Um retrato da educação brasileira: Determinantes multidimensionais do desempenho escolar de estudantes da educação básica no país[12]

Área de interesse de Submissão: Área 7. Economia do Setor Público e Políticas Sociais

Resumo
Estudos sobre economia da educação sugerem que o aprendizado é influenciado por dimensões, abrangendo aspectos individuais, familiares, escolares e socioeconômicos, e contribui para a dinâmica econômica regional. Esta pesquisa investiga os determinantes da qualidade da educação no Brasil e seus efeitos sobre o desempenho de estudantes e o crescimento de municípios, com base nos microdados do Enem e do Censo da Educação do Inep. Em conformidade com a literatura e evidências empíricas, infere-se que desigualdades raciais e socioeconômicas, associadas a contextos familiares, são determinantes do aprendizado, sugerindo externalidade intergeracional entre as condições parentais e os resultados dos filhos. Como estratégia de superação da inacessibilidade dos dados decorrentes da LGPD, adotou-se ênfase em instituições de educação básica localizadas em municípios pequenos, onde há somente uma escola, possibilitando apreender informações dos educandos. Estimou-se um modelo de regressão múltipla com estudantes que realizaram o Enem entre 2015 e 2023, pautado em informações individuais e familiares, e outra estimação com amostra restrita que permitiu cruzamento de informações individuais, familiares e escolares. Posteriormente, estimou-se um modelo de dados em painel hierárquico para compreender o peso do efeito individual e efeito escola. Destacam-se externalidades intergeracionais decorrentes da escolaridade, ocupação e renda dos pais, somando-se ao papel da escola na construção do aprendizado dos estudantes brasileiros, sugerindo a melhoria de indicadores educacionais requer um olhar amplo para políticas públicas no âmbito educacional, de emprego, renda e políticas sociais que amenizem as dificuldades durante e anteriormente à escola.

Palavras-chave: Qualidade da educação; Enem; Modelo hierárquico de dados em painel; Externalidades intergeracionais.

Abstract
Studies on the economics of education suggest that learning is influenced by multiple dimensions, encompassing individual, family, school, and socioeconomic aspects, while also contributing to regional economic dynamics. This research investigates the determinants of educational quality in Brazil and their effects on student performance and municipal growth, based on microdata from the Enem and the Basic Education Census conducted by Inep. In accordance with the literature and empirical evidence, it is inferred that racial and socioeconomic inequalities, associated with family contexts, are determinants of learning, suggesting an intergenerational externality between parental conditions and children's outcomes. The General Data Protection Law restricts access to student information. As a strategy to overcome data inaccessibility, emphasis was placed on basic education institutions located in small municipalities where there is only one school, making it possible to obtain broad information on students. A multiple regression model was estimated using students who took the Enem between 2015 and 2023, based on individual and family information, along with another estimation using a restricted sample that allowed the matching of individual, family, and school information. Subsequently, a hierarchical panel data model was estimated to understand the relative weight of the individual effect and the school effect. Intergenerational externalities stand out arising from the schooling, occupation and income of parents, adding to the role of the school in the construction of the learning of Brazilian students, suggesting the improvement of educational indicators requires a broad look for public policies in the educational scope, of employment, income and social policies that mitigate the difficulties during and prior to school.

[1] João Victor Souza da Silva. Doutor em Economia (UnB). Professor do Departamento de Ciências Econômicas (UFPI).
[2] Andrea Felippe Cabello. Doutora em Economia (UnB). Professora do Departamento de Economia (UnB).

Key-words: Educational quality; Enem (National High School Exam); Multilevel panel modeling; Intergenerational externalities.

1. Introdução

A capacidade de aprendizagem do sujeito é fruto de múltiplos fatores relacionados à sua constituição própria como indivíduo, ao ambiente socioeconômico onde vive e ao contexto familiar onde é criado. Estes aspectos, por sua vez, influenciam-se mutuamente, implicando em um processo árduo a compreensão dos determinantes do seu acúmulo cognitivo. Esta vicissitude é agravada no cenário brasileiro, em decorrência de seu passado colonial e do retrospecto histórico de crescimentos heterogêneos e desigualdades raciais, de gênero, socioeconômicas e espaciais. Não por acaso, a compreensão e a busca por soluções educacionais no país requerem olhares amplos e estratégias rigorosas de apreensão e intervenção que não se restrinjam a fatores estritamente ligados a educação, abrangendo questões ligadas a emprego, renda e inclusão social.

As desigualdades educacionais no país podem ser compreendidas em quatro níveis: longuíssimo, longo, médio e curto prazo. No longuíssimo prazo, o passado colonial escravocrata compromete a posição de indivíduos pretos, pardos e indígenas, somando-se à desigualdade estrutural de gênero nas oportunidades sociais. No longo prazo, a literatura reforça o caráter geracional das desigualdades, decorrentes de diferentes condições parentais e habitacionais, que condicionam distintos acessos a recursos culturais, econômicos e educacionais. No médio prazo, destacam-se desigualdades de estrutura e qualidade escolar, enquanto, no curto prazo, questões conjunturais e incentivos podem alterar o desempenho dos sujeitos. A solução desses problemas requer múltiplas compreensões e intervenções.

