

BACKGROUND FAMILIAR E DESEMPENHO ESCOLAR: UMA ABORDAGEM NÃO PARAMÉTRICA

Wallace Patrick Santos de Farias Souza¹

Victor Rodrigues de Oliveira²

Ana Cláudia Annegues³

Uma ampla literatura teórica e empírica sugere que há uma transferência intergeracional de características socioeconômicas da família, como a renda e a educação. Nesse sentido, o objetivo deste trabalho consiste em investigar a relação entre as características familiares e a proficiência escolar. Para isso, estimou-se um modelo não paramétrico, o *generalized additive model* (GAM), por meio de um estimador em dois estágios, também não paramétrico. Foram utilizados os microdados do Sistema de Avaliação da Educação Básica (Saeb) 2011 e do Censo Escolar, em que as proficiências correspondem às notas do exame de aprendizagem de alunos do 5º e do 9º anos do ensino fundamental de escolas das redes pública e privada. Os resultados sugerem que as variáveis ligadas ao *background* familiar exercem grande influência no desempenho escolar dos alunos em comparação com as variáveis ligadas à escola, aos docentes e aos diretores, sobretudo para os alunos de escola pública.

Palavras-chave: desempenho escolar; modelo GAM; *background* familiar.

FAMILY BACKGROUND AND SCHOOL PERFORMANCE: A NON-PARAMETRIC APPROACH

A large theoretical and empirical literature suggests that there is an intergenerational transfer of socioeconomic characteristics of the family, such as income and education. In this sense, the objective of this work is to investigate a relation between the familiar characteristics and a school proficiency. For this, a non-parametric model, the generalized additive model (GAM), was estimated through a two stages non-parametric estimator. The micro data of Saeb 2011 and the School Census were used, corresponding to the grades of the apprenticeship examination of students in the 5th and 9th grade of public and private schools. The results suggested as variables linked to the family background exert a great influence on the students' school performance in a variable with variables related to the school, the teachers and the principals, especially for the students of the public school.

Keywords: school performance; GAM model; family background.

JEL: 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

1. Doutor em economia aplicada pelo Programa de Pós-Graduação em Economia da Universidade Federal do Rio Grande do Sul (PPGE/UFRGS). Professor no Departamento de Economia da UFPB *E-mail:* <wpsfarias@gmail.com>.

2. Pós-doutorando em desenvolvimento econômico na UFPR. Doutor em economia aplicada pelo PPGE/UFRGS. *E-mail:* <victor5491@gmail.com>.

3. Doutora em economia aplicada pelo PPGE/UFRGS. *E-mail:* <annegues.ana@gmail.com>.

escolar o país ainda permanece em situação desfavorável quando comparado a outros países. Segundo a 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 do poder público e da sociedade, tendo em vista que a educação consiste em um dos pontos-chave para uma nação próspera, com desenvolvimento sustentado.

Várias estratégias para reverter esse quadro têm sido sugeridas 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, na maior parte das vezes, o impacto de fatores ligados à escola e aos professores. Não obstante, outro fator de grande importância para o desempenho dos alunos, embora não inserido diretamente no contexto escolar, é o *background* familiar desses estudantes.

O *background* familiar consiste em um conjunto de características do ambiente familiar no qual os indivíduos se desenvolvem (educação e ocupação dos pais, renda familiar etc.). A importância das características familiares na chance de progresso escolar é um resultado bastante consolidado nas literaturas teórica e empírica, considerado como um dos principais fatores responsáveis pela “manutenção do ciclo intergeracional de pobreza” (Leon e Menezes-Filho, 2002). Tais características tendem a ser repassadas aos indivíduos e exercem forte influência sobre os seus resultados econômicos individuais, como a renda e/ou o nível de escolaridade – ver Conlisk (1974), Becker e Tomes (1979) e Goldberger (1989). Behrman, Birdsall e Székely (2000) mostram que nos países da América Latina, em particular, a educação dos pais tende a ser determinante no nível de educação alcançado pelos indivíduos, o que denota uma baixa mobilidade educacional entre as gerações. No Brasil, o estudo de Ferreira e Veloso (2006) corrobora as evidências encontradas em favor da importância do *background* familiar no nível socioeconômico e educacional dos indivíduos.

Algumas evidências empíricas apontam que esta variável consiste em um insumo tão ou mais importante que os fatores pedagógicos na função de produção educacional. Os resultados encontrados por Souza e Silva (1994), por exemplo, evidenciam que as variáveis ligadas à família tiveram um alto impacto sobre o desempenho escolar dos estudantes no Brasil. 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.

Leon e Menezes-Filho (2002), utilizando dados longitudinais da Pesquisa Mensal de Emprego (PME), verificam que o *background* familiar mostrou-se importante para explicar as chances de progressão escolar. Machado e Gonzaga (2007) analisaram, utilizando a Pesquisa Nacional de Amostra por Domicílios (Pnad), o efeito da renda e da educação dos pais sobre a probabilidade de ocorrer defasagem entre a idade e a série cursada pela criança. Segundo os autores, o jovem de melhor origem social têm melhores condições de permanecer na escola por um período mais longo. Já as crianças cujos pais têm menor nível de instrução e menor renda familiar *per capita* mostraram-se mais suscetíveis a terem defasagem idade-série.

Diante disso, o objetivo do presente artigo é fornecer mais evidências do impacto do *background* familiar sobre o rendimento dos estudantes. Para tanto, a estratégia empírica que será empregada lança mão da econometria não paramétrica para identificar o efeito dessas variáveis sobre a proficiência escolar. Em primeiro lugar, o modelo econométrico a ser estimado na presente análise é um *generalized additive model* (GAM), ou modelo aditivo generalizado, desenvolvido por Horowitz e Mammen (2004), que consiste em uma adição de funcionais desconhecidas, relacionando a variável explicada (desempenho escolar) e as variáveis explicativas. A grande vantagem deste modelo está na possibilidade da inclusão de um número maior de covariáveis, aumentando o poder de explicação e minimizando possíveis vieses de variável omitida sem, no entanto, gerar problemas relacionados à quantidade de parâmetros mensurados. Já a estimação dos parâmetros do modelo baseia-se em um método de inferência não paramétrico em dois estágios, cujas propriedades evitam problemas de dimensionalidade. Além disso, 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.

Em complemento às características familiares, o modelo inclui variáveis relacionadas à escola, ao professor, ao aluno e ao gestor, amplamente utilizadas pela literatura. No primeiro grupo destacam-se análises referentes ao ambiente escolar e às características dos profissionais de educação, tal como o nível de qualificação do professor. Biondi e Felício (2008), por exemplo, 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, Gonzaga e Firpo (2013) encontra 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, o 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. 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).⁴

Além da utilização de uma nova metodologia, esta pesquisa procurar estender a análise a outras variáveis relevantes ao sucesso educacional do aluno, como raça e gênero de professores e diretores, e 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 dos alunos nos Estados Unidos.

Para alcançar tal objetivo, os bancos de dados utilizados são os microdados do Sistema de Avaliação da Educação Básica (Saeb) 2011 e do Censo Escolar, em que se encontram sumarizadas todas as informações referentes às variáveis das escolas, dos professores, dos pais e dos alunos, bem como as notas destes nos testes de aprendizagem em língua portuguesa e matemática. Os resultados de testes padronizados como o Saeb são comumente utilizados como medida do grau de desempenho e aprendizagem dos alunos.

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

O presente estudo utiliza ferramentas de abordagem não paramétrica para investigar como os componentes do *background* familiar estão relacionados ao desempenho dos alunos. Enquanto em muitas situações os pressupostos paramétricos são razoáveis (por exemplo, recorrer à distribuição de Poisson para modelar a demanda por cuidados de saúde), em outras não é possível estabelecer um conhecimento prévio das distribuições subjacentes aos dados. Nessas situações, o uso de métodos paramétricos pode conduzir a resultados viesados, comprometendo a análise do pesquisador. Nesse sentido, é necessário

4. 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.

utilizar procedimentos que sejam insensíveis às premissas do modelo, de forma que os estimadores mantenham as suas propriedades na vizinhança de um determinado ponto da distribuição desconhecida.

