

**Área de Submissão: Métodos Quantitativos**

**Os Determinantes do Desempenho Escolar no Brasil: Novas Evidências Empíricas a partir de Uma Abordagem Não Paramétrica**

**Ana Cláudia Anegues da Silva<sup>△</sup>**

*Doutoranda em Economia Aplicada – Programa de Pós-Graduação em Economia – PPGE/UFRGS.*

Avenida Bento Gonçalves, 1615, Apartamento 505, Partenon

E-mail: anegues.ana@gmail.com

CEP 90650-002 – Porto Alegre/RS - Brasil

Tel: + 55 51 81872589

**Wallace Patrick Santos de Farias Souza<sup>◇</sup>**

*Doutorando em Economia Aplicada – Programa de Pós-Graduação em Economia – PPGE/UFRGS.*

**Victor Rodrigues de Oliveira<sup>§</sup>**

*Doutorando em Economia Aplicada – Programa de Pós-Graduação em Economia – PPGE/UFRGS.*

---

<sup>△</sup> E-mail: anegues.ana@gmail.com

<sup>◇</sup> E-mail: wpsfarias@gmail.com

<sup>§</sup> E-mail: victor5491@gmail.com

## Área de Submissão: Métodos Quantitativos

### Os Determinantes do Desempenho Escolar no Brasil: novas Evidências Empíricas a partir de Uma Abordagem Não Paramétrica

*Resumo:* O objetivo deste trabalho consiste em investigar o impacto de atributos escolares, da origem familiar e de características do aluno em sua proficiência escolar. Para descrever a relação entre as variáveis optou-se pelo modelo não paramétrico *Generalized Additive Model (GAM)*, que possibilita a inclusão de um maior número de variáveis explicativas, tal como descrito em Horowitz e Mammen (2004). Empregou-se ainda um teste de especificação para atestar a robustez do modelo não paramétrico frente às abordagens paramétricas tradicionais. A inferência foi realizada por um estimador não paramétrico em dois estágios, o qual evita problemas de dimensionalidade à medida que se aumenta a ordem da matriz de covariáveis. Utilizaram-se os microdados do SAEB 2011 e do Censo Escolar, onde as proficiências correspondem às notas do exame de aprendizagem de alunos do 5<sup>a</sup> e do 9<sup>a</sup> ano do ensino fundamental de escolas das redes pública e privada. Os resultados encontrados mostram que as variáveis ligadas à família e ao aluno tiveram maior influência no desempenho deste relativamente às variáveis da escola, dos docentes e dos diretores, as quais em sua maioria não foram estatisticamente significantes para os estudantes da rede privada.

*Palavras-chave:* Desempenho Escolar, Modelo GAM, Estimador Não Paramétrico

*Abstract:* The objective of this study is to investigate the impact of school attributes, family origin and characteristics of the students on their school proficiency. To describe the relationship between the variables we chose the nonparametric model Generalized Additive Model (GAM), which enables the inclusion of a larger number of explanatory variables, as described in Horowitz and Mammen (2004). It was used also a test specification to prove the robustness of the nonparametric model against traditional parametric approaches. The inference was performed by a non-parametric estimator in two stages, which avoids problems of dimensional increases as the order of the matrix covariates increases. We used the SAEB 2011 and the School Census microdata, where proficiency corresponds to the learning exam notes of students in 5th and 9th grade of elementary education schools in the public and private systems. The results show that the variables related to the family and the student had a greater influence on his performance in relation to variables of school, teachers and schoolmaster, which mostly were not statistically significant for students from private schools.

*Keywords:* School Performance, GAM Model, Non Parametric Estimator

*JEL Classification:* I20, I25, C14.

## 1 Introdução

Embora o Brasil venha apresentando melhorias em seus indicadores educacionais (taxa de analfabetismo, número de crianças matriculadas), em termos de desempenho escolar o país ainda permanece em situação desfavorável quando comparado a outros países. Segundo a última avaliação feita pelo *Programme for International Student Assessment (PISA)*, no ano de 2012, o Brasil ficou na 55ª posição, ficando abaixo de países da América Latina, como Chile e Uruguai. Esse panorama mostra que o país ainda possui um longo caminho para melhoria da qualidade de ensino, o que desperta preocupação por parte de governos e especialistas, tendo em vista que a educação consiste em um dos pontos-chaves para uma nação próspera com desenvolvimento sustentado.

Várias estratégias para reverter esse quadro têm sido sugeridas, especialmente quando se fala em políticas públicas de educação. Mas antes de tentar definir alguma que possa ser de fato eficaz, em primeiro lugar é necessário identificar os fatores que influenciam a aprendizagem e o desempenho dos alunos. Ao longo dos anos a literatura tem se preocupado em fornecer respostas a essa necessidade, investigando o impacto de fatores que caracterizam a escola, o professor, o aluno e sua origem familiar.

No primeiro grupo de fatores se destacam análises referentes ao ambiente escolar e características dos profissionais de educação, tal como o nível de qualificação do professor e do diretor. Biondi e Felício (2007) utilizam dados em painel para medir o impacto dos atributos escolares sobre o desempenho dos alunos e concluem que o tempo de experiência do docente e o sistema de eleição do diretor possuem grande influência sobre o sucesso educacional. Ainda com relação à qualificação do professor, o estudo de Machado, Firpo e Gonzaga (2013) encontram que esta variável pode reverter problemas de dispersão etária nas turmas. Além de variáveis ligadas à escola, características da vida escolar dos alunos também são consideradas, como por exemplo, seu histórico de repetências. Nesse sentido, pode-se citar o trabalho de Barros e Mendonça (1998), que analisa a probabilidade de reprovação dos estudantes condicionada à repetência, e Leon e Menezes-Filho (2002), que procuram mensurar a influência do *background* familiar sobre o avanço escolar dos alunos, através de dados em painel da Pesquisa Mensal de Emprego (PME)

A origem familiar surge como um fator fortemente relevante ao sucesso educacional. A importância das características familiares na chance de progresso escolar é um resultado bastante consolidado nas literaturas teórica e empírica, e considerado como um dos principais fatores responsáveis pela “manutenção do ciclo intergeracional de pobreza” (Leon e Menezes-Filho, 2002). Alguns estudos mostram que esta variável se mostra tão ou mais importante que as de caráter pedagógico restritas ao ambiente escolar, tal como Souza e Silva (1994), cujos resultados atestam tal importância. Franco, Mandarino e Ortigão (2002), ao verificarem o impacto de projetos pedagógicos implantados pelas escolas, chegam à conclusão de que este fator foi pouco eficaz na melhoria da qualidade de ensino, de tal sorte que a origem familiar exerceu maior influência.

Além das variáveis listadas acima, análises acerca da efetividade de políticas educacionais também vêm sendo realizadas. Menezes-Filho *et al.* (2008) debatem o impacto dos programas de progressão continuada sobre a aprendizagem e mostram que a política contribui positivamente para a permanência do aluno na escola, sem, no entanto, elevar o seu nível de desempenho. A influência de outras características como estrutura física e pedagógica da escola, hábitos de leitura de alunos e professores, região de localização da unidade escolar podem ser encontradas nos trabalhos de Barros *et al.* (2001) e Machado *et al.* (2008).

Do ponto de vista metodológico, os trabalhos citados acima se baseiam em formas funcionais específicas para descrever a relação existente entre a proficiência e as variáveis

explicativas selecionadas. Um problema que pode surgir dessa suposição é a escolha de uma forma funcional que não consiga captar a relação adequada existente entre as variáveis, gerando possivelmente um viés de especificação no modelo. Recentemente, Almeida (2014) utilizou regressões quantílicas na identificação dos determinantes do rendimento escolar, dado que a regressão por quantis não necessita de suposições acerca da distribuição dos erros, conferindo um perfil semiparamétrico ao método.

Para lidar com esse problema de especificação, optamos por investigar os determinantes da proficiência escolar através da utilização de técnicas de estimação não paramétricas. A inferência não paramétrica tem como vantagem a não imposição de restrições à distribuição dos dados, permitindo a identificação de possíveis padrões de comportamento não lineares da variável dependente condicionada aos fatores explicativos.

O modelo econométrico empregado na análise é o *Generalized Additive Model* (GAM) apresentado em Horowitz e Mammen (2004), que consiste em uma adição de funcionais desconhecidas, relacionando as variáveis explicada e explicativas. A grande vantagem deste modelo está na possibilidade da inclusão de um número maior de covariáveis, aumentando o poder de explicação sem, no entanto, ter problemas relacionados à quantidade de parâmetros mensurados. A estimação dos parâmetros do modelo se baseia no método de inferência não paramétrico em dois estágios, com propriedades que evitam problemas de dimensionalidade. Embora a modelagem não paramétrica apresente uma série de vantagens, optou-se pela realização de um teste de especificação para evitar que a escolha do método não paramétrico seja feita de maneira *ad hoc*.

Além da utilização de uma nova metodologia, esta pesquisa procurar estender o espectro a outras variáveis relevantes ao sucesso educacional do aluno, como raça e gênero de professores e diretores e análise de indicadores de violência dentro da escola. A inclusão da raça e do gênero de gestores e educadores visa identificar a ocorrência do fenômeno da “armadilha do estereótipo” (*stereotype threat*), isto é, se a identificação dos alunos com os educadores e com os gestores por esses critérios gera efeitos sobre o seu desempenho. Steele e Aronson (1995) concluíram que diferenças de gênero e raça entre professores e alunos foram relevantes ao processo de aprendizagem nos Estados Unidos.

Os bancos de dados utilizados são os microdados do SAEB 2011 e do Censo Escolar, onde se encontram sumarizadas todas as informações referentes às variáveis das escolas, dos professores, dos pais e dos alunos, bem como as notas destes na avaliação de proficiência em língua portuguesa e matemática. Dado que a habilidade dos estudantes não é diretamente mensurável, comumente se utiliza como *proxy* dessa variável o desempenho em exames de verificação de aprendizagem, como o SAEB.

O artigo contém mais três seções, além desta introdução. A segunda seção traz uma descrição do modelo não paramétrico GAM e do método de estimação em dois estágios. Em seguida, na terceira seção, são apresentados os resultados do estudo, além de uma breve análise descritiva das variáveis consideradas. A quarta e última seção contém as considerações finais.

## **2 Procedimentos Metodológicos**

Esta seção apresenta a estratégia empírica adotada no estudo. Em primeiro lugar serão destacados alguns conceitos e características do modelo econométrico não paramétrico utilizado, o *Generalized Additive Model* (GAM). Em seguida, é descrito o método de inferência pelo estimador de dois estágios adotado na mensuração dos coeficientes; e por fim, apresenta-se a seleção da *bandwidth* empregada nas estimativas.

## 2.1 Generalized Additive Model (GAM)

Os métodos de estimação usualmente empregados pela literatura, como as técnicas de Mínimos Quadrados, requerem que os modelos sejam lineares nos parâmetros, ou seja, baseados em formas funcionais lineares para um conjunto de covariáveis  $X$ . Entretanto, a modelagem paramétrica pode não ser adequada diante da incerteza entre a relação das variáveis explicada e explicativas.

Por outro lado, uma estimação não paramétrica é caracterizada pela não imposição de formas funcionais para a equação de regressão, de modo que se torna possível captar relações não lineares entre as variáveis, diferentemente das abordagens paramétricas tradicionais. Assim, este estudo empregará a classe de *Generalized Additive Model*, que substitui a forma linear por uma soma de funções suavizadas.