Considerando o contexto de jovens concluintes do ciclo da educação básica em escolas brasileiras, este artigo visa compreender os determinantes da capacidade de aprendizagem e desempenho em testes padronizados via Exame Nacional do Ensino Médio (Enem) no país entre 2015 e 2023. Definem-se como principais nortes teórico as contribuições de Hanushek & Woessmann (2008; 2021) sobre qualidade educacional e aprendizado. Além disso, leva-se em conta os aspectos geracionais das desigualdades, especialmente embasado em estudos empíricos como Card & Krueger (1996), Hanushek (2003) e Hanushek *et al* (2005), sobre qualidade educacional e desempenho escolar e Freyer et al (2012) sobre qualidade docente e aprendizado esta proposta se divide em três etapas. Investigações empíricas relevantes dão suporte para compreensão de casos brasileiros, como Moraes et al (2022) que investigam o peso

do efeito escola sobre o aprendizado, Curi & Filho (2013), Pires (2015) e Figueiredo, Nogueira e Santana (2014) que investigam o peso do contexto familiar sobre o desempenho escolar de estudantes da educação básica.

A literatura indica que o desempenho acadêmico dos estudantes é resultado de insumos escolares e de características familiares e da trajetória individual de cada estudante. Todavia, especialmente no Brasil, onde há grande desigualdade de acesso e permanência nas escolas, abre-se espaço para discussão se de fato a elevação da qualidade escolar é capaz de prover ascensão econômica a estudantes de baixa renda, via elevação de sua aprendizagem e acúmulo cognitivo, ou se as melhores escolas estão restritas aos estudantes com melhores condições econômicas e familiares, o que reduziria o potencial transformador da educação a nível de redução da desigualdade social.

Este artigo se divide em três etapas. Primeiro, resgatam-se as principais contribuições de pesquisas teóricas e empíricas que subsidiam a compreensão do problema. Em segundo momento, é descrita a proposta metodológica da análise. Posteriormente, são apresentados os resultados econométricos. Primeiro, é realizado um modelo econométrico de regressões múltiplas para determinar o impacto de efeitos individuais e familiares sobre o desempenho da amostra ampla de estudantes analisados. Haja vista a limitação de dados associativos entre sujeito e escola onde estudou, em segundo momento a análise é focada em estudantes de municípios pequenos onde há apenas uma escola, de tal forma a perceber os efeitos amplos de características individuais, familiares e escolares de estudantes de aproximadamente 2800 municípios no país. Considerações finais sintetizam a abordagem e apontam caminhos para o aprimoramento desta pesquisa.

2. Referencial teórico e bibliográfico

Esforços para compreender a relevância econômica da educação expandiram-se após meados do século XX. Theodore Schultz foi essencial ao definir capital humano como o conjunto de habilidades e conhecimentos econômicos, em dimensões qualitativas e quantitativas, enraizados e sobrepostos à própria constituição dos indivíduos A primeira foca na escolaridade e horas trabalhadas; a segunda considera o nível de conhecimento, saúde e habitação, englobando tudo o que eleva a produtividade (Schultz, 1961).

Esta modalidade de capital apresenta uma singularidade que dificulta sua mensuração e representação em modelos econômicos. Diferentemente de capitais físicos que podem ser notados e comparados, conhecimentos e habilidades não são explícitas e manifestam-se de

maneira distinta não somente pela disparidade de sujeitos, mas também do contexto socioeconômico e produtivo no qual se encontram (Shaffer, 1961). A estratégia básica de superação destas limitações foi a padronização da apreensão via indicadores básicos de acesso à educação básica, notadamente via anos de estudo e sinalização via titulação acadêmica.

Um relatório "*Equality of Educational Oportunity"* publicado por James Coleman em 1966 contribui para um olhar mais abrangente sobre a relevância econômica da educação, que leva a uma compreensão multidimensional dos determinantes da aprendizagem e desempenho escolar. O documento tem como nortes principais para compreensão da qualidade e desempenho educacional, quatro questões: i) O contexto e trajetória de professores e estudantes, com ênfase em questões socioeconômicas e raciais em torno da segregação estadunidense; ii) as características e estrutura da instituição educacional; iii) as condicionantes do desempenho acadêmico e iv) no que cerne à compreensão multidimensional relação entre características da escola e efeito dos pares e o desempenho educacional. O relatório é amplo e acompanhado de metodologias que nortearam estudos educacionais ao longo do século XX (Coleman, 1966).

Ademais, Eric Hanushek reforça as ideias de Coleman, ao passo que reconhece os avanços da abordagem dos insumos educacionais. O processo educacional é dinâmico, complexo e cumulativo, resultado de múltiplos fatores dentro e fora do ambiente educacional, sendo resultante de insumos primários, diga-se influências familiares, efeitos dos pares e características da instituição educacional. Para além disto, a abordagem avançou para além das relação escola-aluno, e passou a se preocupar com questões relacionadas à relação aprendizado-produtividade, desempenho profissional e crescimento econômico (Hanushek, 1997; 2003).