Os métodos ditos não paramétricos apresentam muitas vantagens, tais como: *i)* exigem poucas suposições sobre a população a partir da qual os dados são obtidos; *ii)* permitem obter os p -valores exatos para testes e as probabilidades exatas de “cobertura” dos intervalos de confiança; *iii)* o nível de significância do teste, o poder dos testes para detectar importantes hipóteses alternativas e a variância do estimador não são sensíveis aos pressupostos do modelo; *iv)* são insensíveis à presença de *outliers*; *v)* as técnicas de reamostragem (*jackknife* e *bootstrap*) podem ser utilizadas em muitas situações em que os métodos paramétricos as tornam intratáveis; *vi)* são, geralmente, mais eficientes do que as abordagens paramétricas quando a distribuição dos dados não é normal etc. (Pagan e Ullah, 1999).

Primeiramente, consideramos um GAM. Este modelo substitui a forma linear por uma soma de funções suavizadas.

A literatura econométrica apresenta três tipos de procedimentos para se estimar os modelos GAM: o método de integração marginal (Linton e Nielsen, 1995), o método de dois estágios (Horowitz e Mammen, 2004) e o algoritmo *backfitting*. No trabalho desenvolvido aqui será utilizada a proposta de Horowitz e Mammen (2004).

Assim, seja a seguinte especificação:

$$y = f[\mu + m_1(X^1) + \dots + m_d(X^d)] + \varepsilon, \quad (1)$$

em que X^j ($j = 1, \dots, d$) é o j -ésimo elemento do vetor aleatório $X \in \mathbb{R}^d$ para $d \geq 2$; μ é uma constante desconhecida; f é conhecida; $m_1 \dots m_d$ são funções desconhecidas; e ε é o termo aleatório não observado satisfazendo $E(\varepsilon|X=x) = 0$ 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.

O objetivo empírico deste estudo é mensurar como as variáveis que capturam o *background* familiar estão relacionadas ao desempenho dos alunos. A literatura sobre o tema parece indicar que essa relação é não linear, e o mais adequado seria computá-la por meio de funções suavizadas. Essas funções neste estudo são as funções $m_1(X^1), \dots, m_d(X^d)$ e representam como a variável de interesse (y) é condicionada por meio de cada covariada. Para facilitar a exposição dos resultados, será apresentado o valor predito para cada função com base nos valores assumidos por cada covariável.

Esse estimador contorna o problema do modelo proposto por Linton e Härdle (1996), no qual era necessário que as funções aditivas m_1, \dots, m_d tivessem um número cada vez maior de derivadas à medida que a dimensão do vetor de covariáveis aumentasse.

Uma gama de trabalhos discute a estimação de funções como a apresentada em (1), tais como: Stone (1994) e Newey (1997), utilizando *splines*; e Opsomer (2000), que investiga as propriedades do processo 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 à que será utilizada nesse estudo. Porém, não consideram modelos com funções de ligação e usam o *backfitting* como algoritmo para o primeiro estágio.⁵ No entanto, o estimador utilizado aqui, ao contrário do *backfitting*, tem a vantagem de que a distribuição assintótica de cada componente aditivo é idêntica à que seria encontrada se os demais componentes fossem conhecidos.

Embora a adoção de um modelo não paramétrico forneça vantagens em termos de especificação, por outro lado é importante ter em mente que o uso de um modelo aditivo generalizado apresenta algumas limitações, entre as quais: *i*) pode sofrer uma grande perda de precisão se for mal especificado; *ii*) o modelo tende a produzir estimativas que podem ser, de forma geral, superestimadas; *iii*) é computacionalmente intensivo e muitas vezes pode não convergir.

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. No primeiro estágio, utiliza-se mínimos quadrados não lineares para se obter uma aproximação para cada m_1, \dots, m_d . Será imposta a estrutura aditiva (1) e produzidas estimativas de m_1, \dots, m_d que têm viés assintótico menor que outros procedimentos, tal como os avaliados na integração marginal. Já no segundo estágio é obtida uma estimativa linear local.

Para obter os estimadores do primeiro estágio para uma amostra de tamanho n , minimiza-se a seguinte função objetivo:

$$\min_{\theta \in \Theta_k} S_k(\theta) \equiv \frac{1}{N} \sum_{i=1}^n (y_i - f[p_k(x_i)' \theta])^2, \quad (2)$$

5. 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).

em que $\Theta_k \subset \mathbb{R}^{d+1}$ é um conjunto de parâmetros compacto e p_k são transformações do vetor de covariáveis.

O segundo estágio é a aplicação do estimador *local linear* à função (1), substituindo-se as funções desconhecidas m_1, \dots, m_d pelas estimadas em (2).

Para implementar o estimador *local linear* é necessário definir a *bandwidth*. A determinação do parâmetro de suavização (*bandwidth*) é obtida por meio da minimização do erro quadrático médio assintótico integrado, ou seja, escreve-se o erro quadrático médio (viés ao quadrado + variância), integra-se a expressão resultante e deriva-se com relação ao parâmetro de interesse, a *bandwidth*. Desse modo, estima-se o valor de h que minimiza a seguinte expressão $n^{2/5}[\hat{m}_1(x^1) - m_1(x^1)]$, e é um parâmetro de suavização ótimo que permite um balanceamento adequado entre o viés do estimador e a sua variância.

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 Saeb 2011 e do Censo Escolar. Para tanto, foram utilizados os dados do 5º e do 9º anos 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 e 9º anos 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 competências. 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 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.

Analizou-se a proficiência média em português e em matemática, separadamente para o 5º e o 9º anos, tanto das escolas públicas quanto das escolas privadas, de modo a captar possíveis diferenças nos efeitos entre os níveis, sistemas de ensino e pelo campo de saber avaliado, assim como considerar as características dos professores das áreas específicas. Dessa forma, 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 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 os professores são usadas como variáveis explicativas. A próxima subseção apresenta as estatísticas descritivas de tais variáveis, como forma de fornecer uma visão geral sobre os dados usados nas estimações.

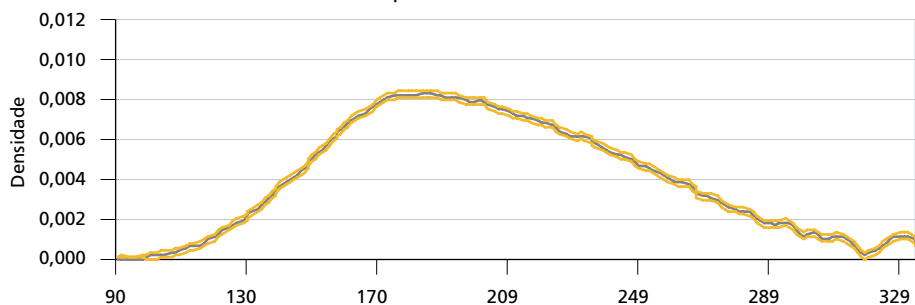
3.2 Estatísticas descritivas

As tabelas A.1 e A.2, no apêndice, apresentam algumas estatísticas descritivas para os dados utilizados. 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 anos. Tais características podem ser observadas no gráfico 1. Neste apresenta-se a densidade *kernel* estimada da proficiência dos alunos por série e tipo de escola.

