um dos pais, têm um desempenho pior.

As regressões foram estimadas ao nível do aluno, mas com erros clusterizados por escolas, assumindo-se que as observações dentro de cada escola são correlacionadas. Além do efeito médio do tratamento nos tratados para a creche e para pré-escola, em português e matemática no 5º e 9º anos do ensino fundamental, esta seção estima também o efeito do ensino infantil sobre as competências socioemocionais. É importante ressaltar que a amostra SENNA/PB é apenas representativa para escolas da rede estadual do Rio de Janeiro.

A tabela 3 apresenta os resultados da regressão para o impacto do ensino infantil (creche e pré-escola) para a amostra SENNA/PB.

**Tabela 3: Resultados das regressões para creche e pré-escola – Habilidades Cognitivas – SENNA/PB.**

Trat.	Creche						Pré-Escola		
Ctrl	Pré-Escola			Primário			Primário		
Método	MQP	ATTS	FMM <sup>14</sup>	MQP	FMM*	ATTS	MPQ	FMM*	ATTS
5º Ano Português	13.02 (1.30)	-12.03* (-3.58)	A: -2.04 (-0.39) B: -0.53 (-0.04)	29.95** (2.63)	A: 23.27* (7.30) B: 36.36* (5.04)	23.49* (5.67)	35.32* (3.66)	A: 29.30* (4.82) B: 44.40* (7.32)	49.41* (10.58)
5º Ano Matemática	19.27** (2.28)	-4.32 (-1.31)	A: 8.54** (2.18) B: 13.76*** (1.94)	31.31* (3.63)	A: 17.17* (4.03) B: 58.77* (5.39)	30.53* (7.87)	28.41* (4.05)	A: 21.31* (3.23) B: 36.47* (7.36)	44.01* (10.70)
Nº de observações – 5º ano	<b>704</b>	<b>1047</b>	<b>827</b>	<b>601</b>	<b>711</b>	<b>740</b>	<b>439</b>	<b>508</b>	<b>726</b>
9º Ano Português	-2.55 (-1.17)	-4.53* (-4.78)	-	6.47* (2.92)	-	6.51* (4.45)	10.26* (6.04)	-	13.51* (9.87)
9º Ano Matemática	0.39 (0.21)	-2.64* (-2.94)	-	8.07* (3.45)	-	8.63* (6.33)	9.83* (4.21)	-	13.30* (10.50)
Nº de observações – 9º ano	<b>9.714</b>	<b>11.996</b>	-	<b>5.922</b>	-	<b>7.256</b>	<b>6.217</b>	-	<b>7.780</b>

Fonte: STATA 13. Elaboração própria. Notas: A estatística *t* está entre parênteses. \* Variáveis significantes a 1%. \*\* Variáveis significantes a 5%. \*\*\* Variáveis significantes a 10%.

Para a estimação do efeito da pré-escola para os dados do 5º Ano do ensino fundamental, o impacto da pré-escola nas notas de português e matemática, foram todos positivos e significativos a 1%, resultando em 35,32 pontos na nota de português e 28,41 pontos em matemática para a regressão MQP. A coluna ATTS apresenta o efeito do tratamento nos tratados, com um impacto de 49,41 pontos para as notas de português e 44,01 pontos nas notas de matemática. No entanto, essa estimativa não controla as variáveis relacionadas ao desempenho. Partindo da hipótese identificada nos gráficos Kernel da seção 3, as estimativas FMM captam esse resultado para cada componente, identificados na tabela A e B. A metodologia FMM controla a heterogeneidade não observada que faz com que a distribuição das notas seja bimodal para o quinto ano. Assim, para o componente A, com 44% de probabilidade na distribuição, identificou-se um impacto de 29,30 pontos em português e 21,31 em matemática. Para o segundo componente, com uma probabilidade de 56%, foi identificado o impacto mais alto do exercício, 44,4 pontos em português e 36,47 pontos em matemática.

Um efeito de 44,4 pontos na nota de português, para um desvio-padrão de 38,06 para a distribuição do componente B, significa que um indivíduo na mediana da distribuição

<sup>14</sup> Na estimação por Misturas Finitas (FMM) dada a hipótese de bimodalidade, apresenta os resultados para dois componentes, na tabela como resultado A e resultado B.

ultrapassaria 12% dos alunos em sua nota de português naquele componente, apenas por ter frequentado a pré-escola. Um aluno na média da distribuição, com aproximadamente 185 pontos em português estaria no nível 3, nível em que ele provavelmente seria capaz de localizar informação explícita em contos e reportagens, em propagandas com ou sem apoio de recursos gráficos, reconhecer relação de causa e consequência em poemas, contos e tirinhas, inferir o sentido de palavra, o sentido de expressão ou o assunto em cartas, contos, tirinhas e histórias em quadrinhos com o apoio de linguagem verbal e não verbal.