Este avanço teórico, bem como das metodologias estatísticas de apreensão, possibilitou a solução de um problema já destacado por Harry Shaffer em 1961. Em virtude da escassez de dados e fragilidade metodológica, Shaffer (1961, p. 1026) destacou que *"were it possible to separate consumption expenditure from investment in man it would still remain a virtual impossibility to allocate a specific return to a specific investment in man"*. Ou seja, seria impossível desagregar o investimento em capital humano e determinar com precisão estatística os diferentes níveis de eficiência. Esta questão tende a ser superada com a maior disponibilidade de dados contemporânea e o avanço de instrumentos estatísticos empregados em abordagens sobre insumos educacionais.

Em avanço às ideias de Coleman e agregando múltiplas contribuições teóricas, metodológicas e instrumentais, a abordagem contemporânea dos insumos educacionais também visa superar o viés de análise de capital humano e capacidades cognitivas pautado meramente na escolaridade. Escolaridade em si não implica em melhoria de habilidades cognitivas e em

retornos econômicos a nível individual e coletivo. É preciso se ater à qualidade e ao tipo de educação que estimule as potencialidades humana em múltiplos aspectos. Aspectos como estrutura do ambiente educacional, efeito de pares, aspectos culturais, familiares e relacionados ao acesso e distribuição de renda são essenciais na determinação dos diferenciais de habilidades cognitivas entre dois indivíduos de mesmo nível de escolaridade (Hanushek; Woessmann, 2008).

Em linhas gerais, Hanushek e Woessmann (2021) mensuram o capital humano através de um modelo básico de insumos educacionais para determinação das habilidades cognitivas individuais, como indicado na Equação 1. No modelo, as habilidades cognitivas são expressão do capital humano e são captadas via desempenho em testes em leitura, matemática e ciências, os quais podem ser utilizados em pesquisas comparativas entre diferentes regiões e países.

$$H = \lambda F + \phi(qS) + \eta A + \alpha Z + v \qquad (1)$$

$$H = C + N + \mu \qquad (1a)$$

Em que H representa as habilidades do indivíduo, F representa os insumos e características familiares, qS expressa insumos e características da instituição educacional, A são características do estudante, Z outros fatores relevantes a serem considerados, v elementos não considerados ou erro. Ademais, λ, ϕ, η e α são parâmetros do modelo. Esta abordagem é interessante porque considera fatores relacionados às características socioeconômicas e institucionais que possam afetar a aprendizagem de indivíduos (Hanushek; Woessmann, 2021).

2.1 Revisão bibliográfica

Hanushek (2003) analisa 89 trabalhos publicados nos EUA até 1995, cuja ênfase se deu sobre a relação entre insumos escolares e desempenho educacional. O autor constatou que os principais insumos abordados nos trabalhos são relação professor-aluno, gasto por aluno, qualificação e experiência do professor. Hanushek reforça que desempenho acadêmico de longo prazo é mais resultado da organização alocativa estratégica dos recursos que do montante em si investido, mas aponta que nem sempre há uma linearidade entre insumos escolares e desempenho educacional, o que pode dizer respeito à fatores extra educacionais como a dinâmica demográfica por exemplo.

Ainda no que se refere à relação entre qualidade educacionais e desempenho acadêmico, destaca-se o trabalho de Hanushek *et al* (2005). Os autores criaram um indicador de valor

adicionado acadêmico que relaciona qualidade do professor e desempenho dos estudantes, compreendendo que docentes são qualificados não somente em virtude de sua titulação e experiência, mas fundamentalmente pelo produto de seu trabalho que é o aprendizado discente. Utilizaram um modelo de dados em painel que considerou o desempenho dos estudantes como variável a ser explicada em função de fatores educacionais, background dos alunos e diferenças individuais. O desempenho dos estudantes foi mensurado com base em testes padronizados em matemática e foram coletados juntamente com as demais informações nos registros administrativos educacionais do estado do Texas (EUA) entre 1989 e 2002.

Os fatores educacionais compreenderam dados dos estudantes como i) ID, um identificador para o aluno; ii) raça; iii) gênero; iv) elegibilidade para bolsas de auxílio estudantil, o qual indicaram vulnerabilidade social. Também foram considerados como aspectos educacionais as características dos professores relacionadas a i) raça; ii) qualificação; iii) anos de experiência; iv) desempenho em testes de certificação; v) vínculo trabalhista; vi) atribuições adequadas ou não a sua qualificação e vii) características do campus ou ambiente educacional. Constatou-se que os melhores alunos apresentam um "rendimento marginal decrescente" e, para corrigir problemas de auto seleção, dividiu-se os estudantes em grupos pautados em seu score inicial médio. Em síntese, constataram que bons professores impactam positivamente sobre alunos com melhores e nos com piores notas iniciais e que professores menos qualificados tendem a sofrer com maior rotatividade.

Fryer et al (2012) também admitem a relação positiva entre a qualidade docente e o desempenho acadêmico, porém consideram a dificuldade de mensuração da produtividade do trabalho docente. Os autores utilizaram um modelo ITT (*intent to treat*), considerando indivíduos randomizados para análise em comparação a um grupo de controle. Baseado em um programa de incentivos docentes realizado na cidade de Chicago Heights (EUA) com 150 professores e 3200 alunos de baixa renda, o qual atribuía bônus salariais aos professores cujas turmas apresentassem maior melhora no desempenho ou ônus àqueles responsáveis por discentes com piora acadêmica, os autores constataram que os efeitos de perda ou ônus foram mais eficazes como estímulo à melhoria dos resultados estudantis, apesar de em ambos os casos o programa apresentar um resultado positivo.

