

# A Relação da Escolaridade dos Pais no Desempenho Acadêmico dos Filhos

Michel Csillag Finger  
Theodoro Nunes Figueiredo Mota

21 de Junho, 2024

## Resumo

Este estudo tem como objetivo avaliar o efeito do nível de escolaridade da mãe e do pai sobre o desempenho acadêmico de alunos brasileiro do 5º ano do ensino fundamental, utilizando microdados do Sistema de Avaliação da Educação Básica (SAEB) de 2019. Foi constatado um efeito positivo significativo.

**Palavras-chaves:** Educação; Escolaridade dos Pais; Desempenho Escolar.

## 1 Introdução

No combate à intensa desigualdade socioeconômica brasileira, a elaboração adequada de políticas públicas voltadas à educação é essencial. Para tanto, é necessário compreender os fatores que influenciam o desempenho escolar dos estudantes brasileiros. Sobre esse tema, buscamos encontrar os efeitos do ambiente domiciliar sobre a educação de crianças, em especial sobre o que diz respeito à condição familiar de crianças brasileiras. Nesse contexto, identificamos estudos prévios que abordavam, em especial, efeitos dos da educação dos pais sobre o desempenho educacional de seus filhos.

No contexto brasileiro, [Feijó, Fraça e Pinho \(2022\)](#) utilizaram microdados do Exame Nacional do Ensino Médio (ENEM) de 2017 e analisaram o efeito conjunto do nível educacional da mãe e do pai de candidatos sobre a nota do vestibular. Sobre os resultados, destaca-se as notas, em média, de alunos com pais com ensino superior são mais do que 50% maiores, mas quando são incluídas variáveis de controle, esse efeito diminuiu drasticamente. A consideração relevante dessa pesquisa é sobre a existência de fatores não observáveis que influenciam o desempenho dos alunos.

Já em um outro estudo, [Bassetto \(2019\)](#) utilizou dados do Sistema de Avaliação do Rendimento Escolar do Estado de São Paulo (Saresp) e identifica que os efeitos de renda familiar e escolaridade da mãe não devem ser interpretados como determinantes

no desempenho de alunos e que outras variáveis devem ser consideradas para encontrar fatores determinantes.

No cenário internacional, [Ermisch e Pronzato \(2010\)](#) utilizaram dados da Noruega para avaliar efeitos da educação dos pais sobre o nível de escolaridade dos filhos, medido em anos de estudo. Em linhas gerais, foram encontradas evidências de que o efeito da educação da mãe é maior quando o nível da educação dos pais é elevado, enquanto o efeito da educação do pai se destaca quando o nível de educação dos pais é menor. Um ponto relevante é que esta pesquisa aborda a questão da mobilidade intergeracional, pois fatores de ambiente domiciliar, não apenas de educação dos pais, como a renda e a ocupação dos mesmos influenciam o comportamento de seus filhos, assim como suas aspirações.

Também destacamos que [Tan e Goldberg \(2009\)](#) fizeram um estudo sobre os efeitos do envolvimento dos pais nos estudos dos filhos em um estado no Oeste dos Estados Unidos. Eles constataram, entre outras coisas, que um envolvimento direto dos pais aumenta o desempenho acadêmico de seus filhos. Além disso, encontraram efeitos distintos para envolvimento entre pais e mães, ou seja, a variação de nota depende se é o pai ou a mãe ajudando o filho na escola. Ademais, eles também constatam que, na maioria das vezes os pais em específico, não mães, se envolvem quando creem que os filhos necessitam de ajuda na escola, o que eleva o nível de ansiedade dos filhos.

Por fim e retornando ao contexto brasileiro, [Reis e Ramos \(2011\)](#) encontraram evidências de que tanto o nível de escolaridade dos pais quanto a renda dos pais influencia não apenas o desempenho acadêmico do aluno, mas também o quão bom será o retorno de seus estudos, o que expõe uma dinâmica de perpetuidade da desigualdade social no Brasil quando comparamos gerações seguintes de famílias com menor nível de escolaridade e renda e daquelas com maior nível de escolaridade e renda.

Sobre a tema, questionamos não apenas o efeito do nível de escolaridade dos pais sobre o desempenho e sucesso acadêmico de crianças brasileiras a partir do uso de dados do Sistema de Avaliação do Sistema Básico de 2019 de alunos do 5o ano do ensino fundamental.

## 2 Método

### 2.1 Hipóteses dos Modelos

Esta seção abrange a teoria econométrica por trás do modelo utilizado na pesquisa. Dadas as limitações impostas pela base de dados (explicitadas na seção 3), o grupo optou por realizar uma regressão múltipla, e a seguir estão apresentadas as quatro principais hipóteses (hipóteses "não estendidas"), de acordo com [Stock e Watson \(2020\)](#).