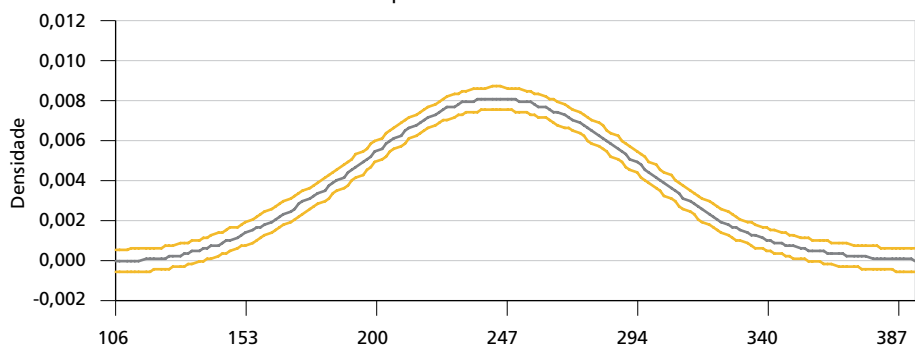
GRÁFICO 1

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

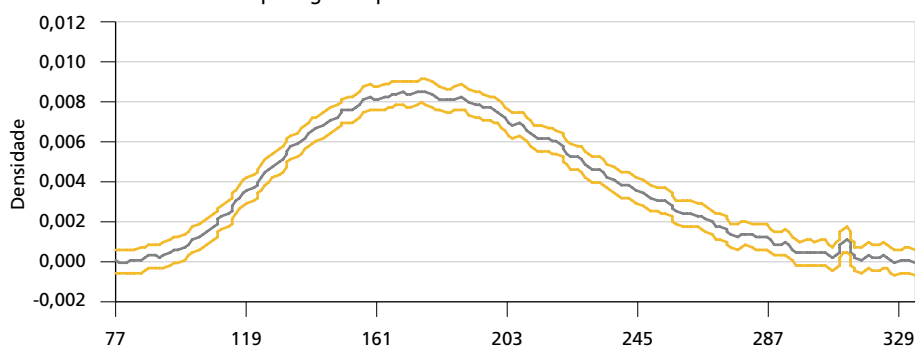
1A – Proficiência em matemática – pública – 5º ano

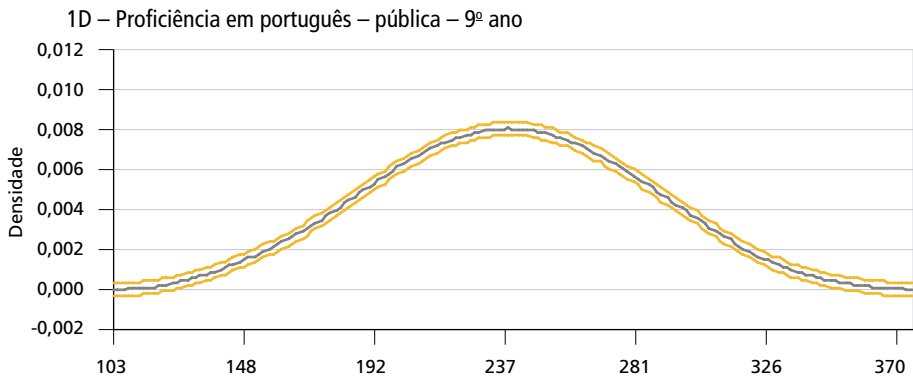


1B – Proficiência em matemática – pública – 9º ano



1C – Proficiência em português – pública – 5º ano





Elaboração dos autores.

Obs.: A densidade é obtida pelo procedimento proposto por Bernacchia e Pigolotti (2011). As linhas em amarelo representam as bandas de variabilidade obtidas por *bootstrap* com 1.000 replicações.

Para garantir a confiabilidade dessas estimações e verificar a significância delas, construíram-se bandas de variabilidade (linhas amarelas) a partir de *bootstrap* não paramétrico. Foram geradas 1.000 amostras e, em seguida, procedeu-se à estimação dos 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 ao *background* familiar, de particular interesse para o presente estudo, algumas características chamam a atenção. Para a rede particular de ensino, a proporção de alunos que têm pais que nunca estudaram é praticamente zero, e para a rede pública essa proporção não ultrapassa 3%. Outro aspecto que chama a 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, nas duas séries. O incentivo dos pais⁶ para os filhos estudarem é também muito maior para os alunos de escolas privadas, indicando uma maior preocupação desses pais com a vida escolar e a importância do ambiente familiar no desempenho dos filhos.

No tocante às características dos alunos, a maioria do 5º ano é do sexo masculino (52,5% na escola pública e 51,5% na escola privada), com uma média de idade de 10,9 anos e um grande percentual de alunos não brancos (71,4% e 58,5% para as escolas pública e privada, respectivamente). Já para o 9º ano, 46% são do sexo masculino na escola pública e 48% na escola privada, enquanto o percentual de alunos não brancos diminui em relação aos estudantes do 5º ano (68,8% para a

6. Foi obtido por meio da técnica de análise de componentes principais utilizando as seguintes variáveis, indicando se os pais: incentivam para estudar, incentivam para fazer o dever de casa, incentivam para ler, incentivam para ir à escola e conversam sobre a escola.

escola pública e 53% para a escola privada). Quanto ao índice socioeconômico,⁷ percebe-se um valor médio muito maior para os alunos de escolas privadas (tanto no 5º quanto no 9º anos) em comparação aos alunos das escolas públicas.

Considerando as variáveis da escola, a grande maioria dos professores possui curso superior, já fez algum curso de atualização e possui dez ou mais anos de experiência em sala de aula. A maior parte dos gestores também é do sexo masculino e tem nível superior. Na rede pública, cerca de 24% dos gestores têm mais de dez anos de experiência, provavelmente pela escolha feita por meio de eleições, o que aumenta a rotatividade dos diretores. Cerca de 1% das escolas públicas e 12% das escolas privadas utilizam provas de seleção; 90% e 97% situam-se em áreas urbanas, respectivamente. Por fim, observa-se que 48% das turmas são homogêneas com relação à idade nas escolas públicas, enquanto nas escolas privadas esse percentual é de 56%.

3.3 Resultados estimados

Esta subseção 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, Zhang e Zhang (2001). Nesse teste compara-se a soma dos quadrados dos resíduos dos modelos paramétrico e não paramétrico sob a hipótese nula de igualdade entre os dois métodos. Fan e Yao (2003) apresentam a distribuição assintótica do teste por meio de *bootstrap*. A rejeição da hipótese nula evidenciada pelo *p-value* na tabela 1 corrobora a utilização do método de estimação não paramétrico proposto na seção 2.

A tabela 2 contém as estimativas do modelo aditivo generalizado. Os parâmetros contidos nessa tabela representam como as proficiências respondem na média a todos os valores possíveis das covariadas. As variáveis consideradas para a análise do desempenho escolar podem ser classificadas de modo geral como associadas às características do aluno, do professor, do gestor e da escola. Dentro das variáveis do aluno destaca-se o *background* familiar. Como o objetivo deste estudo é focar o papel desempenhado pelas características das famílias dos alunos, as estimativas da relação entre a proficiência e todas as variáveis explicativas não relacionadas ao *background* familiar serão apresentadas separadamente.

7. Foi obtido por meio da técnica de análise de componentes principais. Para o índice socioeconômico foram utilizadas as respostas a seis perguntas: número de TVs (zero, um, dois ou mais), número de geladeiras (zero, um ou mais), número de carros (zero, um, dois ou mais), computadores (zero se não tem, um se não tem acesso à internet e dois se tem acesso à internet), número de dormitórios (zero, um, dois, três ou mais) e se há empregada doméstica em sua casa (zero se não tem, um se for diarista, dois se for uma e três para mais de duas empregadas).

TABELA 1
Teste de especificação não paramétrica – H0: perfeita especificação paramétrica

	Pública			
	Matemática		Português	
	5º ano	9º ano	5º ano	9º ano
Estatística	1.237,33	2.055,04	1.500,57	2.438,56
<i>P-value</i>	[0,0000]	[0,0000]	[0,0000]	[0,0000]
	Privada			
	Matemática		Português	
	5º ano	9º ano	5º ano	9º ano
Estatística	277,06	363,70	334,96	527,57
<i>P-value</i>	[0,0000]	[0,0000]	[0,0000]	[0,0000]

Elaboração dos autores.