Adotou-se como variável dependente o desempenho dos estudantes em um conjunto de testes padronizados, os quais foram explicados com base nos incentivos e em dados administrativos que indicam a qualidade educacional. Os testes considerados foram o ISAT (International Student Admissions Test), ITBS (Iowa Test of Basic Skills) e Think Link. A qualidade educacional foi mensurada com base i) na adequação ou não das atribuições dos

professores com base em sua qualificação; ii) gênero do discente; iii) raça; iv) elegibilidade a bolsa de assistência estudantil; v) elegibilidade para serviços educacionais especiais; vi) status de baixa proficiência em língua inglesa e vii) resultados nos testes estudantis.

Ademais, Card e Krueger (1996) buscam relacionar investimentos em insumos educacionais com o acúmulo de capital humano, reconhecendo a relevância do *background* familiar na determinação da escolaridade e dos rendimentos futuros dos indivíduos. Os autores partem da premissa que a mensuração do capital humano via anos de escolaridade é limitada por desconsiderar a relevância da qualidade educacional bem como à bagagem prévia de cada indivíduo. Utilizaram de uma base de dados referente escolaridade e emprego de homens nos estados estadunidenses de Carolina do Sul e Carolina do Norte entre 1900 e 1959. Como explicação para diferenciação de rendimentos consideraram as variáveis i) proporção professor por aluno; ii) anos de escolaridade; iii) *dummies* para distinção entre nascidos em cada um dos estados e iv) raça, entre brancos e negros. Constataram significativa diferença de desempenho entre escolas onde há segregação racial e onde não há. Outra constatação relevante diz respeito à organização dos recursos educacionais, de modo que turmas menores e com mais capital investido por aluno tendem a gerar impactos positivos de médio prazo no rendimento dos estudantes.

Dufflo, Dupas e Kremer (2011) fazem um estudo empírico visando compreender o efeito dos pares sobre o desempenho dos professores, sobre o transbordamento de aprendizagem de aluno para aluno e sobre a escolha do professor sobre o nível de exigência do ensino. Analisando um experimento que durou 18 meses e envolveu 121 escolas no Quênia entre 2005 e 2007, determinaram que o desempenho dos alunos foi mensurado pelo i) score pré testes; ii) média da nota dos outros alunos na turma; iii) esforço do professor; iv) nível de dificuldade determinado pelo professor para suas aulas e v) outros fatores estocásticos específicos aos alunos. Como resultado, percebeu-se que alunos medianos foram beneficiados do contato com alunos melhores, houve maior assiduidade e esforço pelos professores, evidenciando o beneficiamento coletivo para o aprendizado.

3. Proposta metodológica

3.1 Base de dados

O somatório das notas nas áreas do conhecimento Ciências da Natureza (CN), Ciências Humanas (CH), Códigos e Linguagens (LC), Matemática (MT) e na Prova de Redação no Exame Nacional do Ensino Médio (Enem) foi a variável utilizada para mensurar o desempenho,

por considerar que há uma grande base disponível que agrega informações abrangentes no que diz respeito ao perfil individual, familiar e socioeconômico dos estudantes brasileiros. O Enem foi criado em 1998 com o objetivo de referenciar a qualidade dos egressos da educação básica e servir de parâmetro para instituições de ensino superior sobre o perfil de admissão de ingressantes. Após reformulação e padronização, desde 2009 é a principal porta de entrada para o ensino superior no Brasil.

Foram considerados somente indivíduos que realizaram todas as provas, não participaram como treineiros, não anularam e tiveram a redação passível de ser corrigida. Excluíram-se os que não estavam matriculados em turmas do último ano do Ensino Médio em escolas ativas, com ofertas de turmas de Ensino Médio, e reconhecidas pelo MEC. Deste modo, cada observação remete única e exclusivamente a um indivíduo. Foram excluídos da amostra sujeitos com informações omissas, implicando em uma amostra total de 6.672.913 observações de um total de 50.296.609, ou aproximadamente 13,27% das pessoas que realizaram o Enem, entre os anos 2015 e 2023.

Os microdados do Censo da Educação Básica do Inep revelam informações de aproximadamente 231.745 escolas em média por ano, englobando todas as unidades de ensino na educação básica do país e representando uma amostra global de 1.622.216 observações. Considerando apenas escolas ativas com turmas alunos matriculados em uma ou mais turmas do Ensino Médio, a amostra global é reduzida para 200.570. Para o estudo, considerou-se apenas observações que não apresentaram informações faltantes para informações das escolas e dos municípios, de forma que a amostra global incorpora informações de 51.620 escolas, aproximadamente 25,27% da amostra global, das quais 35.944 são públicas urbanas, 6.420 são públicas rurais, 8.687 são privadas urbanas e 569 são escolas privadas rurais, em todos os 5.570 municípios brasileiros entre 2015 e 2021. A inconsistência temporal das bases de dados não é um problema em decorrência das escolhas metodológicas para as três estimações adotadas, como será melhor explicado posteriormente.