Dado um modelo matricial de regressão linear,  $\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{U}$ , em que  $\mathbf{Y}$  é o vetor das observações da variável dependente,  $\mathbf{X}$  é a matriz das observações sobre os regressores,  $\boldsymbol{\beta}$  é o vetor dos coeficientes regressores desconhecidos e  $\mathbf{U}$  é o vetor coluna dos termos de erro. As hipóteses são feitas sobre as observações individuais, que podem ser escritas da seguinte maneira em forma matricial:  $Y_i = \mathbf{X}_i^T \boldsymbol{\beta} + u_i$ . Este é o mesmo modelo, mas especificado para somente a  $i$ -ésima observação, além de  $\mathbf{X}_i^T$  ser a matriz coluna das observações transposta, para que após a multiplicação pelo vetor  $\boldsymbol{\beta}$ , resulte num escalar, tal qual todos os outros elementos da equação.

Listando as hipóteses, temos a primeira hipótese (H1) é a de Exogeneidade do Termo de Erro, ou seja,  $E[u|\mathbf{X}_i] = 0$ , o que no modelo implica o termo de erro não ter relação com as variáveis independentes. A segunda hipótese (H2) é de que  $(\mathbf{X}_i, Y_i)$  são

independentes e identicamente distribuídos (i.i.d), para toda observação. A terceira hipótese (H3) é de que  $\mathbf{X}_i$  e  $u_i$  têm quartos momentos finitos, ou seja,  $E[\mathbf{X}_i^4] < \infty, E[u_i^4] < \infty$ ; outra maneira de descrever essa hipótese é que grandes *ouliers* são improváveis na amostra. A quarta hipótese (H4) é de que  $\mathbf{X}_i$  tem posto cheio, ou seja, não há multicolinearidade perfeita.

Avaliando a validade e interpretando as hipóteses na base de dados e no problema escolhidos, para (H1) é esperado que não haja exogeneidade porque foram incluídas diversas variáveis de controle na tentativa de isolar o efeito do viés de variável omitida, embora não fosse possível controlar por fatores como QI do aluno, nem renda dos pais, que são fatores que podem correlacionar o termo de erro com as variáveis independentes. Sobre (H2), é plausível que não haja independência entre as amostras, por haverem grupos de alunos com professores melhores e piores que transmitam conteúdo de maneira melhor ou pior, afetando o desempenho na prova de maneira a depender do professor. Sobre (H3), é improvável, porém possível que haja grandes *ouliers* pela maneira que o SAEB atribui nota para as provas: a distribuição de notas é padronizada, com média de 250 e desvio padrão de 50. Finalmente, sobre (H4), é perfeitamente plausível que não haja multicolinearidade perfeita já que os regressores são escolhidos de modo a todos serem diferentes, sem sobreposição, com cuidado especial para variáveis binárias categóricas.

## 2.2 Modelos Utilizados

Tendo isso em vista, o grupo decidiu testar a relação principalmente entre a escolaridade dos pais e o desempenho acadêmico de seus filhos. Para isso, o grupo elencou alguns modelos para testar.

Na base de dados, o grupo decidiu fazer modelos que testassem o desempenho acadêmico em Língua Portuguesa (LP) e em Matemática (MT), e a escolaridade dos pais. É notável que o maior impacto no desempenho acadêmico dos filhos se dá na presença de pais com ensino superior cursado. Logo, ao invés de medir o efeito de anos de escolaridade dos pais, a regressão é feita com base em uma variável binária de ensino superior: 1 caso seja presente, e 0 caso contrário. Assim, temos uma estrutura geral para os modelos estudados:

$$\begin{aligned} NOTA_i = & \beta_0 + \beta_1 ES_i^X + \beta_2 EP_i + \beta_3 Urb_i + \beta_4 Cap_i + \\ & \beta_5 LM_i + \beta_6 CE_i + \beta_7 IE_i + \beta_8 ILC_i + \\ & \beta_9 IP_i + \sum_{j=1}^5 \beta_{j+9} Raça_i + \sum_{k=1}^{26} \beta_{k+14} Estado_i + u_i \end{aligned} \quad (1)$$

No modelo, constam as principais variáveis de interesse:  $NOTA_i$  é a nota padronizada do aluno da prova do SAEB, que pode ser tanto a nota da prova de matemática ( $MT_i$ ) quanto a de português ( $PT_i$ ) e  $ES_i^X$  variável de presença de ensino superior de um (ou ambos) pais indicada pelo sobrescrito, podendo ser da mãe ( $ES_i^M$ ) ou do pai ( $ES_i^P$ ). Além disso, o modelo apresenta variáveis binárias para: indicar se o aluno estuda em escola pública ( $EP_i$ ), uma para indicar se o aluno mora em região urbana ( $Urb_i$ ), uma para indicar se o aluno mora numa das capitais ( $Cap_i$ ) e uma para indicar se a língua materna do aluno é português ( $LM_i$ ).

Ademais, com a intenção de isolar o envolvimento direto, ou alto, dos pais na escolaridade do filho, o que comprovadamente tem impacto positivo no desempenho dos

filhos (TAN; GOLDBERG, 2009), constam variáveis binárias de caso os pais têm alta frequência em: conversar com o filho sobre a escola ( $CE_i$ ), incentivar o filho a estudar ( $IE_i$ ), incentivar o filho a fazer lição de casa ( $ILC_i$ ), e incentivar a presença nas aulas ( $IP_i$ ).