Estima-se um modelo aditivo com função *link*, ou seja, preocupado com a estimativa não paramétrica das funções  $m_1, \dots, m_d$  no modelo

$$Y = F[\mu + m_1(X^1) + \dots + m_d(X^d)] + U \quad (1)$$

em que  $X^j (j = 1, \dots, d)$  é o  $j$ -ésimo elemento do vetor aleatório  $X \in \mathbb{R}^d$  para  $d \geq 2$ ,  $\mu$  é uma constante desconhecida,  $F$  é conhecida,  $m_1, \dots, m_d$  são funções desconhecidas e  $U$  é o termo aleatório não observado satisfazendo  $E(U|X = x) = 0$  para quase todo  $x$ . O estimador de cada componente aditivo é normalmente distribuído com a mesma média e variância que teria se os outros componentes fossem conhecidos. A amostra aleatória é *i.i.d.* e o estimador dos componentes aditivos  $m_1, \dots, m_d$  converge em probabilidade à taxa  $n^{-2/5}$  quando  $F$  e  $m_j$ 's são duas vezes continuamente diferenciáveis e a segunda derivada de  $F$  é suficientemente suave. Esse estimador corrige o problema apresentado por Linton e Härdle (1996), que apresentaram um modelo onde era necessário que as funções aditivas  $m_j$ 's tivessem um número cada vez maior de derivadas à medida que a dimensão de  $X$  aumentasse.

Uma gama de trabalhos discute a estimação de funções como em (1), tais como Stone (1994), Newey (1997), utilizando estimadores *splines*, e Opsomer (2000) que investiga as propriedades dos processos de *backfitting*. As propriedades de otimalidade de uma variedade de estimadores não paramétricos para modelos aditivos sem funções de ligação são discutidas em trabalhos como o de Horowitz, Klemelä e Mammen (2002). Esses autores trabalham com uma abordagem de estimação em dois estágios similar a que será utilizada nesse estudo, porém, não consideram modelos com funções *link* e usam *backfitting* para o estimador no primeiro estágio<sup>1</sup>. No entanto, o estimador aqui apresentado, ao contrário o estimador *backfitting*, tem a vantagem de que a distribuição assintótica de cada componente aditivo é a mesma que seria caso os demais componentes fossem conhecidos.

## 2.2 Inferência

O grande número de variáveis explicativas - ou seja, a matriz de covariáveis de dimensão  $d$  suficientemente grande - gera, por sua vez, um maior número de parâmetros a serem estimados. O estimador em dois estágios utilizado pelo presente artigo não requer uma regressão não paramétrica de dimensão  $d$ , evitando, desse modo, o problema da dimensionalidade da matriz. No primeiro estágio, utiliza-se mínimos quadrados não lineares para obter uma aproximação de série para cada  $m_j$ . Será imposta a estrutura aditiva (1) e produzidas estimativas das  $m_j$ 's que têm viés assintótico menor que outros procedimentos, tal

---

<sup>1</sup> Para mais detalhes sobre o procedimento *backfitting* e suas propriedades ver Hastie e Tibshirani (1990), Opsomer e Ruppert (1997), Mammen, Linton e Nielsen (1999) e Opsomer (2000).

como os avaliados por integração marginal. Já no segundo estágio é obtida uma estimativa linear local e derivada, assim como a sua taxa pontual de convergência e distribuição assintótica.

Assuma que o suporte de  $X$  é  $X \equiv [-1,1]^d$  e normalize  $m_1, \dots, m_d$  tal que

$$\int_1^{-1} m_j(v) dv = 0, \quad j = 1, \dots, d \quad (2)$$

A função densidade de probabilidade de  $X$  é delimitada, com uma distância entre zero e duas vezes continuamente diferenciável em  $X$ . Essa suposição garante que o viés do estimador convirja para zero suficientemente rápido.

Para qualquer  $X \in \mathbb{R}^d$  é definido  $m(x) = m_1(X^1) + \dots + m_d(X^d)$ , onde  $x^j$  é o  $j$ -ésimo componente de  $x$ . Deixe  $\{p_k: k: 1, 2, \dots\}$  denotar uma base de funções “suaves” em  $[-1,1]$ , que devem satisfazer as seguintes condições

$$\int_1^{-1} p_k(v) dv = 0 \quad (3)$$

$$\int_1^{-1} p_j(v) p_k(v) dv = \begin{cases} 1 & \text{se } j = k \\ 0 & \text{c. c} \end{cases} \quad (4)$$

$$m_j(x^j) = \sum_{k=1}^{\infty} \theta_{jk} p_k(x^j) \quad (5)$$

para cada  $j = 1, \dots, d$ . Cada  $x^j \in [0,1]$ , os coeficientes adequados são  $\{\theta_{jk}\}$  e para qualquer  $k$  inteiro positivo é definido

$$p_k(x) = [1, p_1(x^1), \dots, p_k(x^1), p_1(x^2), \dots, p_k(x^2), p_1(x^d), \dots, p_k(x^d)] \quad (6)$$

em que  $p_k(x)' \theta_k$  é uma aproximação para a série  $\mu + m(x)$ .

Para obter os estimadores do primeiro estágio dos  $m_j$ 's para uma amostra aleatória  $(Y, X)$ , deixe  $\hat{\theta}_{nk}$  ser a solução para

$$\min_{\theta \in \Theta_k} S_{nk}(\theta) \equiv n^{-1} \sum_{i=1}^n \{Y_i - F[P_k(X_i)' \theta]\}^2 \quad (7)$$

em que  $\Theta_k \subset \mathbb{R}^{d+1}$  é um conjunto de parâmetros compacto. Assim, o estimador da série  $\mu + m(x)$  é

$$\tilde{\mu} + \tilde{m}(x) = P_k(x)' \hat{\theta}_{nk} \quad (8)$$

em que  $\tilde{\mu}$  é o primeiro componente de  $\hat{\theta}_{nk}$ .

Para obter o estimador de segundo estágio de, por exemplo,  $m_1(X^1)$ , deixe  $\tilde{X}_i$  denotar a  $i$ -ésima observação de  $\tilde{X} \equiv (X^2, \dots, X^d)$ . Defina  $\tilde{m}_{-1}(\tilde{X}_i) = \tilde{m}_2(X_i^2) + \dots + \tilde{m}_d(X_i^d)$ , em que onde  $X_i^j$  é a  $i$ -ésima observação do  $j$ -ésimo componente de  $X$  e  $\tilde{m}_j$  é o estimador da série de  $m_j$ .

Seja  $k$  uma função densidade de probabilidade em  $[-1,1]$ , e defina  $K_h(v) = K(v/h)$  para qualquer constante real positiva  $h$ . Seja

$$S'_{nj1}(x^1, \tilde{m}) = -2 \sum_{i=1}^n \{Y_i - F[\tilde{\mu} + \tilde{m}_1(x^1) + \tilde{m}_{-1}(\tilde{X}_i)]\} \times \\ \times F'[\tilde{\mu} + \tilde{m}_1(x^1) + \tilde{m}_{-1}(\tilde{X}_i)](X_i^1 - x^1)^j K_h(x^1 - X_i^1) \quad (9)$$

para  $j=0, 1$  e

$$S'_{nj1}(x^1, \tilde{m}) = 2 \sum_{i=1}^n F'[\tilde{\mu} + \tilde{m}_1(x^1) + \tilde{m}_{-1}(\tilde{X}_i)]^2 (X_i^1 - x^1)^j K_h(x^1 - X_i^1) \\ - 2 \sum_{i=1}^n \{Y_i - F[\tilde{\mu} + \tilde{m}_1(x^1) + \tilde{m}_{-1}(\tilde{X}_i)]\} \times F''[\tilde{\mu} + \tilde{m}_1(x^1) \\ + \tilde{m}_{-1}(\tilde{X}_i)](X_i^1 - x^1)^j K_h(x^1 - X_i^1) \quad (10)$$

para  $j = 0, 1, 2$ . O estimador de segundo estágio de  $m_1(X^1)$  é

$$\hat{m}_1(x^1) = \tilde{m}_1(x^1) - \frac{S''_{n21}(x^1, \tilde{m})S'_{n01}(x^1, \tilde{m}) - S''_{n11}(x^1, \tilde{m})S'_{n11}(x^1, \tilde{m})}{S'_{n01}(x^1, \tilde{m})S''_{n21}(x^1, \tilde{m}) - S''_{n11}(x^1, \tilde{m})} \quad (11)$$

Os estimadores para  $m_2(X^2), \dots, m_d(X^d)$  são obtidos de maneira semelhante.

A taxa prevista de convergência de  $h$  é bem conhecida por ser assintoticamente ideal para a regressão média *kernel* unidimensional, quando a função condicional média é duas vezes continuamente diferenciável. A taxa necessária para que  $\kappa$  assegure o viés assintótico e a variância do estimador no primeiro estágio é suficientemente pequena para alcançar uma taxa de convergência de  $n^{-2/5}$  no segundo estágio.

Assegura-se ainda a existência e não singularidade da matriz de covariância da forma assintótica do estimador de primeiro estágio. Isso é análogo a assumir que a matriz de informação é definida positiva na estimativa de máxima verossimilhança paramétrica.

### 2.3 Seleção da *bandwidth*

Nesta seção apresenta-se o método proposto para a escolha do parâmetro de suavização. A definição da *bandwidth* ótima se dará por meio do método *plug-in*. Este método estima o valor de  $h$  que minimiza o erro quadrático médio assintótico integrado (AMISE) de  $n^{-2/5}[\hat{m}_1(x^1) - m_1(x^1)]$  para  $j = 1, \dots, d$ . Defina as seguintes funções auxiliares:

$$D_0(x^1) = 2 \int F'[\mu + m_1(x^1) + m_{-1}(\tilde{x})]^2 f_X(x^1, \tilde{x}) d\tilde{x} \quad (12)$$

$$D_1(x^1) = 2 \int F'[\mu + m_1(x^1) + m_{-1}(\tilde{x})]^2 \left[ \frac{\partial f_X(x^1, \tilde{x})}{\partial x^1} \right] d\tilde{x} \quad (13)$$

$$A_K = \int_{-1}^1 v^2 K(v) dv \quad (14)$$

$$B_K = \int_{-1}^1 K(v)^2 dv \quad (15)$$

$$g(x^1, \tilde{x}) = F''[\mu + m_1(x^1) + m_{-1}(\tilde{x})]m_1'(x^1) + F'[\mu + m_1(x^1) + m_{-1}(\tilde{x})]m_1''(x^1) \quad (16)$$

$$\beta_1(x^1) = 2C_h^2 A_K D_0(x^1)^{-1} \times \int g(x^1, \tilde{x}) F'[\mu + m_1(x^1) + m_{-1}(\tilde{x})] f_X(x^1, \tilde{x}) d\tilde{x} \quad (17)$$

$$V_1(x^1) = B_K C_h^{-1} D_0(x^1)^{-2} \times \int \text{Var}(U|x^1, \tilde{x}) F'[\mu + m_1(x^1) + m_{-1}(\tilde{x})]^2 f_X(x^1, \tilde{x}) d\tilde{x} \quad (18)$$

O AMISE de  $n^{-2/5}[\hat{m}_1 - m_1]$  é definido como

$$AMISE_1 = n^{4/5} \int_{-1}^1 w(x^1) [\beta_1(x^1)^2 + V_1(x^1)] dx^1 \quad (19)$$

em que  $w(\cdot)$  é um função de ponderação não negativa. Define-se a *bandwidth* ótima assintoticamente para a estimação de  $m_1$  como  $C_{h1} n^{-1/5}$ , e que  $C_{h1}$  minimiza  $AMISE_1$ . Assuma que  $\tilde{\beta}_1(x^1) = \beta_1(x^1)/C_h^2$  e  $\tilde{V}_1(x^1) = C_h V_1(x^1)$ . Disto decorre que

$$C_{h1} = \left[ \frac{1 \int_{-1}^1 w(x^1) \tilde{V}_1(x^1) dx^1}{4 \int_{-1}^1 w(x^1) \tilde{\beta}_1(x^1)^2 dx^1} \right]^{1/5} \quad (20)$$

Assuma que  $L$  seja uma função densidade de probabilidade diferenciável duas vezes no intervalo  $[-1,1]$  e  $\{g_n: n = 1, 2, \dots\}$  uma sequência de números reais positivos que satisfazem:  $g_n \rightarrow 0$  e  $g_n^2 n^{4/5} (\log n)^{-1} \rightarrow \infty$ . As derivadas de  $m_j$ , para  $l = 1, 2$ , podem ser aproximadas como segue:

$$\hat{m}_1^{(l)}(x^1) = g_n^{-1-l} \int_{-1}^1 L^{(l)}[(x^1 - v)/g_n] \hat{m}_1(v) dv \quad (21)$$

Dessa forma, a escolha da *bandwidth* pelo método proposto se mostra assintoticamente eficiente.