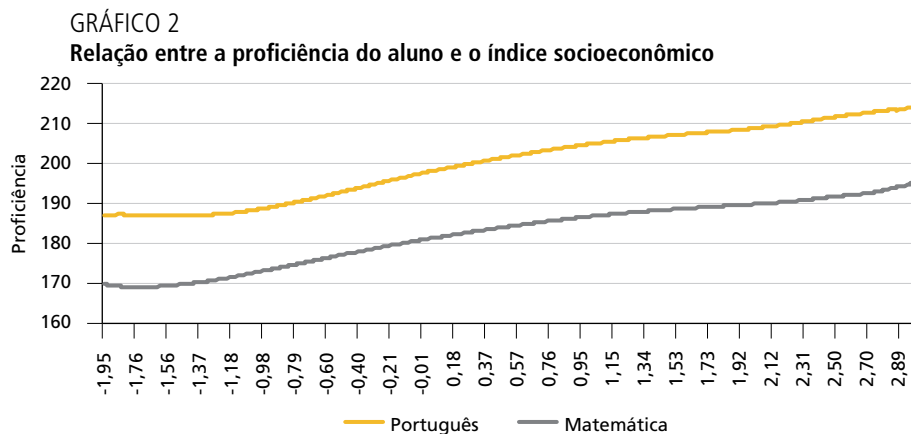
Dessa forma, a subseção 3.3.1 apresenta os resultados alusivos à relação da proficiência dos alunos com as características do aluno e seu *background* familiar, enquanto a subseção 3.3.2 apresenta os resultados referentes às variáveis do ambiente escolar (diretor, professor, escola). As estimativas encontram-se sumarizadas nas tabelas 1 e A.3 (para os métodos não paramétrico e paramétrico, respectivamente), referentes aos alunos de escolas pública e privada.

3.3.1 *Background* familiar

Nesta subseção são analisados os resultados do impacto do *background* familiar sobre as proficiências dos alunos, o qual abrange as condições financeiras da família e a escolaridade dos pais. Além das características familiares, a análise também engloba atributos intrínsecos aos alunos, como raça, sexo e seu histórico escolar. É importante ressaltar que os gráficos construídos nesta subseção são inferidos com base nas estimativas da tabela 2, por meio do modelo aditivo generalizado. A apresentação nesse formato serve para melhor visualizar os atributos familiares e sua relação com a proficiência dos alunos.

Em geral, pais altamente educados podem decidir investir mais do seu tempo e energia na educação dos seus filhos, ou eles podem escolher interagir suas atividades diárias com os seus filhos de forma a ajudá-los a serem bem-sucedidos na escola. Ao mesmo tempo, famílias com melhores condições socioeconômicas geralmente serão capazes de proporcionar mais recursos educacionais em casa ou de escolher as escolas que irão fornecer-lhes esses recursos. Nesse sentido, inclui-se um termo que procura capturar a relação entre condições financeiras e desempenho

acadêmico e encontra-se uma correlação positiva com o desempenho dos alunos, conforme mostra o gráfico 2.



Elaboração dos autores.

Obs.: Dados do 5º ano para as escolas públicas.

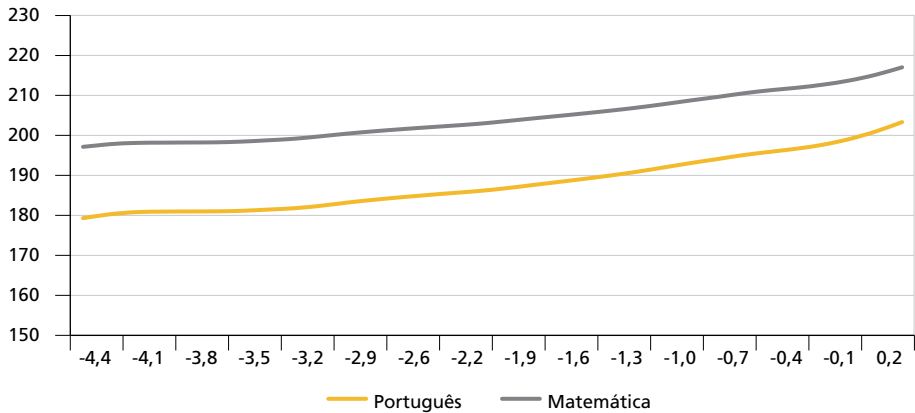
Na rede pública, o *status* socioeconômico e o fato de o aluno ser branco contribuem, em média, para uma maior proficiência em aprendizagem, corroborando resultados conhecidos da literatura. Para os alunos da rede privada, a situação socioeconômica apresentou impacto estatisticamente significativo apenas na proficiência em língua portuguesa daqueles pertencentes à 9ª série.

Na mesma direção, verificou-se que o incentivo dos pais é um elemento fundamental sobre o desempenho dos alunos. O gráfico 3 mostra que existe uma correlação positiva entre essas variáveis. Quando se discrimina por variável, conforme o gráfico 4, observa-se que os pais de alunos de escolas privadas participam de forma mais ativa na vida escolar dos alunos, porém a proporção de pais (escolas públicas e privadas) que incentivam os alunos a estudar é comparativamente menor com relação às demais formas de interação pais-alunos.

Tais resultados corroboram evidências presentes na literatura do papel do *background* familiar. Assim como mostram Leon e Menezes-Filho (2002), nossos resultados evidenciam que crianças com um *background* “ruim” tendem a ter um desempenho abaixo daquelas de melhor nível socioeconômico. Machado e Gonzaga (2007) mostram que estas crianças também estão mais vulneráveis a ter problemas de defasagem entre idade e série.

GRÁFICO 3

Relação entre a proficiência do aluno e o incentivo dos pais: geral

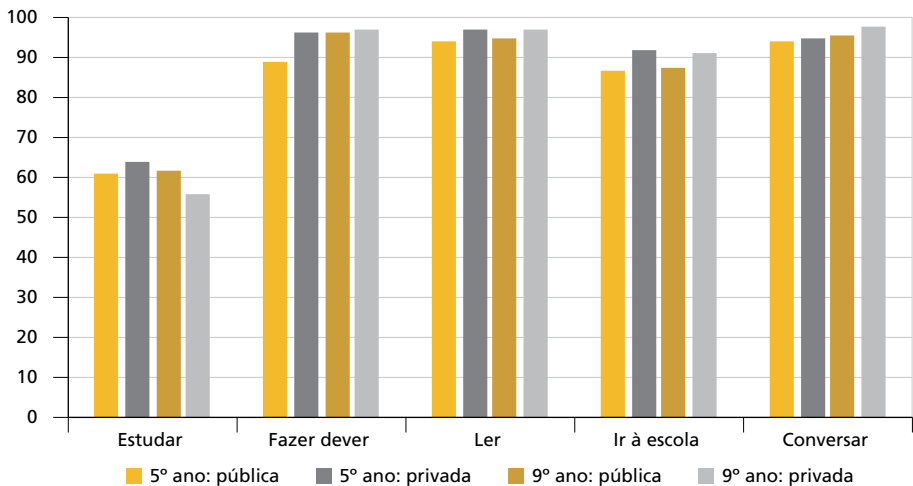


Elaboração dos autores.

Obs.: Dados do 5º ano para as escolas públicas.

GRÁFICO 4

Relação entre a proficiência do aluno e o incentivo dos pais: tipo de incentivo (Em %)

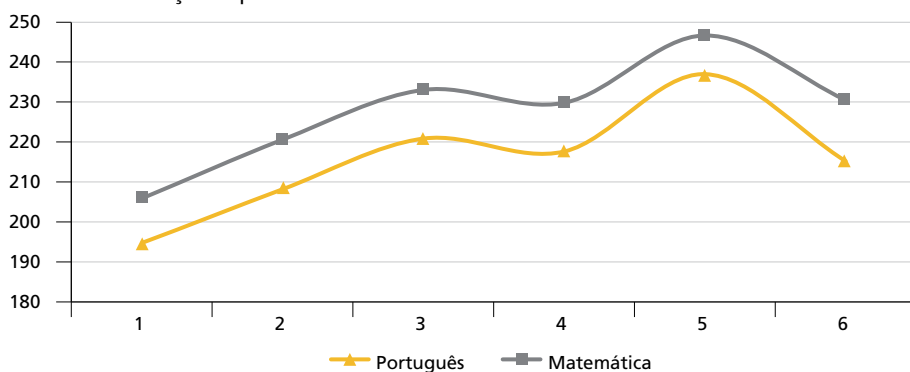
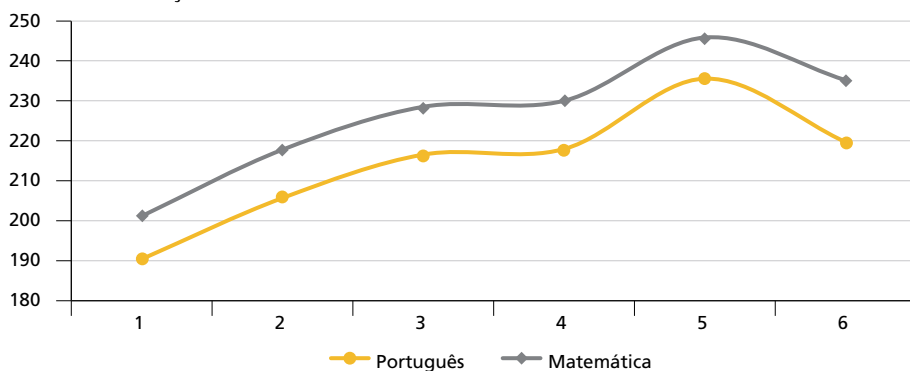


Elaboração dos autores.