No modelo, também constam variáveis categóricas para controlar efeitos do Estado em que o estudante mora e pela Raça do estudante, sem a paridade de uma variável para uma categoria para evitar problemas de multicolinearidade perfeita. Por conta disso, foram retiradas a variável de São Paulo para Estado — por ser o Estado com a maior média nas notas das provas — e Branco para Raça — por conta do racismo estrutural da sociedade brasileira contemporânea —, com intuito de comparar os efeitos das outras categorias com elas em suas respectivas análises.

### 3 Dados

Com a finalidade de analisar os efeitos da condição de domicílio, em especial o nível de educação dos pais, sobre o desempenho escolar, utilizamos microdados do Sistema Nacional de Avaliação Básica (SAEB) de 2019<sup>1</sup>. A base de dados contém a nota de dois exames de proficiência, sendo um em Matemática e outro em Língua Portuguesa, e também respostas de um questionário extenso sobre suas condições de domicílio com respostas categóricas.

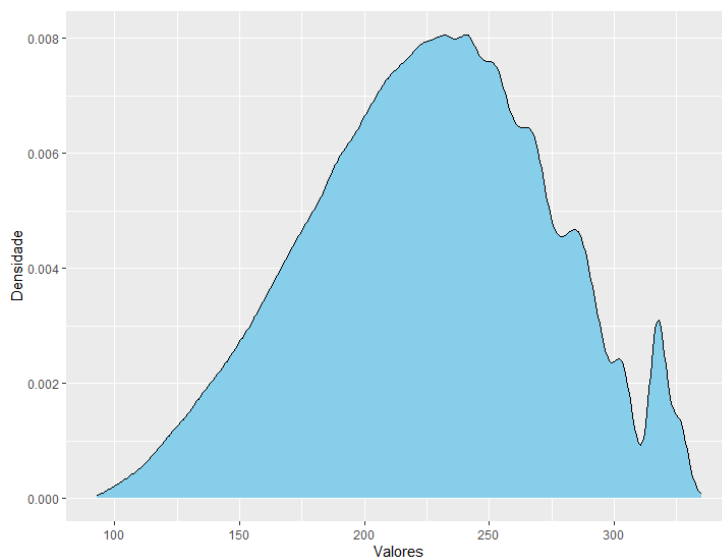
Embora exista dados mais atualizados de 2021, a escolha pelo microdados de 2019 se deve pela consideração pelos efeitos da pandemia e a possibilidade de distorção dos resultados. Sob a premissa de que a pandemia afetou diferentes camadas e tipos de indivíduos da sociedade brasileira de formas diferentes, em média alunos com maior renda em escolas particulares tiveram aulas assíncronas, cobrança de atividades, e mais, em comparação com escolas públicas que não tinham a infraestrutura necessária e interromperam as aulas por semanas, ou até meses.

A base original apresenta mais de 2 milhões de indivíduos de todas as unidades federativas. No entanto, a base com cortes do estudo desconsidera aquelas observações com respostas inválidas resultando em uma amostra com 828.395 observações. A tabela 1 apresenta as estatísticas descritivas das variáveis de interesse de nosso estudo. Por conta dos cortes realizados, seja por não responderem o formulário corretamente, seja por respostas em branco, seja por respostas nulas, a padronização de nota da prova do SAEB foi alterada. A prova do SAEB conta com uma escala única, com média 250 e desvio padrão 50 na amostra completa, em ambas provas de Língua Portuguesa (LP) e de Matemática (MT). Após os cortes, a média da prova de LP é de 225.231, com desvio padrão de 46,517, enquanto a prova de MT tem média 239.025 e desvio padrão de 46.099, como consta nas linhas *Nota LP* e *Nota MT*, respectivamente, da tabela 1.

Como variável dependente escolhemos as notas dos exames padronizados de proficiência em Língua Portuguesa e Matemática. Devido ao formato da pesquisa do SAEB, essas são as nossas únicas variáveis contínuas e ambas variam entre 0 e 375. As notas dos exames de proficiência em Língua Portuguesa e Matemática são padronizadas para uma média de 250 pontos e um desvio padrão de 50. As figuras 1 e 2 apresentam as distribuições de notas da amostra de nossa base com cortes.

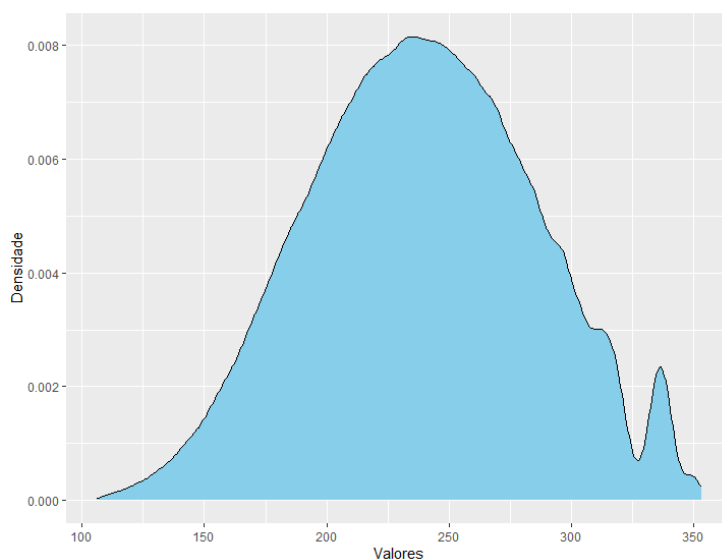
<sup>1</sup> A base pode ser acessada pelo seguinte link, selecionando o ano “2019”, no item “Microdados do Saeb 2019”: <https://www.gov.br/inep/pt-br/areas-de-atuacao/avaliacao-e-exames-educacionais/saeb/resultados>

Figura 1 – Distribuição das Notas de Língua Portuguesa (LP)



Fonte: Elaboração própria.