### 3 Resultados

#### 3.1 Fonte de Dados

Para compreender os determinantes do desempenho educacional dos estudantes do ensino fundamental brasileiro foram empregados os microdados do Sistema Nacional de Avaliação da Educação Básica de 2011 (SAEB 2011) e do Censo Escolar. Para tanto, foram utilizados os dados do 5º ano e do 9º ano da proficiência nas provas de matemática e língua portuguesa dos ensinos público e privado. Para evitar oscilações decorrentes de turmas pequenas e comportamentos específicos de alguma série foram descartadas do estudo as turmas formadas por menos de cinco alunos.



O SAEB é realizado em duas etapas. Em primeiro lugar, são selecionadas escolas com seu conjunto de turmas e alunos do 5º ano e 9º ano do ensino fundamental e da 3ª série do ensino médio. Após, selecionam-se turmas em cada uma das séries dentro das escolas selecionadas. Uma vez selecionada uma turma para participar da avaliação, todos os alunos da turma faziam parte automaticamente da amostra e cada aluno presente no dia da avaliação foi submetido às provas das disciplinas de língua portuguesa e matemática. Os alunos fazem testes que priorizam a competência e a habilidade cognitiva, tendo por finalidade medir a habilidade de leitura em língua portuguesa e na resolução de problemas em matemática. As provas são elaboradas com base na matriz de habilidades. O cômputo do desempenho dos alunos é baseado nas escalas de proficiência. Estas escalas são construídas por meio da Teoria de Resposta ao Item (TRI) e do modelo de Blocos Incompletos Balanceados (BIB).

A TRI é um conjunto de modelos matemáticos que procuram representar a probabilidade de um indivíduo dar uma determinada resposta a um item como função dos parâmetros deste e da habilidade (ou habilidades) do respondente. Essa relação é sempre expressa de tal forma que quanto maior a habilidade, maior a probabilidade de acerto no item. O BIB, por sua vez, é um esquema otimizado para o rodízio de blocos. A utilização da TRI e do BIB permite, assim, a construção de uma escala para cada disciplina, englobando as três séries avaliadas e ordenando o desempenho dos alunos do nível mais baixo para o mais alto. A proficiência dos alunos corresponde a escalas específicas ao assunto, variando de 0 a 425 pontos para o ensino fundamental, permitindo avaliar as competências adquiridas pelos alunos ao longo da trajetória escolar.

O Censo Escolar, por sua vez, é um levantamento de dados estatístico-educacionais de âmbito nacional realizado todos os anos com a participação de todas as escolas públicas e privadas do país. O Censo Escolar coleta dados sobre estabelecimentos, matrículas, funções docentes, movimento e rendimento escolar.

Analisou-se a proficiência média em português e em matemática, separadamente para o 5º e o 9º ano tanto das escolas públicas quanto das escolas privadas, de modo a captar possíveis diferenças nos efeitos entre os níveis, sistemas de ensino, pelo campo de saber avaliado e também considerar as características dos professores das áreas específicas. Assim, foram construídas quatro amostras: a amostra I é constituída pelos alunos do 5º ano das escolas públicas (281.853 alunos); a amostra II, pelos alunos do 9º ano do ensino público (370.716 alunos); a amostra III, pelos alunos do 5º ano das escolas privadas (10.524 alunos); a amostra IV, pelos alunos do 9º ano do ensino privado (14.986 alunos). A Tabela 1 abaixo lista as variáveis utilizadas nas estimações, onde a variável dependente é a proficiência do aluno em português e matemática, e um conjunto de variáveis sobre o *background* familiar do aluno, a escola, o diretor e sobre os professores são usadas como variáveis explicativas.

Tabela 1: Estatísticas descritivas – Descrição das Variáveis Utilizadas

Variáveis	Descrição
<b>Características do Aluno</b>	
<i>Proficiência (variável dependente)</i>	Nota obtida na prova de português/matемática
<i>Aluno homem</i>	Sexo (binária). Feminino = 0
<i>Aluno branco</i>	Cor (binária). Não brancos = 0
<i>Idade do aluno</i>	Entre 8 e 15 anos
<i>Desvio padrão da idade na turma</i>	Dispersão de idade dentro de cada turma
<i>Índice socioeconômico</i>	Indicador de condições socioeconômicas <sup>1</sup>
<i>Pais nunca estudaram</i>	Pais analfabetos
<i>Já foi reprovado</i>	Reprovado pelo menos uma vez
<i>Iniciou estudos na pré-escola</i>	Frequentou a pré-escola
<i>Iniciou estudos na primeira série</i>	Não frequentou a pré-escola
<i>Tamanho da família</i>	Número de moradores
<i>Incentivo para estudar</i>	Os pais incentivam a estudar <sup>1</sup>
<b>Características do Professor</b>	
<i>Professor: homem</i>	Sexo (binária). Feminino = 0
<i>Professor: branco</i>	Cor (binária). Não brancos = 0
<i>Professor com mais de 10 anos de experiência</i>	Tempo de experiência
<i>Professor fez curso de atualização</i>	Capacitação
<i>Professor tem especialização, mestrado ou doutorado</i>	Possui pós-graduação
<i>Professor tem superior</i>	Possui graduação
<b>Características do Diretor</b>	
<i>Diretor: homem</i>	Sexo (binária). Feminino = 0
<i>Diretor: branco</i>	Cor (binária). Não brancos = 0
<i>Diretor tem mais de 10 anos de experiência</i>	Tempo de experiência
<i>Diretor tem curso superior</i>	Possui graduação
<b>Características da Escola e Região</b>	
<i>Qualidade da escola</i>	Infraestrutura escolar
<i>Violência externa</i>	Causada por um agente externo <sup>1</sup>
<i>Violência interna</i>	Causada por um agente interno <sup>1</sup>
<i>Prova de seleção</i>	Os alunos são selecionados para a escola
<i>Tamanho da turma</i>	Número de alunos na turma
<i>Turmas homogêneas quanto à idade</i>	Homogeneidade de idade = 1; caso contrário = 0
<i>Turmas homogêneas quanto ao rendimento escolar</i>	Homogeneidade de resultado = 1; caso contrário = 0
<i>Turmas heterogêneas quanto à idade</i>	Heterogeneidade de idade = 1; caso contrário = 0
<i>Turmas heterogêneas quanto ao rendimento escolar</i>	Heterogeneidade de resultado = 1; caso contrário = 0
<i>Diferença de sexo entre aluno e professor</i>	Mesmo sexo = 0
<i>Diferença de raça entre aluno e professor</i>	Mesma raça = 0
<i>Região Urbana</i>	Rural = 0
<i>Nordeste</i>	Nordeste = 1; caso contrário = 0
<i>Norte</i>	Norte = 1; caso contrário = 0
<i>Sul</i>	Sul = 1; caso contrário = 0
<i>Centro-Oeste</i>	Centro-oeste = 1; caso contrário = 0

Nota: <sup>1</sup> Ver Anexo A.

Fonte: Elaborado pelos autores.

### 3. 2 Estatísticas Descritivas

A estatística descritiva (média, desvio padrão, valores mínimo e máximo) das variáveis apresentadas na Tabela 1 está no Anexo B deste artigo, nas Tabelas 4 a 7, sendo a Tabela 4 correspondente ao 5º ano – Escolas Públicas, Tabela 5 ao 9º ano – Escolas Públicas, Tabela 6 ao 5º ano – Escolas Privadas e Tabela 7 ao 9º ano – Escolas Privadas.

As informações das Tabelas apontam uma média de proficiência em matemática maior do que em português nos quatro níveis analisados, ao mesmo tempo em que a comparação da escola pública com o nível de ensino privado mostra uma vantagem de rendimento para este último, tanto para o 5º quanto para o 9º ano. Tais características podem ser observadas no Gráfico 1. Neste apresenta-se a densidade estimada *kernel* da proficiência dos alunos por série e tipo de escola. Para garantir a confiabilidade dessas estimações e verificar a significância das mesmas construíram-se bandas de variabilidade (linhas tracejadas azuis) a partir de um *bootstrap* não paramétrico. Foram geradas 1000 amostras e em seguida estimaram-se os desvios-padrão das densidades *kernel*. Para os alunos da rede pública percebe-se uma assimetria à esquerda, enquanto para os alunos que estudam em escolas privadas a densidade é deslocada para a direita.

Com relação às características dos alunos, a maioria para o 5º ano são do sexo masculino (52,5% na escola pública e 51,5% na escola privada), com 10,9 anos de média de idade e um grande percentual de alunos não brancos. Já para o 9º ano, 46% são do sexo masculino na escola pública e 48% na escola privada, enquanto que o percentual de alunos de cor branca aumenta em relação aos estudantes do 5º ano.