Tendo feito a pré-escola, com mais 44 pontos na escala SAEB, esse aluno teria uma média de 229 pontos, o que o colocaria no nível 5, ou seja, subiria dois níveis da escala SAEB<sup>15</sup>, em que ele provavelmente seria capaz, além daquelas capacidades, de identificar assunto e opinião em reportagens, contos e cartas, informação explícita em letras de música e contos, reconhecer sentido de conjunções e locuções adverbiais em verbetes, lendas e contos, reconhecer a finalidade de reportagens e cartazes, reconhecer relação de causa em consequência e entre pronome e seu referente em tirinhas, contos e reportagens, inferir elementos de narrativa em fábulas, contos e cartas, inferir finalidade e efeito de sentido decorrente do uso de pontuação e assunto em fábulas, diferenciar opinião de fato em reportagens e etc.

Para o efeito da creche no 5º ano do ensino fundamental, as estimações apresentaram um impacto positivo e significativo quando o grupo de comparação é ter feito o primário. No entanto, quando o grupo de controle é ter feito a pré-escola, os resultados são positivos e significativos apenas nas notas de matemática<sup>16</sup>.

Para os alunos do 9º ano, os impactos da pré-escola são positivos e estatisticamente significantes para todas as estimativas, com uma média de 12 pontos na escala SAEB. As estimações para o efeito da creche resultaram positivas e estatisticamente significantes quando o grupo de controle são os alunos que frequentaram o primário. Em comparação com a pré-escola, o impacto da creche é negativo e estatisticamente significativo quando se observa apenas o efeito do tratamento nos tratados, sem controlar as variáveis de desempenho. Quando estas estão devidamente controladas, como é o caso do modelo por mínimos quadrados ponderados pelos escores de propensão, o efeito da creche resulta estatisticamente significativo.

A tabela 4 apresenta os resultados das estimações de impacto da frequência no ensino infantil sobre as competências socioemocionais:

**Tabela 4: Resultados das regressões para a creche e pré-escola – Competências Socioemocionais – SENNA/PB – Especificação com maior número de observações.**

Competências Socioemocionais/ Tratamento	Creche				Pré-Escola	
	Pré-Escola		Primário		Primário	
	5º Ano	9º Ano	5º Ano	9º Ano	5º Ano	9º Ano
Conscienciosidade	0.25	-0.19	1.01	0.44	1.14	0.48

<sup>15</sup> Cada nível possui um intervalo de 25 pontos.

<sup>16</sup> Apenas quando controlados os efeitos sobre desempenho.



	(0.26)	(-0.53)	(0.71)	(0.91)	(0.87)	(1.00)
Extroversão	-0.30 (-0.46)	-0.24 (-0.96)	0.96 (1.06)	-0.22 (-0.53)	0.79 (0.74)	0.01 (0.02)
Estabilidade Emocional	-0.91 (-0.85)	-0.06 (-0.19)	0.80 (0.82)	0.46 (0.70)	1.94*** (1.70)	0.43 (0.80)
Locus de Controle	-0.42 (-0.83)	-0.23 (-1.06)	0.75 (1.24)	-0.06 (-0.22)	1.32*** (1.70)	0.09 (0.41)
Amabilidade	-0.22 (-0.29)	-0.12 (-0.48)	0.54 (0.56)	0.24 (0.57)	0.85 (0.79)	0.32 (0.83)
Abertura a Novas Experiências	-0.21 (-0.14)	-0.34 (-1.01)	2.53 (1.44)	0.69 (1.25)	2.61*** (1.67)	0.93*** (1.64)
Número mínimo de observações	981	11.359	680	6.762	677	7.255

Fonte: STATA 13. Elaboração própria. Notas: A estatística *t* está entre parênteses; \*Variáveis significantes a 1%. \*\* Variáveis significantes a 5%.

Os resultados da tabela 4 evidenciam que apenas a pré-escola contribui na formação e desenvolvimento das competências socioemocionais. Crianças que frequentaram a pré-escola possuem maior estabilidade emocional com relação às que entraram diretamente no primário, e esse resultado é estatisticamente significativo a 10%. Crianças que frequentaram a pré-escola possuem melhor autoestima, evidência no resultado da variável locus de controle, também para o quinto ano. Crianças que frequentaram a pré-escola são também mais abertas a novas experiências e esse efeito se verifica na etapa do quinto e do nono anos do ensino fundamental.

## 5. Considerações finais

Este trabalho teve como objetivo estimar o impacto da frequência no ensino infantil sobre as habilidades cognitivas e capacidades socioemocionais dos alunos no ensino fundamental para uma amostra de escolas da rede estadual do Rio de Janeiro. Utilizou-se a metodologia de pareamento *Propensity Score Matching*, para controlar a seleção por variáveis observáveis. Utilizou-se o método dos Mínimos Quadrados Ponderados pelos Escores de Propensão para controlar variáveis com relação ao desempenho dos alunos. E por fim um modelo de misturas finitas (FMM) com o objetivo de controlar heterogeneidade não observada, justificado pela evidência de bimodalidade da distribuição da variável resultado para os alunos do quinto ano.