Figura 2 – Distribuição das Notas de Matemática (MT)



Fonte: Elaboração própria.

Quando filtramos por Unidade Federativa, as estatísticas dos exames padronizados revelam que a distribuição de notas e, portanto do nível de escolaridade, é extremamente desigual dentro do território nacional. Enquanto estados como São Paulo destacam por médias elevadas, respectivamente 236.548 e 253.438 para as provas de proficiência de Língua Portuguesa e Matemática, ainda há aqueles como o Maranhão e o Pará que têm médias respectivamente 199.931 e 211.884 para as provas de proficiência de Língua Portuguesa e Matemática. A tabela 2 apresenta essas estatísticas, junto com as médias dos demais estados.

A base de dados escolhidas é caracterizada pela falta de variáveis contínuas que não sejam as notas dos exames padronizados de Matemática e Língua Portuguesa. Como

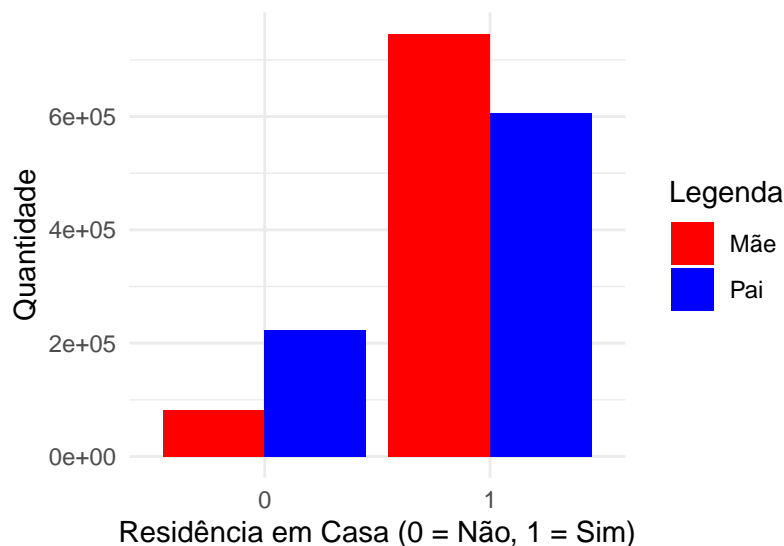
consequência disso, as variáveis independentes de interesse escolhidas são todas categóricas e binárias. Nesse contexto, destacamos algumas observações relevantes sobre as variáveis independentes e a amostra.

A primeira consideração diz respeito à informação presente no gráfico da figura 3. Nota-se que, em geral, a ocorrência de pais ausentes é maior do que a de mães ausentes, ao passo que a presença de mães na casa é superior à presença de pais. Dito isso, para não haver disparidades, as regressões foram feitas filtrando a base para que houvesse presença do pai analisado. Embora haja redução no tamanho da amostra, pelo número de observações permanecer acima de 600.000 para pais e 700.000 para mães após os cortes, consideramos que ainda estamos no assintótico, sem perdas para a inferência.

Além disso, o gráfico da figura 4 expõe que a ocorrência de pais com ensino superior é muito menor do que a de pais sem ensino superior. Destaca-se também que há mais mães com ensino superior do que pais com ensino e o número de pais sem ensino superior é maior do que o de mães sem ensino superior. E, comparando entre os pais, há mais mães com ensino superior do que pais com ensino superior.

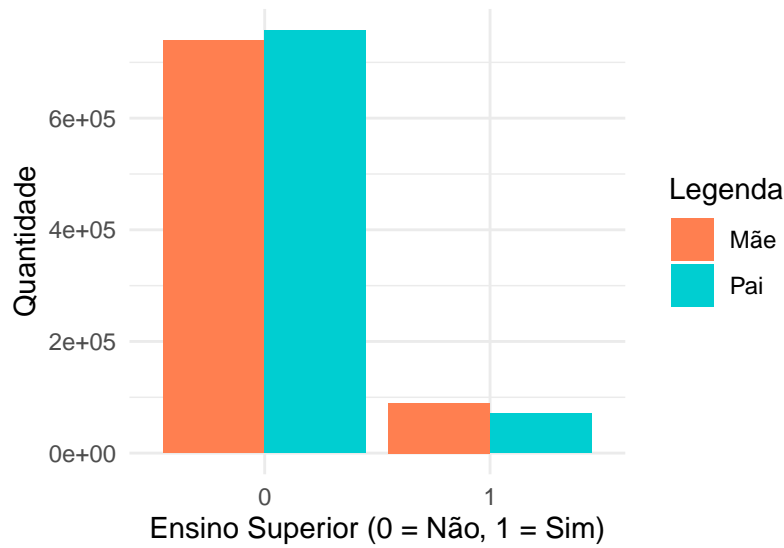
Uma outra observação notável diz respeito às correlações entre as variáveis independentes pode ser observado na matriz de correlação das variáveis independentes da figura 5. As correlações mais intensas são entre as variáveis das notas das provas de PT e MT ( $Corr = 0.67$ ) e entre a escolaridade da mãe e do pai ( $Corr = 0.54$ ). Fora isso, as correlações são menores, em módulo, do que 0.32. Tamanha baixa correlação é um forte indício da ausência de multicolinearidade perfeita, corroborando com a quarta hipótese do modelo (H4). Os valores das correlações podem ser encontrados na figura 5, a matriz de correlação dos regressores excluindo os controles por raça e Estado.