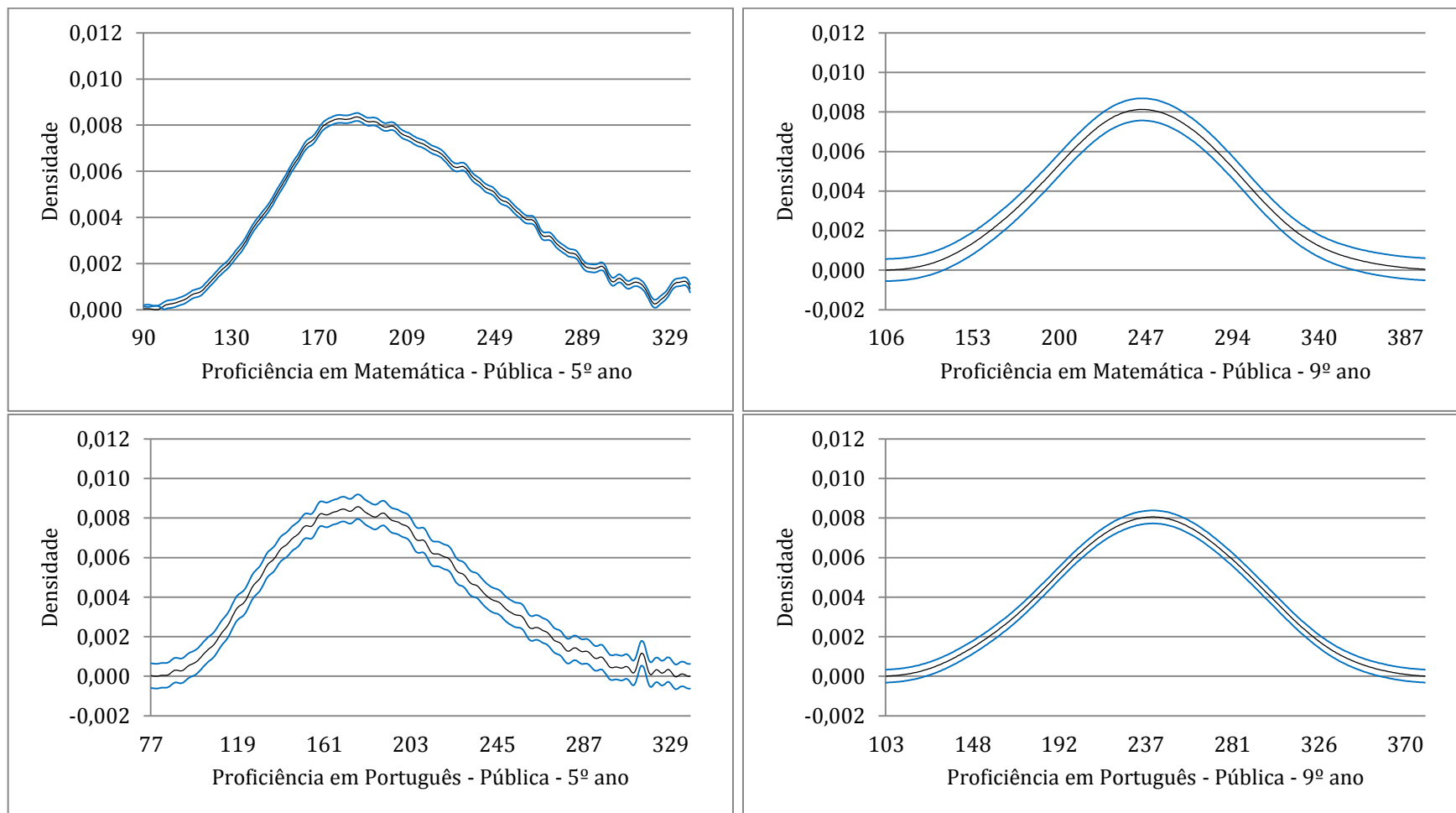
Para a rede particular de ensino, a proporção de alunos que tem pais que nunca estudaram é praticamente zero, enquanto que para a rede pública essa proporção não ultrapassa 3%. Outro aspecto que chama atenção é o percentual de alunos reprovados, com números acima de 30% para a escola pública e entre 9% e 12% para a escola privada, no 5º e no 9º ano, respectivamente.

No que concerne os professores, em todos os níveis a grande maioria possui curso superior, já fez algum curso de atualização ou possui 10 ou mais anos de experiência em sala de aula. Com relação ao gênero, menos de 30% são do sexo masculino, e com relação à cor a maioria são não brancos, a exceção do 9º ano do ensino privado.

As variáveis relacionadas ao diretor também evidenciam um maior percentual com nível superior e uma minoria do sexo masculino, assim como no caso dos docentes. No entanto, com relação à experiência, a situação se inverte quando se compara a rede pública com a privada, tanto no 5º como no 9º ano. Para a pública cerca de 24% dos gestores tem mais de 10 anos de experiência enquanto que na rede particular esse número ultrapassa 50%. Isso se deve ao fato de eleições diretas com a participação da comunidade escolar serem realizadas em boa parte das escolas públicas, o que aumenta a rotatividade e faz com que o mesmo diretor não fique muito tempo no cargo.

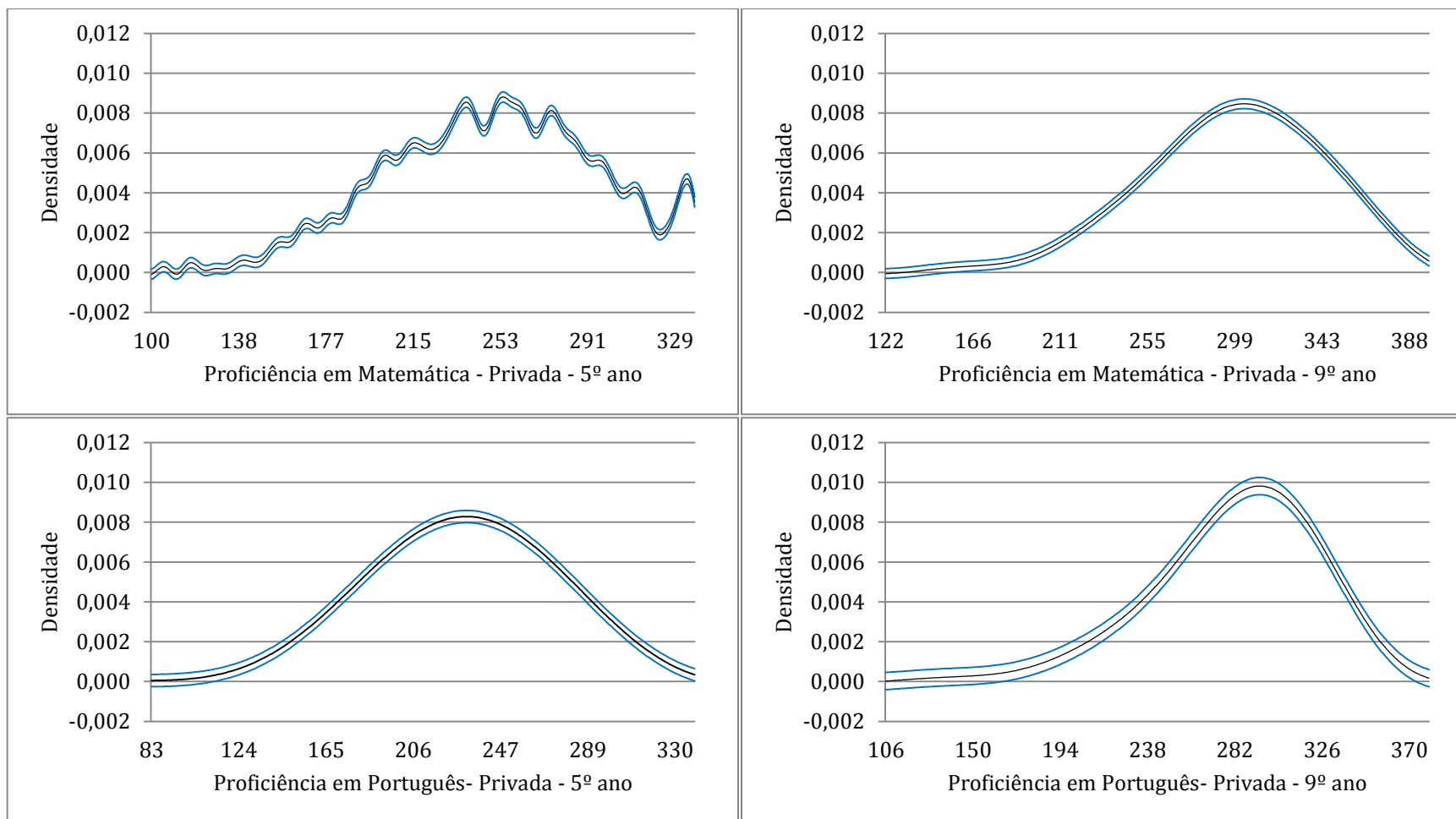
Foi considerada ainda uma *dummy* relatando um pequeno percentual de escolas que utilizam provas de seleção como critério de admissão de alunos (cerca de 1% para a pública e 12% para a privada) e uma *dummy* para a área urbana (cerca de 90% das públicas e 97% das privadas se situam nessa área). A diferença de gênero entre aluno e professor ocorre em 48% das escolas públicas e 50% nas escolas privadas, enquanto que diferenciais de raça ocorrem com percentual de 40% e 41% nas escolas públicas e privadas, respectivamente. Por fim, observa-se que 48% das turmas são homogêneas com relação à idade nas escolas públicas, enquanto que nas escolas privadas esse percentual é de 56%. A homogeneidade das turmas com relação à idade e a homogeneidade/heterogeneidade com relação ao rendimento escolar apresentam proporção bem menor em comparação ao número total de turmas analisadas.

Gráfico 1: Densidades estimadas da proficiência dos alunos por série e por tipo de escola



Nota: A densidade é obtida pelo procedimento proposto por Bernacchia e Pigolotti (2011). As linhas em azul representam as bandas de variabilidade obtidas por *bootstrap* com 1000 replicações.

Fonte: Elaborado pelos autores.



Nota: A densidade é obtida pelo procedimento proposto por Bernacchia e Pigoletti (2011). As linhas em azul representam as bandas de variabilidade obtidas por *bootstrap* com 1000 replicações.

Fonte: Elaborado pelos autores.

### 3.3 Resultados estimados

Esta seção contém os resultados obtidos pela estratégia de estimação não paramétrica descrita anteriormente. Para verificar se a especificação não paramétrica é mais robusta vis-à-vis à paramétrica utilizou-se o teste de razão de verossimilhança generalizada proposto por Fan *et al.* (2001). Nesse teste compara-se a soma dos quadrados dos resíduos dos modelos linear e não paramétrico sob a hipótese nula de igualdade entre os dois métodos. Fan e Yao (2005) apresentam a distribuição assintótica do teste por meio de *bootstrap*. A rejeição da hipótese nula evidenciada pelo *p-value* na Tabela 2 corrobora a utilização do método de estimação não paramétrico proposto na seção 2.

Tabela 2: Teste de especificação não paramétrica –  $H_0$ :  
Perfeita especificação paramétrica

	Pública			
	Matemática		Português	
	5º ano	9º ano	5º ano	9º ano
<i>Estatística</i>	1.237,338	2.055,045	1.500, 57	2.438, 569
<i>P-value</i>	[0,0000]	[0,0000]	[0,0000]	[0,0000]
	Privada			
	Matemática		Português	
	5º ano	9º ano	5º ano	9º ano
<i>Estatística</i>	277,062	363,700	334,962	527,575
<i>P-value</i>	[0,0000]	[0,0000]	[0,0000]	[0,0000]

Fonte: Elaborado pelos autores.

As variáveis consideradas para a análise do desempenho escolar podem ser classificadas de modo geral como associadas a características do aluno, do professor, do gestor e da escola. As estimativas da relação entre a proficiência e todas as variáveis explicativas se encontram sumarizadas na Tabela 3, referentes aos alunos de escolas públicas e privadas.

Com relação ao aluno, as variáveis abrangem características intrínsecas (raça e sexo), seu *background* familiar (renda e escolaridade dos pais), e sua vida escolar. Na rede pública, o status socioeconômico e o fato de o aluno ser branco contribuem, em média, para uma maior proficiência, corroborando resultados conhecidos da literatura. Para os alunos da rede privada, a situação socioeconômica apresentou impacto estatisticamente significativo apenas na proficiência em língua portuguesa daqueles pertencentes ao 9º ano. A variável gênero, em ambas as redes de ensino, apresentou impacto diferenciado entre as notas de Português e Matemática, no qual os homens apresentaram melhor desempenho nesta última disciplina. Outro resultado esperado mostra que o aluno com histórico de reprovações e filho de pais que nunca frequentaram a escola tende a apresentar um pior desempenho, o que reforça a ideia de persistência intergeracional na educação, amplamente discutida em diversos trabalhos sobre o tema, como Ferreira e Veloso (2003, 2006).

A influência da pré-escola é motivo de grande discussão dentro da literatura. Boa parte dos trabalhos alerta para sua importância, uma vez que esta fase seria crucial para o desenvolvimento de habilidades cognitivas do aluno capazes de exercer influência sobre sua capacidade de aprendizagem nas fases seguintes. Heckman e Jacobs (2009) mostram que investimentos em educação na primeira infância tendem a apresentar maior retorno em

comparação com os esforços direcionados ao ensino universitário. Os resultados aqui encontrados foram positivos para o desempenho escolar em ambas as séries e disciplinas avaliadas, consoante com outros testes realizados no Brasil com a base de dados do SAEB, tal qual o desenvolvido por Curi e Menezes-Filho (2009).