A heterogeneidade das escolas foi considera com base na dependência administrativa, se de natureza pública ou de natureza privada. Também se considerou a localização, considerando as peculiaridades de unidades educacionais em zonas urbanas ou rurais dos municípios brasileiros, além da complexidade da gestão escolar, da estrutura e capacidade de oferta de serviços de ensino de qualidade. Foi utilizado o Índice de Complexidade da Gestão Escolar (ICG)[3], um indicador que pode ser utilizado como parâmetro para expressar o padrão

[3] Para mais informações acerca deste indicador, recomenda-se a leitura de https://www.gov.br/inep/pt-br/acesso-a-informacao/dados-abertos/indicadores-educacionais/complexidade-de-gestao-da-escola.

estrutural de uma unidade educacional, possibilitando comparações e diferenciações entre perfis educacionais por segmentos em cada município. Além disso, foram consideradas variáveis para considerar a qualidade estrutural das escolas, referentes à existência de bibliotecas e/ou salas de leituras, laboratórios de ciências, de informática, acesso à internet e disponibilidade de refeição, com base em informações censitárias do Inep.

Por sua vez, a qualidade do trabalho do professor é afetada pelo grau de sobrecarga. O Inep fornece o indicador de Esforço Docente (IED) para mensurar o esforço despendido pelo docente em função da quantidade de turmas, turnos, etapas de ensino e escolas onde atua. Este indicador é interessante porque atribui diferentes pesos aos níveis de esforço, o que possibilita capturar com maior sensibilidade os níveis mais elevados de sobrecarga e permite comparar os diferentes objetos, sejam escolas ou municípios, com base em uma medida tanto de intensidade quanto de distribuição de esforços (Nardo et al, 2005).

No que diz respeito a heterogeneidade dos indivíduos, levou-se em conta particularidades de gênero, raça/cor e condições socioeconômicas dos pais, em função das historicamente reconhecidas desigualdades socioeconômicas do país entre homens e mulheres, entre brancos, negros e indígenas, especialmente, e decorrentes das diferentes trajetórias familiares. A variável mulher é binária, indicando se o indivíduo é do gênero feminino (1) ou masculino (0). No que diz respeito a raça foram criadas variáveis *dummies* para indicar se o indivíduo é branco, preto, pardo, amarelo ou indígena, conforme autodeclaração. Já as condições socioeconômicas foram captadas especialmente via questionários dos microdados do Enem.

Acerca das condições sociohabitacionais dos estudantes, considerar-se-á aqui o uso da variável banheiro, a partir da declaração individual em pesquisa realizada no momento da realização da prova do Enem, com o intuito de abastecer a análise com informações sobre a qualidade da residência e condições de vida básicas dos estudantes. A variável indica a qualidade das condições socio-habitacionais dos indivíduos. Somado a estas variáveis, serão consideradas informações referentes à escolaridade e ocupação dos pais, bem como renda familiar.

O grau de escolaridade, desde Ensino Fundamental (EM), Médio (EM), Superior (ES) e pós-graduação e o perfil da ocupação parental são fatores que influenciam significativamente a capacidade de aprendizagem dos indivíduos. A renda familiar foi considerada em termos de quantidade de salários-mínimos (SM) do ano vigente por família, variando de 0 até renda familiar superior a 20 SM, fim de possibilitar comparação ao longo do tempo.

3.2 Procedimentos econométricos

A investigação sobre os determinantes do desempenho escolar via modelos de regressões múltiplas para o conjunto de estudantes que realizou o Enem no período especificado, captando especialmente o efeito sobre indivíduos que estudam em escolas localizadas em cidades pequenas, por possibilitar alinhar diretamente características individuais, familiares e escolares de maneira plena. Por fim, será estimado um modelo hierárquico, para apreender o peso do "efeito-escola" ou "efeito-individuo" na determinação da aprendizagem.

Wooldridge (1996) considera o modelo de dados em painel adequado para análise de uma série de variáveis sobre um grupo de indivíduos ao longo do tempo. Alia em suas proporções estudos de cortes transversais com séries de tempo, permitindo que se compare a evolução e dinâmica dos indivíduos sob efeito de variáveis independentes na série de tempo trabalhada. Difere do método de cortes transversais, pois não extrai informações de indivíduos aleatórios, o que possibilita que se apreenda a dinâmica das particularidades ao longo do tempo.

Outra importância da utilização de dados em painel é ressaltada por Gujarati e Porter (2009). Os autores ressaltam que o painel fornece o controle das heterogeneidades com características próprias de cada indivíduo ao passo que permite compreensão da dinâmica de seu comportamento ao longo do tempo. A utilização do modelo de dados em painel se justifica também por sua capacidade de fornecer informações com menos colinearidade e mais graus de liberdade, o que possibilita melhor ajustamento, de modo que se adequam bem ao estudo de um grupo fechado de indivíduos que modificam padrões comportamentais ao longo do tempo (Baltagi, 2008).