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). O gráfico 5 exibe a relação que existe entre o desempenho dos alunos no Saeb e o histórico educacional dos pais.

GRÁFICO 5

Relação entre a proficiência do aluno e a educação dos pais**5A – Educação do pai****5B – Educação da mãe**

Elaboração dos autores.

Obs.: 1. Dados do 5º ano para as escolas públicas.

2. Nos eixos x dos gráficos, os números significam: 1 – nunca estudou; 2 – não completou a 4ª série; 3 – completou a 4ª série, mas não completou a 8ª série; 4 – completou a 8ª série, mas não completou o ensino médio; 5 – completou o ensino médio, mas não completou a faculdade; 6 – completou a faculdade.

No que se refere à magnitude do impacto, é possível observar que o método não paramétrico (tabela 1), em geral, apresenta coeficientes menores quando comparado aos coeficientes do método paramétrico observado na tabela A.3. Isso evidencia que o método paramétrico pode estar superestimando o verdadeiro efeito das variáveis que afetam a proficiência, entre outros fatores, por considerar uma forma funcional e supor uma distribuição que não traduz a verdadeira relação entre os dados.

3.3.2 Ambiente escolar

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, consonante 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 mostrou-se 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, seja 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 a 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, Gonzaga e Firpo, 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.

No que concerne ao tamanho da turma, Camargo e Porto Júnior (2012) 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 mostrou-se 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 ela pode oferecer. Dessa forma, famílias com melhores condições econômicas, avaliadas pelo índice socioeconômico, tenderiam a escolher as melhores escolas, constituindo um processo de autosseleção 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 as variáveis *dummy* para características demográficas.

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, porém também positivo (com exceção do 9º ano em português). Com relação à raça, o fato de o professor ser da cor branca impacta negativamente a escola privada e o 9º ano da escola pública, sendo que em ambas as redes de ensino o coeficiente é não significativo. Becker e Barreto (2014) chega à conclusão de que o gênero do professor e sua diferença em relação ao do aluno não aumenta a chance de o aluno no interesse em estudar, sendo mais uma preferência do próprio aluno.

Com relação à experiência do professor, verifica-se uma relação significativa (e negativa) de o professor ter dez anos ou mais de experiência apenas para o 5º ano da rede de ensino pública. A relação negativa parece a princípio contraditória, mas significa que se, por um lado, a experiência denota maior domínio e segurança ao professor, por outro pode criar um abismo de idade, gerando divergência em termos de linguagem ou formas pedagógicas com os alunos. Para o caso de o professor ter pós-graduação (especialização, mestrado ou doutorado), a relação também é

significativa apenas para o mesmo nível (5^o 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.

Com relação às características pessoais do diretor, é verificada significância estatística novamente apenas para a escola pública. Um fato que chama a atenção são os coeficientes significativos para o diretor ser da cor branca e ter mais de dez anos de experiência apresentar sinais negativos. Porém, os resultados fazem sentido se for analisado que a maior parte dos alunos é não branca e que, quando 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 primordialmente o desempenho dos alunos das escolas públicas, dado que estes estão mais suscetíveis a choques devido ao fato de provavelmente possuírem um *background* familiar desfavorável em relação aos alunos da escola privada.

A análise acima é toda baseada na estimação do método não paramétrico com os resultados expostos na tabela 1. No entanto, quando se observa os resultados do método paramétrico na tabela A.3, a grande maioria dos coeficientes relacionados ao impacto de variáveis ligadas ao professor, ao diretor e à escola é, em geral, significativa, ao contrário do que foi exposto para o método aditivo generalizado. Além disso, a magnitude de muitos coeficientes é bastante elevada quando comparada à estimação não paramétrica, o que novamente corrobora para uma possível superestimação e má especificação dos resultados. Isso, por sua vez, leva a conclusões errôneas sobre o real efeito dessas variáveis e interpretações equivocadas sobre a sua importância.

Em suma, os resultados do método paramétrico apontam para coeficientes de maior magnitude e significativos para a grande parte das variáveis utilizadas, enquanto no GAM os efeitos são menores e a significância predomina principalmente nas variáveis referentes ao *background* familiar, sendo muitas outras variáveis (escola, diretor, professor) sem significância estatística. Essas estimativas corroboram a importância das variáveis relacionadas à família como fundamentais no resultado educacional dos indivíduos.

Dado o teste realizado, a maior flexibilização e as vantagens apontadas do método não paramétrico (sobretudo para o presente caso, com uma quantidade considerável de regressores), somos levados a desconfiar dos resultados apresentados pelo modelo paramétrico frente ao primeiro, sobretudo por má especificação da forma funcional e da relação entre as variáveis, bem como de uma possível superestimação dos resultados referentes a determinados grupos de variáveis.

TABELA 2
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	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)				

(Continua)

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	5º ano	9º ano	5º ano	9º ano			
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)	0,0917 (0,166)	-0,0517 (0,155)	0,0132 (0,166)	-0,116 (0,156)	-0,0820 (1,390)	-0,299 (1,425)	-0,160 (1,045)
Professor: branco															
Professor com mais de dez anos de experiência															
Professor fez curso de atualização															
Professor tem especialização, mestrado ou doutorado															
Professor com curso superior															
Diretor: homem															
Diretor: branco															
Diretor tem mais de dez anos de experiência															
Diretor tem curso superior															
Qualidade da escola															
Violência externa															
Violência interna															

(Continua)

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
Prova de seleção	-0,509 (0,704)	-0,819 (0,593)	-0,450 (0,707)	-0,0594 (0,595)	-0,377 (1,868)	0,192 (1,290)	0,00564 (1,915)	0,226 (1,267)
Turmas homogêneas quanto à idade	0,238 (0,188)	0,916*** (0,177)	0,343* (0,188)	0,836*** (0,178)	0,200 (1,536)	-0,233 (1,173)	0,863 (1,574)	0,254 (1,152)
Turmas homogêneas quanto ao rendimento escolar	0,332 (0,309)	0,331 (0,346)	0,310 (0,310)	0,297 (0,347)	-0,225 (2,676)	-1,266 (2,369)	1,413 (2,743)	-1,037 (2,327)
Turmas heterogêneas quanto à idade	0,544** (0,270)	0,521** (0,249)	0,511* (0,271)	0,485* (0,249)	-0,667 (3,393)	0,560 (2,010)	0,270 (3,478)	1,228 (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)

Elaboração dos autores.

Obs.: 1. ***, ** e * = $p < 0,01$, $p < 0,5$ e $p < 0,1$.

2. Erro-padrão entre parênteses.

4 CONSIDERAÇÕES FINAIS

O objetivo deste estudo foi investigar a relação entre o *background* familiar e a proficiência escolar. Para tanto, empregou-se um modelo não paramétrico, permitindo, assim, a identificação de possíveis padrões de comportamento não lineares da variável dependente condicionada aos fatores explicativos. A estimação foi feita por meio de um método proposto por Horowitz e Mammen (2004), que consiste em uma adição de funcionais desconhecidas em dois estágios. A principal vantagem deste estimador é a ausência de problema de dimensionalidade na matriz de covariadas, gerado pelo grande número de variáveis explicativas que dificulta a obtenção de estimativas confiáveis.