Uma contribuição a este debate consistiu em, além de analisar o impacto da pré-escola, investigar as consequências da ausência dessa fase no rendimento escolar, ou seja, mensurar como a proficiência do aluno responde ao início da vida escolar na primeira série do ensino fundamental. Na rede privada, o desempenho dos alunos mostrou possuir uma relação negativa e significativa com esta variável. Já para os estudantes da rede pública, o início relativamente tardio da vida escolar se mostrou desfavorável apenas para os alunos do 5º ano. As evidências empíricas presentes na literatura não possuem explicações para este resultado. Uma hipótese que pode ser levantada é a de que a pré-escola apresentaria maior impacto sobre as séries iniciais, reduzindo seu efeito ao longo do tempo, seja pela adaptação do aluno ao ambiente escolar e/ou pelo seu esforço ao estudar, revertendo as possíveis limitações originadas pela falta do ensino infantil. Tais suposições, porém, ainda carecem de evidências empíricas mais robustas.

O desvio padrão da idade dentro da turma consiste em um indicador de dispersão desta variável. As estimativas para a rede pública mostram que quanto mais dispersa a turma com relação à idade, maior tende a ser a proficiência do estudante. Tal resultado chama atenção, uma vez que contrapõe o argumento consensual de que diferenças de idade afetam negativamente o desempenho dos alunos. Em turmas com maior dispersão de idade, as dificuldades de se implantar projetos comuns de aprendizado são mais expressivas, tendo em vista que há uma diversidade maior de interesses (Machado, Firpo e Gonzaga, 2013). O trabalho de Hidalgo-Hidalgo (2011), entretanto, mostra que o efeito da dispersão nem sempre é conclusivo. A autora desenvolve um modelo teórico que considera a influência da turma no aprendizado individual, definindo critérios para eleger qual o modelo ótimo de alocação dos alunos: um que resulta em uma turma mais homogênea e outro que resulta em uma turma mais heterogênea. Os resultados do modelo mostram que o impacto das diferenças de idade e desempenho sobre a proficiência pode variar conforme o nível inicial de dispersão. Para turmas mais homogêneas os efeitos da redução da dispersão seriam decrescentes.

Com relação aos professores, o fato de este ser do sexo masculino tem um impacto positivo na proficiência dos alunos da escola pública, sendo este impacto não significativo nas escolas privadas, com exceção do 9º ano em português. Já para a variável raça, o fato de o professor ser da cor branca apresenta impacto não significativo para a rede privada e para o 9º ano da escola pública. Becker (2014) chega à conclusão de que o gênero do professor e sua diferença em relação ao do aluno não tende a aumentar o interesse deste último em estudar, o que depende em grande parte de questões de preferência do próprio aluno.

Com relação à experiência do professor, verifica-se uma relação negativa e significativa entre a proficiência e o fato do professor ter 10 anos ou mais de experiência apenas para o 5º ano da rede pública. Tal resultado pode soar a princípio contraditório, entretanto pode significar que a experiência indica maior domínio e segurança do professor, ou pode criar divergência entre geração em termos de linguagem ou métodos pedagógicos com os alunos. Para o caso do professor com pós-graduação (especialização, mestrado ou doutorado) a relação também é significativa apenas para o 5º ano da escola pública. Nesse caso, porém, a relação é positiva, evidenciando que a qualificação do professor pode contribuir para o aprendizado dos alunos, sobretudo nos anos iniciais da vida acadêmica, quando são mais dependentes do auxílio do docente.

Tabela 3: Resultados estimados do modelo aditivo generalizado

Variáveis	Escolas públicas				Escolas privadas			
	Matemática		Português		Matemática		Português	
	5º ano	9º ano	5º ano	9º ano	5º ano	9º ano	5º ano	9º ano
Aluno homem	5,721*** (0,184)	8,978*** (0,145)	-8,078*** (0,185)	-11,48*** (0,146)	7,815*** (1,569)	9,378*** (0,991)	-9,900*** (1,608)	-10,09*** (0,973)
Aluno branco	1,785*** (0,153)	3,141*** (0,145)	2,058*** (0,154)	3,944*** (0,145)	4,190*** (1,185)	2,198** (0,930)	3,475*** (1,214)	3,524*** (0,914)
Idade do aluno	-1,450*** (0,0690)	-3,912*** (0,0724)	-1,417*** (0,0693)	-3,090*** (0,0726)	-4,984*** (0,927)	-4,844*** (0,734)	-3,995*** (0,950)	-3,163*** (0,721)
Desvio padrão da idade na turma	5,555*** (0,170)	7,527*** (0,198)	6,075*** (0,171)	6,980*** (0,198)	7,302*** (2,531)	10,48*** (2,303)	7,062*** (2,595)	9,006*** (2,262)
Índice socioeconômico	1,716*** (0,0916)	1,542*** (0,0924)	0,621*** (0,0919)	1,232*** (0,0927)	0,552 (0,742)	-0,866 (0,601)	-1,012 (0,761)	-2,011*** (0,591)
Pais nunca estudaram	-7,330*** (0,439)	-6,626*** (0,472)	-5,846*** (0,441)	-6,209*** (0,473)	-29,34** (13,39)	-22,22 (14,28)	-33,09** (13,73)	-7,648 (14,03)
Já foi reprovado	-13,04*** (0,173)	-12,57*** (0,173)	-12,99*** (0,174)	-12,20*** (0,173)	-19,42*** (2,052)	-24,30*** (1,619)	-20,83*** (2,103)	-21,64*** (1,591)
Iniciou estudos na pré-escola	4,256*** (0,152)	5,339*** (0,146)	3,413*** (0,153)	5,182*** (0,147)	-0,459 (1,246)	2,089** (0,940)	-1,787 (1,277)	2,215** (0,923)
Iniciou estudos na primeira série	-1,830*** (0,181)	0,675*** (0,183)	-1,710*** (0,182)	0,309* (0,184)	-10,36*** (1,832)	-5,149*** (1,799)	-9,181*** (1,877)	-7,867*** (1,767)
Tamanho da família	-3,371*** (0,0905)	-1,216*** (0,142)	-4,161*** (0,0908)	-2,602*** (0,142)	-3,151*** (0,696)	1,169 (0,912)	-4,560*** (0,713)	-1,298 (0,895)
Incentivo para estudar	5,312*** (0,0795)	1,559*** (0,0893)	6,104*** (0,0798)	1,829*** (0,0896)	3,984*** (0,899)	0,528 (0,731)	5,992*** (0,921)	2,028*** (0,718)
Tamanho da turma	-0,146*** (0,0160)	-0,246*** (0,0123)	-0,176*** (0,0161)	-0,261*** (0,0123)	0,0509 (0,118)	0,0246 (0,0844)	0,0105 (0,121)	-0,000467 (0,0829)
Professor: homem	0,348* (0,191)	0,444*** (0,148)	0,409** (0,192)	0,384*** (0,148)	0,233 (1,665)	0,160 (1,103)	0,754 (1,706)	-0,440 (1,083)
Professor: branco	0,0917	-0,0517	0,0132	-0,116	-0,0820	-0,0603	-0,299	-0,160



	(0,166)	(0,155)	(0,166)	(0,156)	(1,390)	(1,064)	(1,425)	(1,045)
Professor com mais de 10 anos de experiência	-0,292**	0,0213	-0,272*	0,0129	-0,369	-0,243	-0,0487	0,0395
	(0,148)	(0,138)	(0,148)	(0,138)	(1,272)	(1,015)	(1,303)	(0,997)
Professor fez curso de atualização	-0,0476	-0,108	0,0187	-0,0104	0,0260	-1,010	0,379	-1,357
	(0,193)	(0,166)	(0,194)	(0,167)	(1,851)	(1,317)	(1,897)	(1,294)
Professor tem especialização, mestrado ou doutorado	0,401***	0,0461	0,527***	0,158	-0,599	0,117	0,0973	0,568
	(0,139)	(0,136)	(0,139)	(0,136)	(1,249)	(0,983)	(1,280)	(0,966)
Professor tem superior	0,182	0,149	0,167	0,221	0,561	-0,330	0,126	-0,322
	(0,138)	(0,188)	(0,138)	(0,188)	(1,253)	(1,240)	(1,284)	(1,218)
<hr/>								
Diretor: homem	0,461**	0,174	0,501***	0,172	-0,721	0,129	-0,355	0,386
	(0,185)	(0,144)	(0,185)	(0,144)	(1,718)	(1,205)	(1,761)	(1,184)
Diretor: branco	-0,705***	-0,637***	-0,604***	-0,500***	-0,0405	-0,115	0,536	-0,611
	(0,155)	(0,146)	(0,156)	(0,147)	(1,311)	(1,092)	(1,344)	(1,073)
Diretor tem mais de 10 anos de experiência	-0,417**	-0,251*	-0,364**	-0,165	-0,366	0,215	-0,763	0,128
	(0,166)	(0,151)	(0,166)	(0,151)	(1,247)	(0,946)	(1,278)	(0,929)
Diretor tem curso superior	0,0412	-0,423	0,0723	-0,460	1,374	-3,358	1,091	-4,551
	(0,482)	(0,532)	(0,484)	(0,533)	(5,770)	(7,184)	(5,914)	(7,056)
<hr/>								
Qualidade da escola	-0,301***	-0,0542	-0,179*	-0,0478	-0,390	0,209	-0,0276	0,604
	(0,0909)	(0,0892)	(0,0913)	(0,0894)	(0,854)	(0,684)	(0,876)	(0,672)
Violência externa	0,176*	0,0855	0,180*	0,124	0,975	0,0822	1,126	-0,365
	(0,0925)	(0,0843)	(0,0928)	(0,0846)	(1,346)	(1,222)	(1,380)	(1,200)
Violência interna	0,236**	0,0503	0,228**	0,0712	0,484	0,760	0,328	0,524
	(0,0940)	(0,0810)	(0,0943)	(0,0812)	(1,486)	(1,078)	(1,523)	(1,058)
Prova de seleção	-0,509	-0,819	-0,450	-0,0594	-0,377	0,192	0,00564	0,226
	(0,704)	(0,593)	(0,707)	(0,595)	(1,868)	(1,290)	(1,915)	(1,267)
Turmas homogêneas quanto à idade	0,238	0,916***	0,343*	0,836***	0,200	-0,233	0,863	0,254
	(0,188)	(0,177)	(0,188)	(0,178)	(1,536)	(1,173)	(1,574)	(1,152)
Turmas homogêneas quanto ao rendimento escolar	0,332	0,331	0,285	0,297	-0,225	-1,266	1,413	-1,037
	(0,309)	(0,346)	(0,310)	(0,347)	(2,676)	(2,369)	(2,743)	(2,327)
Turmas heterogêneas quanto à idade	0,544**	0,521**	0,511*	0,485*	-0,667	0,560	0,270	1,228
	(0,270)	(0,249)	(0,271)	(0,249)	(3,393)	(2,010)	(3,478)	(1,974)

Turmas heterogêneas quanto ao rendimento escolar	-0,119 (0,219)	-0,0629 (0,214)	-0,0456 (0,220)	0,0194 (0,214)	0,00932 (2,079)	0,00287 (1,580)	0,886 (2,131)	0,307 (1,552)
Diferença de sexo entre aluno e professor	-0,122 (0,183)	-0,587*** (0,143)	-0,272 (0,184)	-0,441*** (0,144)	0,0302 (1,561)	-1,545 (0,983)	0,402 (1,600)	-1,587 (0,966)
Diferença de raça entre aluno e professor	-1,664*** (0,150)	-1,439*** (0,140)	-1,483*** (0,151)	-1,062*** (0,141)	-1,879 (1,162)	0,341 (0,915)	0,237 (1,191)	0,221 (0,899)
Região Urbana	-1,143*** (0,251)	-0,193 (0,251)	-1,035*** (0,252)	-0,236 (0,252)	-3,659 (3,730)	-3,343 (3,002)	-0,898 (3,823)	-2,465 (2,949)
Nordeste	0,174*** (0,00981)	0,136*** (0,00946)	0,151*** (0,00984)	0,0998*** (0,00949)	0,0711 (0,0908)	0,00279 (0,0746)	0,0920 (0,0930)	0,00257 (0,0733)
Norte	3,022*** (0,231)	2,500*** (0,230)	2,682*** (0,232)	2,304*** (0,230)	0,551 (1,830)	-0,444 (1,504)	0,943 (1,875)	0,284 (1,478)
Sul	-0,163 (0,270)	-1,395*** (0,240)	-0,108 (0,271)	-1,281*** (0,241)	-2,810 (2,737)	-0,930 (2,190)	-2,275 (2,805)	-0,500 (2,152)
Centro-Oeste	0,356 (0,362)	0,0761 (0,333)	0,110 (0,363)	0,241 (0,334)	-0,623 (2,468)	-0,0925 (1,943)	-0,783 (2,529)	0,520 (1,909)

Nota: \*\*\*  $p < 0,01$ , \*\*  $p < 0,5$ , \*  $p < 0,1$ . Erro-padrão entre parênteses.