Será considerada a possibilidade de o modelo de dados em painel ser *pooled*, efeitos fixos ou efeitos aleatórios, conforme ressalta Greene (2002), de modo que testes serão realizados para considerar o mais adequado à pesquisa, em conformidade com a base de dados levantada. Ainda conforme Greene (2002), a regressão é *pooled* quando há apenas um intercepto $\beta 0$ para todos os sujeitos analisados, o que significa que somente os parâmetros de inclinação variam resultando em estimadores consistentes e eficientes. Regressões por efeitos fixos por sua vez apresentam uma constante βi que não varia ao longo do tempo, para cada indivíduo. Já efeitos aleatórios são modelos com uma variável erro ui aleatória para cada indivíduo.

Os modelos hierárquicos ou multiníveis podem ser utilizados para lidar com estudos com características peculiares, onde os objetos de análise apresentam similaridades intragrupos

e particularidades entre diferentes grupos. Modelos convencionais utilizam-se de preditores específicos para captar efeitos aleatórios sobre padrões de indivíduos, enquanto modelos hierárquicos possibilitam apreender efeitos aleatórios em diferentes níveis de influência sobre os elementos analisados. O modelo apresenta-se como um modelo linear misto, de modo que a variável dependente é afetada em diferentes níveis pelas características das variáveis independentes (Dedrick et al, 2009).

Modelos multiníveis são usualmente utilizados em investigações sobre educação por sua capacidade de apreender efeitos aleatórios sobre diferentes padrões de grupos de indivíduos ou escolas. Por exemplo, permitem captar as características internas às classes ou às escolas, ou mesmo em cada município, ao passo que expressam as particularidades de cada sujeito. Captam a variância de cada efeito aleatório e permitem apreender com rigor quais níveis são mais significativos na explicação da variável dependente, enquanto ainda fornecem parâmetros referentes a cada variável individualmente (Dedrick et al, 2009).

O modelo hierárquico é importante porque considera não somente os efeitos fixos dos parâmetros de um modelo convencional, mas destaca efeitos aleatórios em níveis que podem ter pesos hierarquicamente distintos na estimação sobre a variável dependente. No modelo considerado, adotou-se três efeitos aleatórios: individuais, que consideraram o peso de parâmetros raciais, de gênero, condições residenciais e familiares, como escolaridade, ocupação e renda dos pais, enquanto informações escolares englobaram aspectos estruturais da escola, indicadores escolares como IED, ICG, bem como aspectos municipais, haja vista que cada município abrange única e exclusivamente uma escola. Além disso, destaca-se aspectos aleatórios sobre cada ano, com o intuito de melhor apreender efeitos de flutuação na amostra. Destaca-se a incorporação da *dummy* Covid, referente ao contexto pandêmico da Covid-19 que perdurou entre 2020 e 2022, abrangendo 2 anos na amostra e ressaltando a relevância da captação da variância no tempo com vistas a perceber padrões expressos a cada ano entre as variáveis observadas.

Primeiramente, estima-se um modelo econométrico de regressões múltiplas que considerou como variável dependente a nota individual no Enem no período entre 2015 e 2023, de cada um dos 6.672.913 estudantes analisados, com base nas informações dos microdados da educação básica e microdados do Enem, disponibilizados pelo Inep. As variáveis foram trabalhadas em logaritmo por duas razões. Primeiro, para limitar problemas de heterocedasticidade, ao linearizar os dados e reduzir suas dispersões, e também para melhor captar a elasticidade entre variáveis dependentes e independentes (Wooldridge, 1996; Greene, 2002; Baltagi, 2005).

$$\begin{aligned}\log\left(Enem_{id\,it}\right) = {} & \beta_0 + \beta_1 Mulher + \beta_2 Preta + \beta_3 Parda + \beta_4 Amarela \\ & + \beta_5 Indigena + \beta_6 publica + \beta_7 rural + \beta_8 log\left(escolaridade_{pai} + 1\right) \\ & + \beta_9 log(escolaridade_{mae} + 1) + \beta_{10} \log\left(renda_{familiar} + 1\right) \\ & + \beta_{11} \log\left(ocupacao_{pai} + 1\right) + \beta_{12} \log(ocupacao_{mae} + 1) \\ & + \beta_{13} \log(banheiro + 1) + \beta_{14} \log(computador + 1) \\ & + \beta_{15} internet + COVID + u_{it}\end{aligned} \quad (1)$$

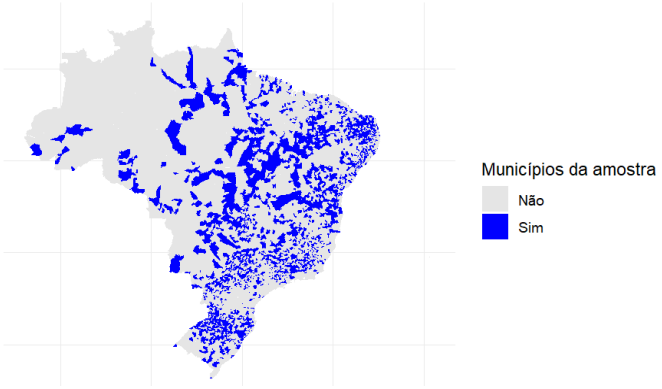
A Regressão 2 é simultaneamente mais completa e mais restrita. Mais completa porque abrange mais informações por indivíduo, vinculando-o não somente a elementos familiares, mas especialmente a aspectos sobre a qualidade da escola onde concluiu o Ensino Médio. Isto favorece uma compreensão ampla sobre as múltiplas determinações do aprendizado. Por sua vez, são limitantes ao conjunto de indivíduos que concluíram seu ciclo na educação básica em pequenos municípios onde há somente uma escola em funcionamento. O período engloba os anos de 2015 a 2021, em alinhamento ao censo da educação básica que apresenta informações completas a nível escolar para cada estudante que realizou o Enem. Isto foi necessário como estratégia de superação da dificuldade de acesso aos dados identificados no país[4]. Como é impossível definir, por meio dos microdados do Enem ou do Censo da Educação Básica, a escola de origem da população ampla, adotou-se como escolha estudantes de municípios pequenos, pois como há somente uma escola no município e sabe-se o município de origem, foi possível fazer uma conexão direta entre estudante e escola.