Os resultados indicaram que, para as escolas privadas, somente o conjunto de variáveis relacionadas ao *background* familiar mostraram-se significativas. Para os alunos da rede pública, foram encontradas evidências que confirmam o processo de persistência intergeracional na educação, uma vez que os alunos com histórico de reprovações e filhos de pais que nunca frequentaram a escola tendem a apresentar um pior desempenho. Outra evidência importante diz respeito à influência da violência sobre o desempenho escolar. Os resultados demonstram que existe uma correlação negativa entre violência e proficiência para os alunos do 5º ano das escolas públicas.

Em linhas gerais, a pesquisa apontou a relevância que questões socioeconômicas, como as condições das famílias e a violência nas escolas, têm sobre o aprendizado dos alunos. Embora reformas nos currículos escolares e uma melhor formação dos professores sejam ações importantes para melhorar a qualidade da educação brasileira, discussões acerca do papel da família e de soluções para a violência constituem aspectos fundamentais para que estas ações convertam-se em benefícios reais ao desempenho dos alunos.

REFERÊNCIAS

- BARROS, R. P. *et al.* **Determinantes do desempenho educacional no Brasil**. Rio de Janeiro: Ipea, 2001. (Texto para Discussão, n. 834).
- BARROS, R. P.; MENDONÇA, R. **Consequências da repetência sobre o desempenho educacional**. Brasília: MEC, 1998. (Projeto de Educação Básica para o Nordeste).
- BECKER, J. C.; BARRETO, M. Ways to go: men's and women's support for aggressive and nonaggressive confrontation of sexism as a function of gender identification. **Journal of Social Issues**, v. 70, n. 4, p. 668-686, 2014.
- BECKER, G. S.; TOMES, N. An equilibrium theory of the distribution of income and intergenerational mobility. **Journal of Political Economy**, v. 87, p. 1153-1189, 1979.

BEHRMAN, J.; BIRDSALL, N.; SZÉKELY, M. Intergenerational mobility in Latin America: deeper markets and better schools make a difference. *In*: BIRDSALL, N.; GRAHAM, C. (Eds.). **New markets, new opportunities? Economic and social mobility in a changing world**. Washington: Brookings Institution Press and the Carnegie Endowment for International Peace, 2000.

BERNACCHIA, A.; PIGOLOTTI, S. Self-consistent method for density estimation. **Journal of the Royal Statistical Society B**, v. 73, n. 3, p. 407-422, 2011.

BIONDI, R.; FELÍCIO, F. **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.; PORTO JÚNIOR, S. D. S. **O efeito do tamanho da turma sobre o desempenho escolar: uma avaliação do impacto da “enturmação” no Ensino Fundamental do Rio Grande do Sul**. 2012. Dissertação (Mestrado) – Universidade Federal do Rio Grande do Sul, Porto Alegre, 2012.

CONLISK, J. Can equalization of opportunity reduce social mobility? **American Economic Review**, v. 64, p. 80-90, 1974.

CURI, A. Z.; MENEZES-FILHO, N. A. A relação entre educação pré-primária, salários, escolaridade e proficiência escolar no Brasil. **Estudos Econômico**, v. 39, n. 4, p. 811-850, 2009.

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

FAN, J.; YAO, Q. **Nonlinear time series: nonparametric and parametric methods**. New York: Springer, 2003.

FAN, J.; ZHANG, C.; ZHANG, J. Generalized likelihood ratio statistics and Wilks phenomenon. **The Annals of Statistics**, v. 29, p. 153-193, 2001.

FERREIRA, S.; VELOSO, F. Mobilidade intergeracional de educação no Brasil. **Pesquisa e Planejamento Econômico**, v. 33, n. 3, p. 481-513, 2003.

_____; _____. Intergenerational mobility of wages in Brazil. **Brazilian Review of Econometrics**, v. 26, n. 2, p. 181-211, 2006.

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

GOLDBERGER, A. Economic and mechanical models of intergenerational transmission. **American Economic Review**, v. 74, p. 504-513, 1989.

HASTIE, T. J.; TIBSHIRANI, R. J. **Generalized additive models**. London: Chapman and Hall, 1990.

HECKMAN, J.; JACOBS, B. **Policies to create and destroy human capital in Europe**. Bonn: IZA, 2009. (IZA Discussion Papers, n. 4680).

HIDALGO-HIDALGO, M. On the optimal allocation of students when peer effects are at work: tracking vs. mixing. **Series**, v. 2, n. 1, p. 31-52, 2011.

HOROWITZ, J. L.; KLEMELÄ, J.; MAMMEN, E. **Optimal estimation in additive regression models**. Mannheim: University of Mannheim, 2002. (Working Paper).

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

LEON, F. L.; MENEZES-FILHO, N. A. Reprovação, avanço e evasão escolar no Brasil. **Pesquisa e Planejamento Econômico**, v. 32, n. 3, p. 417-452, 2002.

LINTON, O. B.; HÄRDLE, W. Estimating additive regression models with known links. **Biometrika**, v. 83, p. 529-540, 1996.

LINTON, O.; NIELSEN, J. P. A kernel method of estimating structured nonparametric regression based on marginal integration. **Biometrika**, v. 82, n. 1, p. 93-100, 1995.

MACHADO, A. F. *et al.* Qualidade do ensino em matemática: determinantes do desempenho de alunos em escolas públicas estaduais mineiras. **Economia**, v. 9, n. 1, p. 23-45, 2008.

MACHADO, D. C.; GONZAGA, G. O impacto dos fatores familiares sobre a defasagem idade-série de crianças no Brasil. **Revista Brasileira de Economia**, v. 61, n. 4, p. 449-476, 2007.

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

MAMMEN, E.; LINTON, O. B.; NIELSEN, J. P. The existence and asymptotic properties of backfitting projection algorithm under weak conditions. **The Annals of Statistics**, v. 27, p. 1443-1490, 1999.

MENEZES-FILHO, N. *et al.* Avaliando o impacto da progressão continuada nas taxas de rendimento e desempenho escolar do Brasil. *In: SEMINÁRIO DE ECONOMIA DE BELO HORIZONTE*, 6., Belo Horizonte, 2008. **Anais...** Belo Horizonte: SEBH, 2008.

NEWBY, W. K. Convergence rates and asymptotic normality for series estimators. **Econometrics**, . 79, p. 147-168, 1997.

OPSOMER, J. D. Asymptotic properties of backfitting estimators. **Journal of Multivariate Analysis**, v. 73, p. 166-179, 2000.

OPSOMER, J. D.; RUPPERT, D. **On the existence and asymptotic properties of backfitting estimators**. [s.l.]: [s.n.], 1997. Preprint.

PAGAN, A.; ULLAH, A. **Nonparametric econometrics**. Cambridge: Cambridge University Press, 1999.

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

STEELE, C. M.; ARONSON, J. Stereotype threat and the intellectual test-performance of African Americans. **Journal of Personality and Social Psychology**, v. 69, n. 5, p. 797-811, 1995.

STONE, C. J. The use of polynomial splines and their tensor products in multivariate function estimation. **The Annals of Statistics**, v. 22, p. 118-184, 1994.

BIBLIOGRAFIA COMPLEMENTAR

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

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

CARROLL, B. R. **The effects of school violence and crime on academic achievement**. Davidson: Davidson College, 2006.

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

GROGGER, J. Local violence and educational attainment. **Journal of Human Resources**, v. 32, n. 4, p. 659-682, 1997.

MILLER, A.; CHANDLER, K. **Violence in US public schools: 2000 School Survey on Crime and Safety**. Washington: National Center for Education Statistics, 2005.