Figura 3 – Comparação de Frequência de Residência dos Pais



Fonte: Elaboração própria.

Figura 4 – Comparação de Presença de Ensino Superior dos Pais



Fonte: Elaboração própria.

Figura 5 – Matriz de Correlação das Variáveis Independentes

Capital	1.00	-0.08	0.14	0.04	0.02	-0.02	-0.00	-0.04	0.01	0.02	0.01	-0.01	-0.00	-0.00
Pública	-0.08	1.00	-0.03	-0.06	-0.05	0.01	-0.02	-0.02	-0.06	-0.07	-0.01	-0.00	-0.00	0.01
Urbano	0.14	-0.03	1.00	0.14	0.13	-0.00	0.03	-0.03	0.05	0.05	0.02	0.00	0.02	0.02
Nota LP	0.04	-0.06	0.14	1.00	0.67	0.04	0.10	0.07	0.15	0.11	0.03	0.06	0.07	0.09
Nota MT	0.02	-0.05	0.13	0.67	1.00	0.03	0.10	0.08	0.15	0.12	0.02	0.03	0.04	0.07
LM_PT	-0.02	0.01	-0.00	0.04	0.03	1.00	0.01	0.01	-0.02	-0.03	0.00	0.03	0.03	0.04
Pres_Mae	-0.00	-0.02	0.03	0.10	0.10	0.01	1.00	0.28	0.05	0.02	0.04	0.03	0.04	0.03
Pres_Pai	-0.04	-0.02	-0.03	0.07	0.08	0.01	0.28	1.00	0.02	0.09	0.04	0.03	0.04	0.02
Esc_Mae	0.01	-0.06	0.05	0.15	0.15	-0.02	0.05	0.02	1.00	0.54	0.11	0.06	0.05	0.03
Esc_Pai	0.02	-0.07	0.05	0.11	0.12	-0.03	0.02	0.09	0.54	1.00	0.10	0.05	0.04	0.02
Conv_Esc.	0.01	-0.01	0.02	0.03	0.02	0.00	0.04	0.04	0.11	0.10	1.00	0.15	0.19	0.11
IEst. Alto	-0.01	-0.00	0.00	0.06	0.03	0.03	0.03	0.03	0.06	0.05	0.15	1.00	0.32	0.32
ILC Alto	-0.00	-0.00	0.02	0.07	0.04	0.03	0.04	0.04	0.05	0.04	0.19	0.32	1.00	0.28
IP Alto	-0.00	0.01	0.02	0.09	0.07	0.04	0.03	0.02	0.03	0.02	0.11	0.32	0.28	1.00

Fonte: Elaboração própria.

## 4 Resultados

Os resultados das regressões estão na tabela 3, contida no apêndice 5<sup>2</sup>. Dentre eles, ressaltamos que os únicos regressores não significativos ao nível de 1%, são significativos ao nível de 5%. A seguir, apresentamos os principais resultados das regressões, relacionados à pergunta original envolvendo a escolaridade dos pais (denotados por ES Mãe e ES Pai na tabela) sobre o envolvimento dos pais na educação dos filhos, extremamente relevante na literatura (ERMISCH; PRONZATO, 2010; TAN; GOLDBERG, 2009), e sobre as medidas

<sup>2</sup> O código pode ser encontrado no seguinte repositório do GitHub: [https://github.com/Micsifin/Econometrics\\_I](https://github.com/Micsifin/Econometrics_I)

de ajuste e avaliação da regressão.

A respeito da relação entre a presença de ensino superior dos pais e as notas nas provas do SAEB dos alunos, o fato de mães terem ensino superior aumenta a nota da prova de matemática, em média, em 15.365 pontos, *ceteris paribus*; enquanto, na prova de língua portuguesa, há um aumento de 15.109 pontos, em média, tudo mais constante. Por outro lado, a presença de ensino superior dos pais está relacionada a um aumento na nota da prova de matemática de 12.46 pontos, em média, *ceteris paribus*; ao passo que a presença de ensino superior paterno aumenta a nota de LP, em média, 12.149 pontos, tudo mais constante.