Esta amostra engloba um pouco mais de 2800 municípios e de 500 mil estudantes. Após a conjugação de dados completos entre todas as escolas e municípios, computou-se mais 400 mil estudantes na amostra da Regressão 2, o que garante uma representação estatística para parte relevante do território e população nacionais. A Estimação 3 refere-se ao modelo hierárquico e abrange a mesma amostra da Estimação 2.

A Figura 2 apresenta a abrangência dos municípios que propiciam uma análise ampla dos estudantes, englobando aspectos individuais, familiares, escolares e socioeconômicos. Sua visualização demonstra a distribuição espacial dos municípios analisados em todo o território brasileiro.

[4] Em decorrência da Lei Geral de Proteção de Dados (LGPD), de 2018, há restrições no que diz respeito ao acesso a dados identificados, seja a nível de indivíduos ou mesmo de escolas. Em decorrência disso, não é possível auferir a qualidade de uma escola com base no desempenho médio de seus estudantes para todo o período amostral dos microdados do Enem, pois estes dados não são de acesso público. Na mesma linha, não é possível vincular estudantes às instituições onde estudaram. Portanto, a estratégia adotada objetivou minimizar estas limitações, sem comprometer o rigor metodológico.

Figura 2: Distribuição espacial dos municípios com somente uma escola de Ensino Médio cadastrada no Censo da Educação Básica



Fonte: Elaboração própria com Base em Inep (2025).

Nesta linha, tem-se a Estimação II

$$\begin{aligned}\log(Enem_{id_{it}}) = {} & \beta_0 + \beta_1 Mulher + \beta_2 Preta + \beta_3 Parda + \beta_4 Amarela \\ & + \beta_5 Indigena + \beta_6 publica + \beta_7 rural + \beta_8 log(escolaridade_{pai} + 1) \\ & + \beta_9 log(escolaridade_{mae} + 1) + \beta_{10} \log(renda_{familiar} + 1) \\ & + \beta_{11} \log(ocupacao_{pai} + 1) + \beta_{12}(ocupacao_{mae} + 1) \\ & + \beta_{13} \log(banheiro + 1) + \beta_{14} \log(computador + 1) \\ & + \beta_{15} internet_{casa} + \beta_{16} \log(Remuneracao) \\ & + \beta_{17} \log(IED) + \beta_{18} Biblioteca + \beta_{19} Lab_{Ciencias} \\ & + \beta_{20} Lab_{informatica} + \beta_{21} Refeitorio + \beta_{22} Internet_{escola} \\ & + \beta_{23} \log(Aluno_{turma}) + \beta_{24} \log(Horas_aula_{med}) \\ & + \beta_{25} \log(Prof_{turma}) + \beta_{26} \log(ICG) + \beta_{27} \log(PIB) \\ & + \beta_{28} \log(POPULACAO) + \beta_{29} \text{COVID} + u_{it}\end{aligned} \tag{2}$$

Por sua vez, a Estimação III refere-se ao modelo hierárquico

$$\begin{aligned}\log(Enem_{id_{it}}) = {} & \beta_0 + \beta_1 Mulher + \beta_2 Preta + \beta_3 Parda + \beta_4 Amarela \\ & + \beta_5 Indigena + \beta_6 banheiro + \beta_7 computador + \beta_8 internet_{casa} \\ & + \beta_9 internet_{escola} + \beta_{10} Biblioteca + \beta_{11} Lab_{Ciencias} \log(ocupacao_{pai} + 1) \\ & + \beta_{12} Lab_{informatica}(ocupacao_{mae} + 1) + \beta_{13} Refeitorio \\ & + \beta_{14} \log(renda_{familiar} + 1) + \beta_{15} \log(Escolaridade_{pai}) \\ & + \beta_{16} \log(Escolaridade_{mae}) \\ & + \beta_{17} \log(Ocupacao_{pai}) + \beta_{18} \log(Ocupacao_{mae}) \\ & + \beta_{19} \log(Remuneracao) + \beta_{20} \log(IED_{MED}) \\ & + \beta_{21} \log(Aluno_{turma}) + \beta_{22} \log(horas_{aula_{MED}}) \\ & + \beta_{23} \log(Aluno_{turma}) + \beta_{24} \log(Horas_{aula_{med}}) \\ & + \beta_{25} \log(Prof_{turma}) + \beta_{26} \log(ICG) + \beta_{27} \log(PIB) \\ & + \beta_{28} \log(POPULACAO) + \beta_{29} \text{COVID} + i_j + e_k + a_l + u_{it}\end{aligned} \tag{3}$$