APÊNDICE

TABELA A.1
Estatísticas descritivas: 5º ano do ensino fundamental

	Ensino público		Ensino privado	
	Média	Desvio-padrão	Média	Desvio-padrão
Proficiência em português	188,80	45,78	229,00	45,78
Proficiência em matemática	208,00	47,28	249,20	47,28
Aluno homem	0,525	0,499	0,515	0,499
Aluno branco	0,284	0,451	0,415	0,451
Idade do aluno	10,910	1,206	10,270	1,206
Desvio-padrão da idade na turma	0,933	0,459	0,574	0,459
Índice socioeconômico	-0,053	0,804	0,924	0,804
Pais nunca estudaram	0,025	0,155	0,002	0,155
Já foi reprovado	0,312	0,463	0,095	0,463
Iniciou estudos na pré-escola	0,385	0,487	0,338	0,487
Iniciou estudos na primeira série	0,219	0,413	0,116	0,413
Tamanho da família	3,527	0,750	3,213	0,750
Incentivo para estudar	-0,070	0,863	0,105	0,863
Tamanho da turma	10,820	4,268	14,530	4,268
Professor: homem	0,160	0,366	0,152	0,366
Professor: branco	0,426	0,494	0,484	0,494
Professor com mais de dez anos de experiência	0,695	0,461	0,641	0,461
Professor fez curso de atualização	0,856	0,351	0,889	0,351
Professor tem especialização, mestrado ou doutorado	0,450	0,498	0,403	0,498
Professor com curso superior	0,520	0,500	0,664	0,500
Diretor: homem	0,273	0,446	0,136	0,446
Diretor: branco	0,499	0,500	0,573	0,500
Diretor tem mais de dez anos de experiência	0,248	0,432	0,503	0,432
Diretor tem curso superior	0,830	0,120	0,910	0,120
Qualidade da escola	0,154	0,847	0,228	0,847
Violência externa	0,119	0,972	-0,531	0,972
Violência interna	0,165	1,017	-0,621	1,017
Prova de seleção	0,012	0,107	0,128	0,107
Turmas homogêneas quanto à idade	0,480	0,500	0,563	0,500
Turmas homogêneas quanto ao rendimento escolar	0,041	0,198	0,062	0,198
Turmas heterogêneas quanto à idade	0,099	0,299	0,035	0,299
Turmas heterogêneas quanto ao rendimento escolar	0,178	0,383	0,150	0,383
Diferença de sexo entre aluno e professor	0,482	0,500	0,503	0,500
Diferença de raça entre aluno e professor	0,404	0,491	0,413	0,491
Região urbana	0,906	0,292	0,973	0,292
Nordeste	0,312	0,463	0,237	0,463
Norte	0,160	0,367	0,298	0,367
Sul	0,104	0,305	0,055	0,305
Centro-Oeste	0,043	0,203	0,078	0,203

Elaboração dos autores.

TABELA A.2
Estatísticas descritivas: 9º ano do ensino fundamental

	Ensino público		Ensino privado	
	Média	Desvio-padrão	Média	Desvio-padrão
Proficiência em português	240,20	45,78	282,50	45,78
Proficiência em matemática	246,60	47,28	298,30	47,28
Aluno homem	0,460	0,499	0,480	0,499
Aluno branco	0,312	0,451	0,470	0,451
Idade do aluno	10,850	1,206	10,320	1,206
Desvio-padrão da idade na turma	0,925	0,459	0,675	0,459
Índice socioeconômico	0,010	0,804	0,991	0,804
Pais nunca estudaram	0,019	0,155	0,001	0,155
Já foi reprovado	0,337	0,463	0,129	0,463
Iniciou estudos na pré-escola	0,460	0,487	0,429	0,487
Iniciou estudos na primeira série	0,210	0,413	0,073	0,413
Tamanho da família	3,820	0,750	3,758	0,750
Incentivo para estudar	0,050	0,863	0,099	0,863
Tamanho da turma	16,300	4,268	20,650	4,268
Professor: homem	0,268	0,366	0,280	0,366
Professor: branco	0,474	0,494	0,557	0,494
Professor com mais de dez anos de experiência	0,669	0,461	0,646	0,461
Professor fez curso de atualização	0,814	0,351	0,854	0,351
Professor tem especialização, mestrado ou doutorado	0,399	0,498	0,386	0,498
Professor com curso superior	0,861	0,500	0,824	0,500
Diretor: homem	0,273	0,446	0,136	0,446
Diretor: branco	0,499	0,500	0,573	0,500
Diretor tem mais de dez anos de experiência	0,248	0,432	0,503	0,432
Diretor tem curso superior	0,830	0,120	0,910	0,120
Qualidade da escola	0,154	0,847	0,228	0,847
Violência externa	0,119	0,972	-0,531	0,972
Violência interna	0,165	1,017	-0,621	1,017
Prova de seleção	0,012	0,107	0,128	0,107
Turmas homogêneas quanto à idade	0,480	0,500	0,563	0,500
Turmas homogêneas quanto ao rendimento escolar	0,041	0,198	0,062	0,198
Turmas heterogêneas quanto à idade	0,099	0,299	0,035	0,299
Turmas heterogêneas quanto ao rendimento escolar	0,178	0,383	0,150	0,383
Diferença de sexo entre aluno e professor	0,482	0,500	0,503	0,500
Diferença de raça entre aluno e professor	0,404	0,491	0,413	0,491
Região urbana	0,906	0,292	0,973	0,292
Nordeste	0,312	0,463	0,237	0,463
Norte	0,160	0,367	0,298	0,367
Sul	0,104	0,305	0,055	0,305
Centro-Oeste	0,043	0,203	0,078	0,203

Elaboração dos autores.

TABELA A.3
Resultados estimados do modelo paramétrico

Variáveis	Escolas públicas						Escolas privadas						
	Matemática			Português			Matemática			Português			
	5º ano	9º ano	9º ano	5º ano	9º ano	9º ano	5º ano	9º ano	9º ano	5º ano	9º ano	9º ano	
Aluno homem	6,015*** (0,214)	9,012*** (0,163)	-9,610*** (0,207)	-13,178*** (0,160)	7,155*** (1,718)	9,976*** (1,094)	-11,865*** (1,737)	-10,433*** (1,042)	5,268*** (0,978)	4,962*** (1,312)	9,976*** (1,094)	-11,865*** (1,737)	-10,433*** (1,042)
Aluno branco	3,310*** (0,177)	5,169*** (0,162)	3,267*** (0,172)	5,442*** (0,159)	5,774*** (1,298)	4,458*** (1,027)	4,962*** (1,312)	5,268*** (0,978)	5,268*** (0,978)	4,962*** (1,312)	4,458*** (1,027)	4,962*** (1,312)	5,268*** (0,978)
Idade do aluno	-0,969*** (0,080)	-4,771*** (0,081)	-1,064*** (0,078)	-3,783*** (0,079)	-4,104*** (1,016)	-4,122*** (0,810)	-3,338*** (1,027)	-2,407*** (0,772)	-2,407*** (0,772)	-4,122*** (0,810)	-3,338*** (1,027)	-3,338*** (1,027)	-2,407*** (0,772)
Desvio-padrão da idade na turma	-6,445*** (0,197)	-5,104*** (0,221)	-6,129*** (0,191)	-6,324*** (0,217)	-4,378 (2,773)	-18,403*** (2,542)	-2,599 (2,803)	-18,498*** (2,422)	-18,498*** (2,422)	-4,378 (2,773)	-18,403*** (2,542)	-2,599 (2,803)	-18,498*** (2,422)
Índice socioeconômico	4,669*** (0,106)	4,762*** (0,103)	3,589*** (0,103)	4,509*** (0,101)	4,552*** (0,813)	4,586*** (0,664)	3,030*** (0,822)	1,677*** (0,632)	1,677*** (0,632)	4,552*** (0,813)	4,586*** (0,664)	3,030*** (0,822)	1,677*** (0,632)
Pais nunca estudaram	-11,118*** (0,509)	-10,021*** (0,528)	-10,001*** (0,494)	-10,031*** (0,518)	-19,810 (14,669)	-21,404 (15,764)	-28,288* (14,829)	-17,759 (15,018)	-17,759 (15,018)	-19,810 (14,669)	-21,404 (15,764)	-28,288* (14,829)	-17,759 (15,018)
Já foi reprovado	-18,310*** (0,200)	-14,718*** (0,193)	-18,315*** (0,195)	-14,588*** (0,190)	-25,138*** (2,247)	-28,204*** (1,788)	-25,223*** (2,272)	-25,519*** (1,703)	-25,519*** (1,703)	-25,138*** (2,247)	-28,204*** (1,788)	-25,223*** (2,272)	-25,519*** (1,703)
Iniciou estudos na pré-escola	5,943*** (0,176)	6,398*** (0,164)	4,536*** (0,171)	5,880*** (0,160)	-2,091 (1,364)	0,767 (1,037)	-2,677* (1,379)	1,379 (0,988)	1,379 (0,988)	-2,091 (1,364)	0,767 (1,037)	-2,677* (1,379)	1,379 (0,988)
Iniciou estudos na primeira série	-3,647*** (0,210)	0,140 (0,205)	-3,530*** (0,204)	-0,886*** (0,201)	-14,313*** (2,006)	-9,534*** (1,986)	-12,333*** (2,028)	-11,707*** (1,892)	-11,707*** (1,892)	-14,313*** (2,006)	-9,534*** (1,986)	-12,333*** (2,028)	-11,707*** (1,892)
Tamanho da família	-5,636*** (0,105)	-2,560*** (0,158)	-6,463*** (0,102)	-3,943*** (0,155)	-5,166*** (0,762)	0,118 (1,006)	-6,398*** (0,770)	-1,955** (0,959)	-1,955** (0,959)	-5,166*** (0,762)	0,118 (1,006)	-6,398*** (0,770)	-1,955** (0,959)
Incentivo para estudar	8,012 (0,092)	2,114*** (0,100)	8,810*** (0,089)	2,450*** (0,098)	4,877*** (0,984)	1,233 (0,807)	6,414*** (0,995)	2,988*** (0,769)	2,988*** (0,769)	4,877*** (0,984)	1,233 (0,807)	6,414*** (0,995)	2,988*** (0,769)
Tamanho da turma	0,509*** (0,019)	0,614*** (0,014)	0,424*** (0,018)	0,549*** (0,014)	0,463*** (0,129)	0,483*** (0,093)	0,286** (0,130)	0,234*** (0,089)	0,234*** (0,089)	0,463*** (0,129)	0,483*** (0,093)	0,286** (0,130)	0,234*** (0,089)