Sendo assim, é notável o impacto positivo da presença de ensino superior dos pais no desempenho acadêmico dos filhos. Por mais que números como 12 e 15 pontos na escala única do SAEB que padroniza as notas com média de 250 e desvio padrão de 50, essa pontuação é aproximadamente metade de um nível na escala de proficiência do SAEB: cada nível é um intervalo de 25 pontos na escala do SAEB (BRASIL, 2020). Ou seja, supondo que a análise ocorre dentro de um nível em específico, 12 ou 15 pontos adicionais pode ser a diferença entre perceber e identificar com clareza, por exemplo, a localização de informação explícita em contos, reportagens e propagandas com ou sem apoio de recursos gráficos, entre outros, por exemplo, como consta no Nível 3 da Escala de Proficiência de Língua Portuguesa para 5º ano do Ensino Fundamental (BRASIL, 2020).

A respeito do envolvimento dos pais na educação dos filhos, por mais que ele seja reconhecido como um fator importante na educação dos filhos, neste presente estudo não foi constatado um efeito muito grande na nota dos alunos. Mesmo que todos os coeficientes sejam significantes a nível de confiança 1%, a magnitude dos efeitos constatada não é tão grande quanto o efeito da presença de ensino superior dos pais.

Embora pareça contraintuitivo, uma alta quantidade de conversas sobre a escola entre pais e filhos só diminui a nota: em média, as notas das provas variaram  $-0.986$ ,  $-0.663$ ,  $-0.636$ ,  $-0.328$  nas quatro regressões (1), (2), (3) e (4), tudo mais constante, como consta na tabela 3. Além disso, sobre o alto incentivo ao estudo dos filhos, foi constatado um maior aumento na média das notas das provas de LP do que de MT: em média, as notas de MT aumentam 0.425 (regressão 1) e 0.558 (regressão 2), tudo mais constante; enquanto as notas de LP aumentam, em média, 2.699 (regressão 3) e 2.865 (regressão 4), *ceteris paribus*.

O alto incentivo a fazer a lição de casa também conta valores similares aos do alto incentivo ao estudo: a nota de MT aumenta, em média, 0.89 (regressão 1) e 1.021 (regressão 2), enquanto a nota de LP aumenta 2.389 (regressão 3) e 2.523 (regressão 4) na presença do incentivo a realização de lição de casa (Incentivo LC na tabela 3. Por outro lado, o alto incentivo a presença conta com números mais expressivos: a nota de MT aumenta, em média, 7.685 (regressão 1) e 7.421 (regressão 2), enquanto a nota de LP aumenta 9.245 (regressão 3) e 8.943 (regressão 4) na presença do incentivo à presença.

Sobre as medidas de ajuste, como estamos tratando de uma regressão múltipla, usaremos o  $R^2$  ajustado na análise, embora seja possível constatar que tanto o  $R^2$  quanto o  $R^2$  ajustado apresentem os mesmos valores na tabela 3. O modelo das notas de MT na presença de ensino superior da mãe é capaz de explicar 12.6% da variabilidade dos erros, enquanto 12.4% da variabilidade dos erros são explicados pelo modelo com ensino superior do pai e notas da prova de MT. Além disso, 11.3% e 11.2% da variabilidade dos erros são explicados pelos modelos de LP com presença de ensino superior da mãe e do pai, respectivamente. Resultados inexpressivos eram relativamente esperados levando em conta



que os modelos se tratam somente de variáveis categóricas e binárias, sem espaço para maiores variabilidades de variáveis contínuas. Além disso, a estatística F nos 4 modelos é próxima de 2 milhões, o que é extremamente relevante a qualquer nível de significância. Esses valores indicam que há ao menos algum coeficiente que é não nulo nas quatro regressões.

## 5 Conclusão

Este trabalho teve como objetivo avaliar os efeitos do ambiente domiciliar sobre o desempenho escolar, em especial ao que diz respeito à situação familiar de crianças brasileiras. Sobre esse tema, trabalhos anteriores já relevaram questões sobre o nível educacional dos pais sobre o desempenho de alunos brasileiro.

Isolamos os efeitos do envolvimento direto dos pais na educação dos filhos, dado que [Tan e Goldberg \(2009\)](#) já haviam encontrado evidências de que a participação direta dos pais nos estudos dos filhos é um fator relevante sobre o desempenho acadêmico dos alunos e tem um efeito positivo sobre o mesmo.

A respeito dos resultados encontrados, a presença de ensino superior nos pais acarreta num efeito significativo na média das notas das provas dos filhos, tudo mais constante. Além disso, encontramos resultados congruentes com os de [Tan e Goldberg \(2009\)](#). Entretanto, pais conversarem sobre a escola com seus filhos tem um efeito negativo nas notas. Uma explicação possível é que isso leva à ansiedade dos filhos, o que pioraria o desempenho acadêmico (como já constatado na mesma pesquisa), ou até que a conversa sobre a escola somente viria em momentos de desempenho acadêmico preocupante.

Uma consideração relevante sobre o estudo é que diz respeito a uma limitação da própria base de dados escolhidas é o fato de não termos dados sobre a condição financeira dos pais dos alunos que realizaram os exames padronizados. Esse fator foi abordado no trabalho de [Reis e Ramos \(2011\)](#) e é fundamental para a análise da desigualdade socioeconômica brasileira e suas perpetuações.