Fonte: Elaborado pelos autores.

No que concerne às características pessoais do diretor, é verificada significância estatística novamente apenas para a escola pública. Um fato que chama atenção são os coeficientes negativos e significantes para o impacto do diretor de cor branca e com mais de 10 anos de experiência. Porém, os resultados fazem sentido, uma vez que a maior parte dos alunos são não brancos e se o diretor ocupa por muito tempo o cargo pode gerar vícios de gestão que comprometam o desempenho dos alunos.

Em suma, tanto as características dos professores quanto as do diretor impactam de maneira significativa o desempenho dos alunos das escolas públicas, dado que estes estão mais suscetíveis devido ao fato de provavelmente possuírem um *background* familiar desfavorável em relação aos alunos da escola privada.

No que concerne o tamanho da turma, Camargo e Porto Junior (2014) mostram que existe uma grande controvérsia na literatura quanto ao real impacto desta variável sobre o desempenho escolar. Os autores, ao analisarem a relação entre proficiência e o tamanho da turma encontram uma relação negativa, porém não significativa. Alguns trabalhos como o de Duso e Sudbrack (2009) partem do argumento de que mais alunos na sala de aula comprometeriam uma estratégia de ensino mais individualizada e compatível com as necessidades de cada aluno. O tamanho da turma no presente artigo se mostrou significativo apenas para o desempenho dos alunos das escolas públicas. Talvez esses alunos, por apresentarem um conjunto de fatores adversos (*background* familiar desfavorável, histórico de reprovações), tenham seu aprendizado já fragilizado, de modo que o efeito do número de alunos em sala poderia ser mais expressivo. O tamanho da família apresentou impacto negativo e significativo, tanto na rede pública quanto na rede privada.

A escolha das escolas por parte das famílias constitui um processo de decisão não aleatório e, portanto, a manutenção dos alunos em uma escola depende diretamente da qualidade que a mesma pode oferecer. Dessa forma, famílias com melhores condições econômicas, avaliadas pelo índice socioeconômico, tenderiam a escolher as melhores escolas, constituindo um processo de auto-seleção e permitindo a presença de correlação entre esse e os fatores responsáveis pela proficiência do aluno. Uma forma de mitigar esta correlação é por meio da inclusão de um conjunto de covariadas relacionadas ao ambiente escolar, a saber, um índice de qualidade escolar, de violência escolar (causada por agentes internos e externos), o critério de admissão na escola, o critério utilizado pelo diretor para alocar os estudantes entre as turmas e variáveis *dummy* para características demográficas.

Com relação à qualidade da escola, obtida por meio de *principal component analysis*, observou-se uma relação negativa entre essa e o desempenho dos alunos do 5º ano do ensino fundamental para as escolas de ensino público. Esta medida indica que melhores condições de infraestrutura escolar estão, de forma geral, associadas a maiores notas nos exames padronizados, uma vez que se utilizou uma escala inversa. Esta correlação é maior, em módulo, quando se avaliam os alunos por meio da prova de Língua Portuguesa. Também se observou que não há uma relação significativa para os alunos do 9º ano e para os alunos do ensino privado.

O problema da violência escolar também merece destaque. Os dados indicaram que pelos menos 47% e 42% das escolas públicas reportaram problemas de consumo e de tráfico escolar, respectivamente. Quando se observam estes dados para os domínios da escola percebe-se que este relato aumenta. No mínimo 66,74% dos diretores das escolas públicas disseram que houve a incidência de agressão verbal a professores e 46% das mesmas relatou que 47% dos seus alunos se envolveram em agressões físicas. Para as escolas privadas estes números se reduzem consideravelmente: somente 16% destas relataram a ocorrência de furtos causados por agentes externos. Em seus domínios, menos de 32% das escolas privadas indicou que houve algum tipo de agressão entre alunos, professores e funcionários (física ou verbal). Grogger (1997), em seu estudo seminal, analisou os efeitos da violência escolar sobre

o resultado dos alunos<sup>2</sup>. Além de classificar cada tipo de violência ocorrido nas escolas como sério, moderado, menor, ou não existente, o autor concluiu que a violência apresentou efeitos expressivos: níveis moderados desta reduziram a probabilidade de formatura no ensino médio em 5,1% e diminuíram a probabilidade de um aluno ingressar no ensino superior em 6,9%. Além disso, escolas com violência moderada remuneraram os professores 2,4% a mais do que escolas sem violência. Para o caso brasileiro, esta literatura ainda é escassa, exceto pelo estudo de Severnini e Firpo (2009). As estimativas obtidas pelos autores demonstraram que o principal problema nas escolas é a presença do tráfico de drogas e/ou seu consumo nas instalações, de forma que os alunos que frequentam escolas com tais problemas tendem a ter proficiência 1% menor. Encontrou-se que esta correlação é negativa para os alunos do 5º ano das escolas públicas brasileiras, sendo esta maior no caso da violência interna. Já para os alunos do 9º ano não se encontrou nenhuma relação significativa, podendo indicar que os alunos tornam-se indiferentes à ocorrência de algum ato violento, fazendo com que os efeitos destas sobre os rendimentos escolares sejam inócuos.

No tocante ao processo de seleção dos alunos inclui-se uma variável *dummy* para reduzir a correlação do processo de auto-seleção por parte da família dos alunos. Considerando-se a prova seleção como uma etapa para a inserção do aluno na escola, não se encontrou nenhuma evidência significativa que sustente uma possível correlação entre essa e a sua performance na prova do SAEB, a despeito do sinal negativo desta variável. Esta estimativa indica que as escolas que não adotaram critérios de segregação dos alunos *a priori*, com o objetivo de maximizar seus desempenhos médios nos testes padronizados, apresentaram melhores gestões organizacionais. Dessa forma, um processo de aleatoriedade na escolha dos alunos por parte das escolas permite que aqueles com piores desempenhos não percam as boas influências que podem ser adquiridas a partir da presença de alunos com maiores níveis de habilidade.

Contrastando-se os resultados dos parâmetros relacionados à alocação dos alunos entre as turmas notou-se que aqueles que foram alocados em turmas com base na homogeneidade e na heterogeneidade quanto à idade apresentaram, em média, um maior desempenho escolar. Apesar da possível contradição nestes resultados deve ser observado que a idade do aluno apresentou um papel central sobre seu desempenho. Assim, os que foram alocados em turmas com base na homogeneidade (ou heterogeneidade) poderiam ter estudado em outras escolas, ou na atual, que também utilizavam esse critério. Neste caso, também é possível perceber uma possível correlação entre o desempenho do aluno e do grupo do qual faz parte, isto é, o *peer effects*.

Por fim, no que diz respeito às variáveis demográficas as escolas localizadas nas zonas urbanas apresentaram, em média, um menor nível de proficiência. Apesar das escolas rurais possuírem, em média, menos recursos e relativamente menos insumos do que aquelas que se encontram nas áreas urbanas, esse fator não representa uma limitação à gestão das escolas rurais. Este resultado também foi encontrado por Delgado e Machado (2007) e parece indicar que as escolas rurais, a despeito de seus menores níveis de insumos, apresentam melhores práticas de gestão. As variáveis *dummy* indicaram que as escolas das regiões Norte e Nordeste concentraram os alunos com melhores desempenhos acadêmicos vis-à-vis os da região Sudeste.

#### **4 Considerações Finais**

Quando se observa a situação do ensino no Brasil nota-se que muitas reformas como a adoção de políticas redistributivas e as transformações na organização e na gestão da educação

---

<sup>2</sup> Para mais detalhes ver Miller e Chandler (2005), Ammermüller (2007) e Carroll (2006).

pública (descentralização dos recursos educacionais concomitantemente à expansão das matrículas a partir de 1996, que culminaram na criação do Fundo de Desenvolvimento do Ensino Fundamental e de Valorização do Magistério) ocorreram no sistema de ensino nas últimas décadas com o objetivo de melhorar a excelência da educação e ampliar o acesso da população à escolaridade. Entretanto, estas ações não foram suficientes para que a sociedade possa desfrutar de uma educação de qualidade. A partir disso, observa-se que é fundamental investigar os determinantes da proficiência escolar (variável *proxy* da qualidade do ensino).

Neste sentido, o objetivo deste estudo foi investigar os determinantes da proficiência escolar por meio de um modelo aditivo generalizado, permitindo assim a identificação de possíveis padrões de comportamento não lineares da variável dependente condicionada aos fatores explicativos. Para tanto, empregou-se o método proposto por Horowitz e Mammen (2004), que consiste em uma adição de funcionais desconhecidas em dois estágios. A principal vantagem deste estimador é não sofrer do problema da dimensionalidade.

Para cumprir o objetivo proposto empregaram-se os microdados do SAEB e do Censo Escolar, ambos para o ano de 2011. É importante destacar que se incluíram variáveis para capturar o incentivo dos pais, a “armadilha do estereótipo”, as relações de violência da comunidade escolar, dentre outras.

Os resultados indicaram que para as escolas privadas somente o conjunto de variáveis relacionadas ao *background* familiar se mostraram significativos. Para os alunos da rede pública essas relações não são homogêneas. Encontramos evidências para estes últimos que confirmam o processo de persistência intergeracional na educação, uma vez que os alunos com histórico de reprovações e filho de pais que nunca frequentaram a escola tendem a apresentar um pior desempenho. No tocante à trajetória escolar observa-se que os investimentos em educação na primeira infância tendem a apresentar uma correlação positiva com o desempenho dos alunos, corroborando assim as evidências encontradas por Heckman e Jacobs (2009) e Curi e Menezes-Filho (2009). Uma hipótese que pode ser levantada é a de que a pré-escola apresentaria maior impacto sobre as séries iniciais, reduzindo seu efeito ao longo do tempo.

As estimativas para a rede pública também mostraram que quanto mais dispersa a turma quanto à idade, maior tende a ser a proficiência do estudante. Tal resultado chama atenção, uma vez que contrapõe o argumento consensual de que diferenças de idade afetam negativamente o desempenho dos alunos. Entretanto, Hidalgo-Hidalgo (2011) mostra que o efeito da dispersão nem sempre é conclusivo. Os resultados do modelo proposto pela autora mostram que o impacto das diferenças de idade e desempenho sobre a proficiência pode variar conforme o nível inicial de dispersão.