(Continua)

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	5º ano	9º ano	5º ano	9º ano		
Professor: homem	-1,439*** (0,221)	-0,705*** (0,165)	-2,094*** (0,215)	-0,688*** (0,162)	-4,462** (1,823)	2,384* (1,217)	-7,792*** (1,843)	-0,787 (1,160)	3,403*** (0,192)	2,683*** (0,174)	2,198*** (0,186)	1,989*** (0,170)	3,592** (1,539)	1,635 (1,119)
Professor: branco	1,130*** (0,171)	0,957*** (0,154)	0,802*** (0,166)	0,748*** (0,151)	1,434 (1,393)	6,083*** (1,121)	2,470* (1,408)	4,348*** (1,068)	1,443*** (0,223)	1,267*** (0,186)	1,291*** (0,217)	0,997*** (0,183)	-1,807 (2,027)	-4,219** (2,050)
Professor com mais de dez anos de experiência	2,079*** (0,160)	-2,269*** (0,152)	-1,614*** (0,156)	-1,509*** (0,149)	-0,967 (1,369)	3,823*** (1,085)	0,713 (1,383)	1,914* (1,034)	0,544*** (0,159)	-0,785*** (0,210)	0,334*** (0,155)	2,088 (1,373)	0,406 (1,388)	2,040 (1,304)
Professor com curso superior	-1,213*** (0,214)	-0,004 (0,161)	-1,192*** (0,208)	-0,587*** (0,158)	2,911 (1,882)	4,289*** (1,330)	2,513 (1,903)	3,508*** (1,267)	1,423*** (0,179)	0,941*** (0,164)	0,948*** (0,174)	0,717*** (0,160)	-0,225 (1,452)	0,113 (1,149)
Diretor: homem	1,911*** (0,192)	-0,934*** (0,169)	1,724*** (0,186)	-0,199 (0,166)	5,871*** (1,366)	1,120 (1,044)	3,162** (1,381)	-0,911 (0,995)	2,640*** (0,559)	-5,486*** (0,595)	-2,496*** (0,543)	-5,467*** (0,584)	-13,237** (6,390)	-41,565*** (7,554)
Diretor: branco	1,243*** (0,105)	2,159*** (0,100)	1,371*** (0,102)	2,090*** (0,098)	2,431*** (0,936)	2,696*** (0,755)	3,104*** (0,946)	1,960*** (0,719)	-1,851*** (0,107)	-2,267*** (0,094)	-1,020*** (0,104)	-1,123*** (0,093)	-4,594*** (1,491)	1,133 (1,285)
Diretor tem mais de dez anos de experiência	-1,109*** (0,109)	-1,033*** (0,091)	-0,868*** (0,106)	-0,780*** (0,089)	-0,377 (1,628)	-7,076*** (1,189)	1,039 (1,646)	-4,768*** (1,133)						
Diretor tem curso superior														
Qualidade da escola														
Violência externa														
Violência interna														

(Continua)

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	5º ano	9º ano	5º ano	9º ano
Prova de seleção	-2,626*** (0,816)	12,760*** (0,664)	-1,712** (0,792)	10,332*** (0,651)	9,106*** (2,047)	5,092*** (1,424)	8,770*** (2,069)	4,618*** (1,357)				
Turmas homogêneas quanto à idade	-0,787*** (0,217)	-0,735*** (0,198)	-0,782*** (0,211)	-0,227 (0,194)	6,187*** (1,682)	-2,043 (1,294)	0,820 (1,701)	-1,813 (1,233)				
Turmas homogêneas quanto ao rendimento escolar	0,606* (0,357)	4,322*** (0,388)	0,796** (0,347)	3,945*** (0,380)	-2,398 (2,932)	-2,579 (2,614)	-3,978 (2,964)	-0,780 (2,491)				
Turmas heterogêneas quanto à idade	-2,236*** (0,313)	-0,151 (0,278)	-1,641*** (0,304)	0,319 (0,273)	-0,569 (3,717)	-2,625 (2,219)	-4,247 (3,758)	-2,939 (2,114)				
Turmas heterogêneas quanto ao rendimento escolar	1,387*** (0,253)	3,169*** (0,239)	1,886*** (0,246)	2,978*** (0,235)	10,241*** (2,277)	5,282*** (1,745)	6,527*** (2,302)	4,347*** (1,662)				
Diferença de sexo entre aluno e professor	-0,034 (0,212)	-0,739*** (0,161)	-0,269 (0,206)	-0,460*** (0,157)	0,806 (1,710)	-1,940* (1,085)	1,135 (1,729)	-2,103** (1,034)				
Diferença de raça entre aluno e professor	-2,942*** (0,174)	-2,198*** (0,157)	-2,486*** (0,169)	-1,774*** (0,154)	-3,888*** (1,272)	0,338 (1,010)	-0,559 (1,286)	-0,046 (0,962)				
Região urbana	4,049*** (0,291)	2,610*** (0,281)	4,018*** (0,283)	4,174*** (0,276)	8,834** (4,086)	15,496*** (3,314)	4,816 (4,130)	15,166*** (3,157)				
Nordeste	-16,263*** (0,238)	-8,685*** (0,222)	-11,248*** (0,232)	-8,299*** (0,218)	-18,976*** (2,088)	-5,780*** (1,729)	-12,320*** (2,110)	-3,971** (1,647)				
Norte	-12,758*** (0,267)	-4,621*** (0,257)	-7,074*** (0,260)	-2,082*** (0,252)	-16,826*** (2,004)	-12,408*** (1,661)	-8,919*** (2,026)	-6,301*** (1,582)				
Sul	-0,206 (0,312)	1,170*** (0,268)	-0,371 (0,303)	-1,399*** (0,263)	-6,920** (2,998)	-4,411* (2,418)	-12,080*** (3,031)	-4,790** (2,303)				
Centro-Oeste	-2,741*** (0,419)	1,802*** (0,373)	-0,395 (0,407)	1,849*** (0,366)	-12,184*** (2,703)	-9,498*** (2,145)	-3,745 (2,733)	-3,481* (2,043)				

Elaboração dos autores.

Obs.: 1. ***, ** e * = $p < 0,01$, $p < 0,05$ e $p < 0,1$.

2. Erro-padrão entre parênteses.