Além disso, pelo fato de haverem máscaras nos identificadores dos municípios, não foi possível utilizar de um ferramental econométrico mais aprofundado, como variáveis instrumentais, ou a junção de diversas bases de dados para mais análises estatísticas e outras inferências. Possíveis próximas etapas para pesquisas do gênero seriam realizar trabalhos similares com controles para renda per capita municipal, por exemplo, com bases que não mascaram seus dados.

Seguindo [Stock e Watson \(2020\)](#) sobre ameaças a validade interna, temos que o nossos estimadores de mínimos quadrados ordinários são consistentes e não viesados, além de rejeitarmos a hipótese nula do testes de hipótese para todos os casos com, ao menos, 5% de nível de significância. Assim, nossas inferências estatísticas devem ser validas para a amostra.

Além disso, de acordo com a própria nota técnica do SAEB, a amostra de alunos deve ser representativa da população brasileira e foi definida utilizando tanto dados do censo escolar de 2018, quanto dados preliminares do censo escolar de 2019. Sob a premissa de que isto é válido e que não há diferenças significativas entre a amostra e a população de interesse, não haveria problemas relacionados à validade externa. Porém, como a base sofreu muitos cortes, a validade externa do estudo pode ter sido comprometida.

## Referências

- BASSETTO, Camila Fernanda. Background familiar e desempenho escolar: uma abordagem com variáveis binárias a partir dos resultados do Saresp. **Revista Brasileira de Estudos de População**, SciELO Brasil, v. 36, e0077, 2019.
- BRASIL. Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira (INEP). **Escala de proficiência do SAEB**, Brasília, DF: INEP, 2020.
- ERMISCH, John; PRONZATO, Chiara. **Causal effects of parents' education on children's education**. [S.l.], 2010.
- FELJÓ, Janaína Rodrigues; FRAÇA, João Mário Santos De; PINHO, Valdemar Rodrigues De. Desempenho dos estudantes ao final do ensino médio: Mensurando a influência direta e indireta da educação dos pais. **Revista Brasileira de Economia**, SciELO Brasil, v. 76, n. 1, p. 30–56, 2022.
- REIS, Mauricio Cortez; RAMOS, Lauro. Escolaridade dos pais, desempenho no mercado de trabalho e desigualdade de rendimentos. **Revista Brasileira de Economia**, SciELO Brasil, v. 65, p. 177–205, 2011.
- STOCK, James H; WATSON, Mark W. **Introduction to econometrics**. [S.l.]: Pearson, 2020.
- TAN, Edwin T; GOLDBERG, Wendy A. Parental school involvement in relation to children's grades and adaptation to school. **Journal of applied developmental psychology**, Elsevier, v. 30, n. 4, p. 442–453, 2009.

## A Tabelas

Tabela 1 – Descritivas das Variáveis Utilizadas no Modelo

Variáveis	N	Média	Desv. Pad.	Mín.	Máx.
Ensino Superior Mãe	828,395	0.109	0.311	0	1
Ensino Superior Mãe	828,395	0.085	0.279	0	1
Conversa Escola	828,395	0.417	0.493	0	1
Alto Incentivo ao Estudo	828,395	0.852	0.355	0	1
Alto Incentivo LC	828,395	0.791	0.406	0	1
Alto Incentivo Presença	828,395	0.906	0.291	0	1
Estadi	828,395	32.051	9.920	11	53
Capital	828,395	0.186	0.389	0	1
Escola Pública	828,395	0.986	0.119	0	1
Urbano	828,395	0.888	0.315	0	1
Preencheu Prova LP	828,395	1.000	0.017	0	1
Preencheu Prova MT	828,395	1.000	0.018	0	1
Nota LP	828,395	225.231	47.517	92.790	334.808
Nota MT	828,395	239.025	46.099	105.998	353.090
PT Língua Materna	828,395	0.983	0.130	0	1
Raça	828,395	2.072	1.223	0	5
Presença Mãe	828,395	0.901	0.298	0	1
Presença Pai	828,395	0.731	0.443	0	1
Escolaridade Mãe	828,395	1.524	1.918	0	5
Escolaridade Pai	828,395	1.209	1.794	0	5
Conversa Escola	828,395	1.322	0.639	0	2
Incentivo ao Estudo	828,395	1.822	0.455	0	2
Incentivo LC	828,395	1.740	0.542	0	2
Incentivo Presença	828,395	1.877	0.409	0	2

Fonte: Elaboração própria.

Tabela 2 – Média de Notas nas Provas por UF

UF	Média Notas LP	Média Notas MT
Rondônia	214.743889536695	229.621228167489
Acre	226.01095481435	242.671342022204
Amazonas	217.358796320676	229.305359310162
Roraima	218.292718326656	231.499469001635
Pará	200.888452915824	211.7781200614
Amapá	206.011119556089	214.654757145239
Tocantins	215.241275180889	230.730215749333
Maranhão	199.931755891098	211.884528785632
Piauí	220.069565697556	233.967266955945
Ceará	234.465500207968	247.068918755729
Rio Grande do Norte	206.019271294802	216.014629155001
Paraíba	210.954468266244	223.982169455553
Pernambuco	209.836731607634	224.958587552691
Alagoas	213.764495759705	231.041272484754
Sergipe	201.134252202996	214.008323825195
Bahia	208.756058445058	220.581527502804
Minas Gerais	233.842786310549	246.972731402131
Espírito Santo	228.015440684563	241.259902659688
Rio de Janeiro	222.82599813477	234.438764779706
São Paulo	236.54896182034	253.438960383442
Paraná	234.62950755947	251.944301277111
Santa Catarina	235.541976275842	247.777718761202
Rio Grande do Sul	229.888752013812	240.00048106486
Mato Grosso do Sul	223.067773573966	234.496393827477
Mato Grosso	221.342609599076	233.02547972642
Goiás	228.540772605611	239.835149551743
Distrito Federal	232.178514251929	244.640765989628

Fonte: Elaboração própria.