Com relação à raça, o fato do professor ser da cor branca impacta negativamente na escola privada e no 9º ano da escola pública, sendo que em ambas as redes de ensino o coeficiente é não significativo. Com relação à experiência do professor, verifica-se uma relação significativa e negativa apenas para o 5º ano da rede de ensino pública. Embora a relação negativa pareça a princípio contraditória, ela pode apresentar uma ambiguidade; por um lado a experiência denota maior domínio e segurança ao professor e por outro pode criar um abismo de idade divergência em termos de linguagem ou formas pedagógicas com os alunos. Para o caso do professor ter pós-graduação (especialização, mestrado ou doutorado) a relação também é significativa apenas para o mesmo nível (5º ano da escola pública). No entanto, nesse caso a relação é positiva, o que é uma evidência de que a qualificação do professor pode contribuir para o aprendizado dos alunos, sobretudo nos anos iniciais da vida acadêmica quando são mais dependentes do auxílio docente.

Encontraram-se resultados que demonstram que a correlação entre proficiência e violência é negativa para os alunos do 5º ano das escolas públicas brasileiras, sendo esta maior no caso da violência interna. Já para os alunos do 9º ano não se encontrou nenhuma

relação significativa, podendo indicar que os alunos tornam-se indiferentes à ocorrência de algum ato violento, fazendo com que os efeitos destas sobre os rendimentos escolares sejam inócuos.

Também se verificou que as escolas que não adotaram critérios de segregação dos alunos *a priori*, com o objetivo de maximizar seus desempenhos médios nos testes padronizados, apresentaram melhores gestões organizacionais. Dessa forma, um processo de aleatoriedade na escolha dos alunos por parte das escolas permite que aqueles com piores desempenhos não percam as boas influências que podem ser adquiridas a partir da presença de alunos com maiores níveis de habilidade.

Finalmente, no que diz respeito às variáveis demográficas as escolas localizadas nas zonas urbanas apresentaram, em média, um menor nível de proficiência. Apesar das escolas rurais possuírem, em média, menos recursos e relativamente menos insumos do que aquelas que se encontram nas áreas urbanas, esse fator não representa uma limitação à gestão das escolas rurais. Este resultado também foi encontrado por Delgado e Machado (2007) e parece indicar que as escolas rurais, a despeito de seus menores níveis de insumos, apresentaram melhores práticas de gestão.

## 5 Referências Bibliográficas

ALMEIDA, A. T. C. D. (2014). Determinantes dos piores e melhores resultados educacionais dos alunos da rede pública de ensino fundamental no Brasil. *Planejamento e Políticas Públicas*, 42: 147-187.

AMMERMÜLLER, A. (2006). Violence in European schools: victimization and consequences. Centre for European Economic Research (ZEW), *Discussion Paper* n. 07-004.

BARROS, R. P.; MENDONÇA, R.; SANTOS, D. D. dos; QUINTAES, G. (2001). *Determinantes do desempenho educacional no Brasil*. Texto para discussão do IPEA. Rio de Janeiro: IPEA, Nº 834.

BARROS, R. P.; MENDONÇA, R. (1998). *Consequências da repetência sobre o desempenho educacional*. Projeto de educação básica para o Nordeste. Brasília, DF: MEC.

BECKER, K. L. (2014). *A “armadilha do estereótipo”: a identificação étnico-racial ou de gênero entre alunos e professores pode incentivar o aluno a gostar de estudar?* Disponível em < <https://drive.google.com/folderview?id=0B4RxCaDzonzReFJPOTINejY3d3M&usp=sharing&tid=0B4RxCaDzonzRWmp0ZHk1TFNJMmc> >. Acesso em 10 jul. 2014.

BERNACCHIA, A.; PIGOLOTTI, S. (2011). Self-Consistent Method For Density Estimation. *Journal of the Royal Statistical Society B*, 73(3): 407-422.

BIONDI, R.; FELÍCIO, F. (2008). *Atributos escolares e o desempenho dos estudantes: uma análise em painel dos dados do SAEB*. Rio de Janeiro: UFF, 2008. (Texto para Discussão, n. 236).

CAMARGO, J.; JÚNIOR, S. D. S. P. (2014). *O efeito do tamanho da turma sobre o desempenho escolar: uma avaliação do impacto da “enturmação” no Ensino Fundamental do Rio Grande do Sul*. Disponível em < [http://www.ufrgs.br/fce/wordpress/wp-content/uploads/2014/04/TD14\\_2013\\_camargo\\_portojr.pdf](http://www.ufrgs.br/fce/wordpress/wp-content/uploads/2014/04/TD14_2013_camargo_portojr.pdf) >. Acesso em 13 jul. 2014.

CARROLL, B. R. (2006). *The effects of school violence and crime on academic achievement*. Davidson College, 33 p.

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

DELGADO, V. M. S.; MACHADO, A. F. (2007). Eficiência das escolas públicas estaduais de Minas Gerais. *Pesquisa e Planejamento Econômico*, 37(3): 427-464.

DUSO, A. P.; SUDBRACK, E. M. (2009). Política educacional: para além da racionalidade econômica - questionando a enturmação. *Revista de Ciências Humanas*, 9(15): 1-50.

FAN, J.; YAO, Q. (2003). *Nonlinear Time Series: Nonparametric and Parametric Methods*. Springer.

FAN, J.; ZHANG, C.; ZHANG, J. (2001). Generalized likelihood ratio statistics and Wilks phenomenon. *The Annals of Statistics*, 29: 153–193.

FERREIRA, S.; VELOSO, F. (2003). Mobilidade intergeracional de educação no Brasil. *Pesquisa e Planejamento Econômico*, 33(3): 481–513.

FERREIRA, S.; VELOSO, F. (2006). Intergenerational mobility of wages in Brazil. *Brazilian Review of Econometrics*, 26(2): 181-211.

FRANCO, C; MANDARINO, M.; ORTIGÃO, M. I. (2002). O projeto pedagógico e os resultados escolares. *Pesquisa e Planejamento Econômico*, 32(3): 477-494.

GROGGER, J. (1997). Local violence and educational attainment. *Journal of Human Resources*, 32(4): 659-682.

HASTIE, T. J.; TIBSHIRANI, R. J. (1990). *Generalized Additive Models*. Chapman and Hall, London.

HECKMAN, J.; JACOBS, B. (2009). Policies to Create and Destroy Human Capital in Europe. *IZA Discussion Papers* 4680, Institute for the Study of Labor (IZA).

HOROWITZ, J. L.; MAMMEN, E. (2004). Nonparametric estimation of an additive model with a link function. *The Annals of Statistics*, 32(6): 2412-2443.

HIDALGO-HIDALGO, M. (2011). On the optimal allocation of students when peer effects are at work: tracking vs. mixing. *SERIES*, 2(1), 31-52.

HOROWITZ, J. L., KLEMELA, J. and MAMMEN, E. (2002). Optimal estimation in additive regression models. *Working paper*, Germany.

LEON, F. L.; MENEZES-FILHO, N. A. (2002). Reprovação, avanço e evasão escolar no Brasil. *Pesquisa e planejamento econômico*, 32(3): 417-452.

LINTON, O. B.; HARDLE, W. (1996). Estimating additive regression models with known links. *Biometrika*, 83: 529–540.

MACHADO, D. C.; GONZAGA, G.; FIRPO, S. P. (2013). A relação entre proficiência e dispersão de idade na sala de aula: a influência do nível de qualificação do professor. *Pesquisa e Planejamento Econômico*, 43(3): 419-445.

MACHADO, A. F.; MORO, S.; MARTINS L.; RIOS, J. (2008). Qualidade do Ensino em Matemática: Determinantes do Desempenho de Alunos em Escolas Públicas Estaduais Mineiras. *Economia*, 9(1): 23-45.

MAMMEN, E., LINTON, O. B.; NIELSEN, J. P. (1999). The existence and asymptotic properties of backfitting projection algorithm under weak conditions. *Ann. Statist.* 27: 1443–1490.

MENEZES-FILHO, N.; VASCONCELLOS, L.; WERLANG, S. R. da C.; BIONDI, R. L. (2008). *Avaliando o Impacto da Progressão Continuada nas Taxas de Rendimento e Desempenho Escolar do Brasil*. 2008. Disponível em <  
[http://www.fundacaoitausocial.org.br/\\_arquivosstaticos/FIS/pdf/10\\_-\\_relatorio\\_de\\_avaliacao\\_progressao\\_continuada\\_-\\_atualizado.pdf](http://www.fundacaoitausocial.org.br/_arquivosstaticos/FIS/pdf/10_-_relatorio_de_avaliacao_progressao_continuada_-_atualizado.pdf) > Acesso em 17 jul. 2014.

MILLER, A.; CHANDLER, K. (2005). *Violence in U.S. Public Schools: 2000 School Survey on Crime and Safety*. National Center for Education Statistics, Statistical Analysis Report, 114 p.

NEWAY, W. K. (1997). Convergence rates and asymptotic normality for series estimators. *Econometrics*, 79: 147–168.

OPSOMER, J. D. (2000). Asymptotic properties of backfitting estimators. *J. Multivariate Anal.* 73: 166–179.

SOUZA, A. M.; SILVA, N. V. (1994). Origem familiar, qualidade da educação e escolas públicas em São Paulo: relações e efeitos nas transições escolares. *Pesquisa e Planejamento Econômico*, 24(1): 97-114.

STEELE, C. M. ARONSON, J. (1995) Stereotype Threat and the Intellectual Test-Performance of African Americans. *Journal of Personality and Social Psychology*, 69(5): 797-811.

STONE, C. J. (1994). The use of polynomial splines and their tensor products in multivariate function estimation. *Ann. Statist.* 22: 118–184.



## Anexo A

O índice socioeconômico utilizado neste estudo como variável *proxy* para as condições de vida da família dos alunos, a medida de incentivo aos estudos, o índice de qualidade da escola e os índices de violência interna e externa são obtidos por meio da técnica de análise de componentes principais. Assim, assume-se que existe uma relação linear entre um conjunto de perguntas respondidas pelos alunos com um conjunto de fatores latentes, isto é,

$$A - \mu = RF + \varrho$$

em que  $A$  é o conjunto de perguntas do questionário dos alunos,  $\mu$  é a média de  $A$ ,  $R$  representa uma matriz de ponderação e  $F$  é o fator. Assume-se que  $F$  são variáveis ortogonais independentes de  $\varrho$ . Disto decorre que  $\mathbb{E}[F] = 0$  e  $\text{Cov}[F] = I$ . Finalmente,  $\mathbb{E}[\varrho] = 0$  e  $\text{Cov}[\varrho] = \Omega$ . Uma vez que se assumiu que  $\varrho$  e  $F$  são independentes é possível estabelecer que

$$Y = \text{Cov}[A] = RR' + \Omega$$

A análise de componentes principais permite decompor  $Y$ , de tal forma que  $Y \sim \Gamma \Lambda \Gamma^t + \Omega$ , em que  $\Lambda$  é uma matriz diagonal que contém os autovalores e  $\Gamma$  é uma matriz com os autovetores associados. Os fatores  $F$  de interesse são obtidos pela aplicação dos seguintes passos:

1. Construa a matriz  $A$ ;
2. Subtraia a média de cada coluna de  $A$ ;
3. Calcule a matriz de covariância de  $A$ ;
4. Calcule os autovalores e os autovetores da matriz de covariância;
5. Escolha o maior autovetor que é associado ao maior autovalor. Este autovetor é a base de construção do componente principal. O componente principal é o produto da raiz quadrada do autovalor pelo autovetor.

Para o índice socioeconômico foram utilizadas a resposta de 6 perguntas: número de TVs (0, 1, 2 mais), número de geladeiras (0, 1 ou mais), número de carros (0, 1, 2 ou mais), computadores (0 se não tem, 1 se não tem acesso à internet e 2 se tem acesso à internet), número de dormitórios (0, 1, 2, 3 ou mais) e se há empregada doméstica em sua casa (0 se não tem, 1 se for diarista, 2 se for uma e 3 se for mais de duas empregadas).