Encontrou-se efeitos positivos e significativos da pré-escola nas notas de português e matemática em todas as estimativas MQP, FMM e ATTS para todos os anos avaliados, 5º e 9º anos do ensino fundamental. Encontrou-se também efeitos positivos e significativos para a creche nas notas de matemática no quinto ano do ensino fundamental quando os alunos que frequentaram a pré-escola são o grupo de controle, e positivos e estatisticamente significantes para todas as estimativas em que os alunos que iniciaram diretamente no primário são o grupo de controle. Partindo da hipótese identificada nos gráficos *Kernel* da seção 3, as estimativas FMM, controlando heterogeneidade não observada resultaram para o componente A, com 44% de probabilidade na distribuição, um impacto de 29,30 pontos em português e 21,31 em matemática. Para o componente B, com uma probabilidade de 56%, foi identificado o impacto mais alto do exercício, 44,4 pontos em português e 36,47 pontos em matemática.

Um efeito de 44,4 pontos na nota de português, para um desvio-padrão de 38,06 para a distribuição do componente B, significa que um indivíduo na mediana da distribuição ultrapassaria 12% dos alunos em sua nota de português naquele componente, apenas por ter frequentado a pré-escola. Um aluno na média da distribuição, com aproximadamente 185 pontos em português estaria no nível 3. Tendo feito a pré-escola, com mais 44 pontos na escala SAEB, esse aluno teria uma média de 229 pontos, o que o colocaria no nível 5, ou seja, subiria dois níveis da escala SAEB.

Para as capacidades socioemocionais, alguns resultados estimados foram positivos e estatisticamente significante a 10% para as crianças que frequentaram a pré-escola. Estas crianças possuem maior estabilidade emocional com relação às que entraram diretamente no primário, melhor autoestima, evidência no resultado da variável *locus de controle* (ambos para o quinto ano) e são também mais abertas a novas experiências (efeito para as etapas do quinto e do nono anos do ensino fundamental).

## **Referências**

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

Borghans, L., Duckworth, A. L., Heckman, J. J., & ter Weel, B. (2008). The economics and psychology of personality traits. *IZA Discussion Papers*, No. 3333.

Bowles, S., Gintis, H. (2002). The Inheritance of Inequality. *Journal of Economic Perspectives*, 16(3).

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

Caspi, A., Williams, B., Kim-Cohen, J., Craig, I.W., Milne, B.J., Poulton, R., et al. (2007). Moderation of breastfeeding effects on the IQ by genetic variation in fatty acid metabolism. *Proceedings of the National Academy of Sciences, USA*, 104, 18860–18865.

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

Cunha, F., Heckman, J. J., Schennach, S. (2010) Estimating the technology of cognitive and noncognitive skill formation. *National Bureau of Economic Research*.

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

Farkas, G. (2003). Cognitive Skills and Noncognitive Traits and Behaviors in Stratification Processes. *Annual Review of Sociology*, vol. 29, pp. 541-562.

Goldberg, L.R., Norman, W.T., and Schwartz, E. (1980). The comparative validity of questionnaire data and objective test data in predicting five peer rating criteria. *Applied Psychological Measurement*, 4, 183–194.

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

Heckman, J. (2006). Skill Formation and the Economics of Investing in Disadvantaged Children. *Science* 30.

Heckman, James. (2008). The case for investing in disadvantaged young children. In: *First Focus* (Ed.) *Big ideas for children: Investing in our nation's future*, p. 49-58. Washington, DC.

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

Jencks, C. (1979). *Who gets ahead?: The determinants of economic success in America*. New York, NY: Basic Books.

Lleras, C. (2008). Do skills and behaviors in high school matter? The contribution of noncognitive factors in explaining differences in educational attainment and earnings. *Social Science Research*, 37(3), 888-902.

McLachlan, G. J., Peel, D. (2000). *Finite Mixture Models*. Wiley Series in Probability and Mathematical Statistics. John Wiley & Sons.

Menezes Filho, N. (2007). *Os determinantes do desempenho escolar do Brasil*. São Paulo: IBMEC.

Pinto, C. C. X. (2012). Pareamento. In: Menezes Filho, N. (org.). *Avaliação econômica de projetos sociais*. 1 ed. São Paulo: Dinâmica, p. 85-105.

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

Rosenblatt, M. (1956) Remarks on some nonparametric estimates of a density function, *Ann. Math. Statist.* 27 832–835.

Santos, D. D. (2014) *A Importância Socioeconômica das Características de Personalidade*. Instituto de Personalidade.

Soares, J. F.; Alves, M. T. G. (2003). Desigualdades raciais no sistema brasileiro de educação básica. *Educação e Pesquisa*, São Paulo, v. 29, n. 1, p. 147-165.

Soares, J. F.; Collares, A. C. M. (2006) Recursos familiares e o desempenho cognitivo dos alunos do ensino básico brasileiro. DADOS: Revista de Ciências Sociais, Rio de Janeiro, v. 49, n. 3, p. 615-481.

Turkheimer, E., Haley, A., Waldron, M. D'Onofrio, B. e Gottesman, I. I. (2003). Socioeconomic Status Modifies Heritability of IQ in Young Children. Psychological Science 14(6): 623-628.