Tabela 3 – Resultados das Regressões

	<i>Variáveis Dependentes:</i>			
	MT e Mãe (1)	MT e Pai (2)	LP e Mãe (3)	LP e Pai (4)
ES Mãe	15.365*** (0.160)		15.109*** (0.166)	
ES Pai		12.460*** (0.195)		12.149*** (0.202)
Conversa Escola	-0.986*** (0.104)	-0.663*** (0.115)	-0.636*** (0.107)	-0.328*** (0.119)
Incentivo Estudo	0.425*** (0.155)	0.558*** (0.174)	2.699*** (0.161)	2.865*** (0.180)
Incentivo LC	0.890*** (0.134)	1.021*** (0.151)	2.389*** (0.139)	2.523*** (0.156)
Incentivo Presença	7.685*** (0.187)	7.421*** (0.209)	9.245*** (0.194)	8.943*** (0.217)
Constante	250.204*** (0.627)	251.601*** (0.688)	225.070*** (0.649)	225.855*** (0.712)
Escola Pública	-20.504*** (0.416)	-21.647*** (0.451)	-20.414*** (0.431)	-21.490*** (0.467)
Cidade Urbana	9.348*** (0.168)	9.581*** (0.183)	11.536*** (0.175)	11.630*** (0.189)
Capital	0.382*** (0.140)	0.536*** (0.160)	2.668*** (0.145)	2.695*** (0.165)
PT Língua Materna	10.746*** (0.391)	11.537*** (0.434)	12.098*** (0.405)	13.301*** (0.449)
Não quis declarar	-11.932*** (0.179)	-12.301*** (0.198)	-14.994*** (0.185)	-15.268*** (0.205)
Presença das seguintes variáveis categóricas de Estado nas regressões:				
Acre	Sim	Sim	Sim	Sim
Alagoas	Sim	Sim	Sim	Sim
Amapá	Sim	Sim	Sim	Sim
Amazonas	Sim	Sim	Sim	Sim
Bahia	Sim	Sim	Sim	Sim
Ceará	Sim	Sim	Sim	Sim
Distrito Federal	Sim	Sim	Sim	Sim
Espírito Santo	Sim	Sim	Sim	Sim
Goiás	Sim	Sim	Sim	Sim
Maranhão	Sim	Sim	Sim	Sim
Mato Grosso	Sim	Sim	Sim	Sim
Mato Grosso do Sul	Sim	Sim	Sim	Sim
Minas Gerais	Sim	Sim	Sim	Sim
Pará	Sim	Sim	Sim	Sim
Paraíba	Sim	Sim	Sim	Sim
Paraná	Sim	Sim	Sim	Sim
Pernambuco	Sim	Sim	Sim	Sim

*Continua na próxima página*

Tabela 3 – (Continuação)

	<i>Variáveis Dependentes:</i>			
	MT e Mãe (1)	MT e Pai (2)	LP e Mãe (3)	LP e Pai (4)
Piauí	Sim	Sim	Sim	Sim
Rio de Janeiro	Sim	Sim	Sim	Sim
Rio Grande do Norte	Sim	Sim	Sim	Sim
Rio Grande do Sul	Sim	Sim	Sim	Sim
Rondônia	Sim	Sim	Sim	Sim
Roraima	Sim	Sim	Sim	Sim
Santa Catarina	Sim	Sim	Sim	Sim
São Paulo*	Não	Não	Não	Não
Sergipe	Sim	Sim	Sim	Sim
Tocantins	Sim	Sim	Sim	Sim
<i>* o Estado de São Paulo foi usado como referência nas regressões</i>				
Presença das seguintes variáveis categóricas de Raça nas regressões:				
Branca*	Não	Não	Não	Não
Preta	Sim	Sim	Sim	Sim
Parda	Sim	Sim	Sim	Sim
Amarela	Sim	Sim	Sim	Sim
Indígena	Sim	Sim	Sim	Sim
Não quis declarar	Sim	Sim	Sim	Sim
<i>* a raça Branca foi usada como referência nas regressões</i>				
Observações	746,301	605,701	746,333	605,727
R <sup>2</sup>	0.126	0.124	0.113	0.112
R <sup>2</sup> Ajustado	0.126	0.124	0.113	0.112
F Estatística	2,694.054***	2,150.708***	2,382.551***	1,903.670***
(df = 40, __)	(df = 746260)	(df = 605660)	(df = 746292)	(df = 605686)
<i>Observações:</i>				
*p<0.1; **p<0.05; ***p<0.01				
df é graus de liberdade, em inglês				
Fonte: Elaboração Própria				