Para a medida de incentivo aos estudos considerou-se se os pais incentivam o aluno: a estudar (0, não; 1, sim), a fazer o dever de casa (0, não; 1, sim), a ler (0, não; 1, sim), a ir a escola (0, não; 1, sim) e se conversam sobre a escola (0, não; 1, sim).

O índice de qualidade da escola foi construído utilizando-se a seguintes variáveis para indicar se a escola tem (0, não; 1, sim): acesso a rede água pública, acesso a rede de energia elétrica, acesso a coleta de lixo, tem cozinha, sala de diretoria, sala de professores, laboratório de informática, laboratório de ciência, quadra de esportes descoberta, quadra de esportes coberta, biblioteca, sala de leitura, sanitário dentro do prédio e sanitário fora do prédio.

O índice de violência externa considera a incidência de atentado à vida de professores, alunos ou funcionários, furtos, roubos, consumo de bebida e drogas, tráfico de drogas e atuação de gangues (causados por agente externo).

O índice de violência interna considera a incidência de atentado à vida de professores, alunos ou funcionários, furtos, roubos, consumo de bebida e drogas, tráfico de

drogas, posse de armas, agressões físicas e verbais entre alunos, funcionários e professores e atuação de gangues (causados por agente externo).

## Anexo B: Estatísticas descritivas

Tabela 4: Estatísticas descritivas – 5º ano do ensino fundamental – ensino público

	<b>Média</b>	<b>Desvio-padrão</b>	<b>Mínimo</b>	<b>Máximo</b>
Proficiência em Português	188,80	45,78	77,20	339,50
Proficiência em Matemática	208,00	47,28	90,13	338,20
Aluno homem	0,525	0,499	0	1
Aluno branco	0,284	0,451	0	1
Idade do aluno	10,910	1,206	8	15
Desvio padrão da idade na turma	0,933	0,459	0	3,536
Índice socioeconômico	-0,053	0,804	-1,951	3,005
Pais nunca estudaram	0,025	0,155	0	1
Já foi reprovado	0,312	0,463	0	1
Iniciou estudos na pré-escola	0,385	0,487	0	1
Iniciou estudos na primeira série	0,219	0,413	0	1
Tamanho da família	3,527	0,750	1	4
Incentivo para estudar	-0,070	0,863	-4,402	0,377
Tamanho da turma	10,820	4,268	1	32
Professor: homem	0,160	0,366	0	1
Professor: branco	0,426	0,494	0	1
Professor com mais de 10 anos de experiência	0,695	0,461	0	1
Professor fez curso de atualização	0,856	0,351	0	1
Professor tem especialização, mestrado ou doutorado	0,450	0,498	0	1
Professor tem superior	0,520	0,500	0	1
Diretor: homem	0,273	0,446	0	1
Diretor: branco	0,499	0,500	0	1
Diretor tem mais de 10 anos de experiência	0,248	0,432	0	1
Diretor tem curso superior	0,830	0,120	0	1
Qualidade da escola	0,154	0,847	-6,890	1,270
Violência externa	0,119	0,972	-0,793	5,695
Violência interna	0,165	1,017	-1,013	7,624
Prova de seleção	0,012	0,107	0	1
Turmas homogêneas quanto à idade	0,480	0,500	0	1
Turmas homogêneas quanto ao rendimento escolar	0,041	0,198	0	1
Turmas heterogêneas quanto à idade	0,099	0,299	0	1
Turmas heterogêneas quanto ao rendimento escolar	0,178	0,383	0	1
Diferença de sexo entre aluno e professor	0,482	0,500	0	1
Diferença de raça entre aluno e professor	0,404	0,491	0	1
Região Urbana	0,906	0,292	0	1
Nordeste	0,312	0,463	0	1
Norte	0,160	0,367	0	1
Sul	0,104	0,305	0	1
Centro-Oeste	0,043	0,203	0	1

Fonte: Elaborado pelos autores.

Tabela 5: Estatísticas descritivas – 9º ano do ensino fundamental – ensino público

	<b>Média</b>	<b>Desvio-padrão</b>	<b>Mínimo</b>	<b>Máximo</b>
Proficiência em Português	240,20	45,78	77,20	339,50
Proficiência em Matemática	246,60	47,28	90,13	338,20
Aluno homem	0,460	0,499	0	1
Aluno branco	0,312	0,451	0	1
Idade do aluno	10,850	1,206	8	15
Desvio padrão da idade na turma	0,925	0,459	0	3,536
Índice socioeconômico	0,010	0,804	-1,951	3,005
Pais nunca estudaram	0,019	0,155	0	1
Já foi reprovado	0,337	0,463	0	1
Iniciou estudos na pré-escola	0,460	0,487	0	1
Iniciou estudos na primeira série	0,210	0,413	0	1
Tamanho da família	3,820	0,750	1	4
Incentivo para estudar	0,050	0,863	-4,402	0,377
Tamanho da turma	16,300	4,268	1	32
Professor: homem	0,268	0,366	0	1
Professor: branco	0,474	0,494	0	1
Professor com mais de 10 anos de experiência	0,669	0,461	0	1
Professor fez curso de atualização	0,814	0,351	0	1
Professor tem especialização, mestrado ou doutorado	0,399	0,498	0	1
Professor tem superior	0,861	0,500	0	1
Diretor: homem	0,273	0,446	0	1
Diretor: branco	0,499	0,500	0	1
Diretor tem mais de 10 anos de experiência	0,248	0,432	0	1
Diretor tem curso superior	0,830	0,120	0	1
Qualidade da escola	0,154	0,847	-6,890	1,270
Violência externa	0,119	0,972	-0,793	5,695
Violência interna	0,165	1,017	-1,013	7,624
Prova de seleção	0,012	0,107	0	1
Turmas homogêneas quanto à idade	0,480	0,500	0	1
Turmas homogêneas quanto ao rendimento escolar	0,041	0,198	0	1
Turmas heterogêneas quanto à idade	0,099	0,299	0	1
Turmas heterogêneas quanto ao rendimento escolar	0,178	0,383	0	1
Diferença de sexo entre aluno e professor	0,482	0,500	0	1
Diferença de raça entre aluno e professor	0,404	0,491	0	1
Região Urbana	0,906	0,292	0	1
Nordeste	0,312	0,463	0	1
Norte	0,160	0,367	0	1
Sul	0,104	0,305	0	1
Centro-Oeste	0,043	0,203	0	1

Fonte: Elaborado pelos autores.

Tabela 6: Estatísticas descritivas – 5º ano do ensino fundamental – ensino privado

	<b>Média</b>	<b>Desvio-padrão</b>	<b>Mínimo</b>	<b>Máximo</b>
Proficiência em Português	229,00	45,78	77,20	339,50
Proficiência em Matemática	249,20	47,28	90,13	338,20
Aluno homem	0,515	0,499	0	1
Aluno branco	0,415	0,451	0	1
Idade do aluno	10,270	1,206	8	15
Desvio padrão da idade na turma	0,574	0,459	0	3,536
Índice socioeconômico	0,924	0,804	-1,951	3,005
Pais nunca estudaram	0,002	0,155	0	1
Já foi reprovado	0,095	0,463	0	1
Iniciou estudos na pré-escola	0,338	0,487	0	1
Iniciou estudos na primeira série	0,116	0,413	0	1
Tamanho da família	3,213	0,750	1	4
Incentivo para estudar	0,105	0,863	-4,402	0,377
Tamanho da turma	14,530	4,268	1	32
Professor: homem	0,152	0,366	0	1
Professor: branco	0,484	0,494	0	1
Professor com mais de 10 anos de experiência	0,641	0,461	0	1
Professor fez curso de atualização	0,889	0,351	0	1
Professor tem especialização, mestrado ou doutorado	0,403	0,498	0	1
Professor tem superior	0,664	0,500	0	1
Diretor: homem	0,136	0,446	0	1
Diretor: branco	0,573	0,500	0	1
Diretor tem mais de 10 anos de experiência	0,503	0,432	0	1
Diretor tem curso superior	0,910	0,120	0	1
Qualidade da escola	0,228	0,847	-6,890	1,270
Violência externa	-0,531	0,972	-0,793	5,695
Violência interna	-0,621	1,017	-1,013	7,624
Prova de seleção	0,128	0,107	0	1
Turmas homogêneas quanto à idade	0,563	0,500	0	1
Turmas homogêneas quanto ao rendimento escolar	0,062	0,198	0	1
Turmas heterogêneas quanto à idade	0,035	0,299	0	1
Turmas heterogêneas quanto ao rendimento escolar	0,150	0,383	0	1
Diferença de sexo entre aluno e professor	0,503	0,500	0	1
Diferença de raça entre aluno e professor	0,413	0,491	0	1
Região Urbana	0,973	0,292	0	1
Nordeste	0,237	0,463	0	1
Norte	0,298	0,367	0	1
Sul	0,055	0,305	0	1
Centro-Oeste	0,078	0,203	0	1

Fonte: Elaborado pelos autores.

Tabela 7: Estatísticas descritivas – 9º ano do ensino fundamental – ensino privado

	<b>Média</b>	<b>Desvio-padrão</b>	<b>Mínimo</b>	<b>Máximo</b>
Proficiência em Português	282,50	45,78	77,20	339,50
Proficiência em Matemática	298,30	47,28	90,13	338,20
Aluno homem	0,480	0,499	0	1
Aluno branco	0,470	0,451	0	1
Idade do aluno	10,320	1,206	8	15
Desvio padrão da idade na turma	0,675	0,459	0	3,536
Índice socioeconômico	0,991	0,804	-1,951	3,005
Pais nunca estudaram	0,001	0,155	0	1
Já foi reprovado	0,129	0,463	0	1
Iniciou estudos na pré-escola	0,429	0,487	0	1
Iniciou estudos na primeira série	0,073	0,413	0	1
Tamanho da família	3,758	0,750	1	4
Incentivo para estudar	0,099	0,863	-4,402	0,377
Tamanho da turma	20,650	4,268	1	32
Professor: homem	0,280	0,366	0	1
Professor: branco	0,557	0,494	0	1
Professor com mais de 10 anos de experiência	0,646	0,461	0	1
Professor fez curso de atualização	0,854	0,351	0	1
Professor tem especialização, mestrado ou doutorado	0,386	0,498	0	1
Professor tem superior	0,824	0,500	0	1
Diretor: homem	0,136	0,446	0	1
Diretor: branco	0,573	0,500	0	1
Diretor tem mais de 10 anos de experiência	0,503	0,432	0	1
Diretor tem curso superior	0,910	0,120	0	1
Qualidade da escola	0,228	0,847	-6,890	1,270
Violência externa	-0,531	0,972	-0,793	5,695
Violência interna	-0,621	1,017	-1,013	7,624
Prova de seleção	0,128	0,107	0	1
Turmas homogêneas quanto à idade	0,563	0,500	0	1
Turmas homogêneas quanto ao rendimento escolar	0,062	0,198	0	1
Turmas heterogêneas quanto à idade	0,035	0,299	0	1
Turmas heterogêneas quanto ao rendimento escolar	0,150	0,383	0	1
Diferença de sexo entre aluno e professor	0,503	0,500	0	1
Diferença de raça entre aluno e professor	0,413	0,491	0	1
Região Urbana	0,973	0,292	0	1
Nordeste	0,237	0,463	0	1
Norte	0,298	0,367	0	1
Sul	0,055	0,305	0	1
Centro-Oeste	0,078	0,203	0	1

Fonte: Elaborado pelos autores.