Em que i_j, e_k e a_l são respectivamente os interceptos aleatórios para escola, estudante e ano, respectivamente, indicando os três níveis considerados no modelo da Estimação 3. Com o objetivo de corrigir problemas de escala, adotou-se a estratégia de se converter as variáveis Banheiro e Computador em binárias e logaritimizar as demais variáveis numéricas. Além disso, as variáveis foram reescalonadas a partir da subtração da pela média e divisão pelo desvio padrão.

Referente à ambas as estimações, tem-se a variável $Mulher$ como uma dummy tal que é igual a 1 se o indivíduo for do gênero feminino ou 0 caso seja masculino[5]. São consideradas as variáveis binárias $Preta,\ Parda,\ Amarela$ e $Indigena$ para indicar a raça/cor do indivíduo, tal que são iguais a 1 se autodeclarados da raça/cor mencionada e igual a 0 caso contrário; a cor branca foi mantida como *benchmark* no modelo. A *dummy* $publica = 1$ se a escola onde o indivíduo estudou é pública e igual a 0 caso seja privada, enquanto a variável $rural = 1$ caso a escola esteja localizada em zona rural e é igual a 0 caso esteja em perímetro urbano.

As variáveis $Lab_{Ciencias}$, $Lab_{Informatica}$, $refeitorio$, $biblioteca$ e $internet$ são variáveis binárias que indicam a existência destes elementos na estrutura das escolas brasileiras. Por sua vez, ICG indica o grau de complexidade da gestão docente, simbolizando o porte e dimensão da escola, IED_{MED} refere ao índice de esforço docente, expressando o grau de sobrecarga destes profissionais em seus ambientes de trabalho.

As variáveis $Aluno_{turma}$ e $Prof_{aluno}$ indicam respectivamente a proporção de alunos por turmas de ensino médio e a proporção de professores por turmas de ensino médio em cada escola na amostra. As variáveis PIB_{it} e $POPULACAO_{it}$ indicam respectivamente o Produto Interno Bruto do município *i* no tempo *t* e população do município *i* no tempo *t*, $\beta_{12}Covid$ é uma variável binária para os anos de 2020 e 2021, que foram marcados pelo contexto pandêmico e u_{it} representa os resíduos.

As variáveis $log(escolaridade_{pai} + 1)$ e $log(escolaridade_{mae} + 1)$ referem-se, respectivamente, ao maior grau de escolaridade alcançado pelo pai e pela mãe do indivíduo, variando de 0 a 5. Por sua vez, $\log(ocupacao_{pai} + 1)$ e $(ocupacao_{mae} + 1)$ referem-se ao grupo ocupacional do pai e da mãe do educando, em escala de 1 a 6 respectivamente. A variável $\log\ (renda_{fam} + 1)$ indica a renda familiar do indivíduo, em quantidade de salários-mínimos.

[5] As informações disponíveis nos questionários educacionais do INEP não consideram demais categorias de gêneros.

Por sua vez, $\log(banheiro + 1)$ indica a quantidade de banheiros na residência dos estudantes, como proxy da qualidade habitacional, enquanto $\log(computador + 1)$ indica quantos computadores são disponíveis para o estudante em sua residência. Reforça-se que as variáveis estão em logaritmo a fim de estabilizar a variância e mitigar efeitos de multicolinearidade dos dados. Ademais, as variáveis logarítmicas somam 1 unidade para evitar a possibilidade de indefinição do log de 0. Por fim, $internet$ é uma variável binária igual a 1 caso o estudante tenha acesso à internet em sua residência e igual a 0 caso não tenha, enquanto u_{it} representa os resíduos.

Espera-se com estas investigações, compreender melhor os determinantes da desigualdade educacional no país, considerando as heterogeneidades individuais e a nível municipal, atentando-se aos efeitos por categorias de grupos escolares, bem como aspectos socioeconômicos e técnicos que afetam o aprendizado e desempenho acadêmico.

4. Análise de resultados econométricos

A verificação da validade da Estimação I se deu pela testagem e eventual correção dos problemas de heterocedasticidade, via teste de Breusch-Pagan, cuja hipótese nula é homoscedasticidade. Também se utilizou o teste de variância inflacionária do fator (VIF) para verificar multicolinearidade, ou seja, a correlação entre duas ou mais variáveis dependentes. Foi verificada a existência de endogeneidade, ou seja, a correlação entre variáveis independentes e resíduos no modelo, a partir da realização do teste de Hausman. Pelos testes, verificou-se a existência de heterocedasticidade, o que foi corrigido via erros padrão robustos e não foram constatados problemas de multicolinearidade ou endogeneidade.

No que diz respeito à Estimação II, utilizou-se o teste de Breusch-Pagan e verificou-se o problema de heterocedasticidade, o qual foi corrigido via erros padrão robustos. Foram utilizados os testes de VIF e de Hausman e não se constatou problemas de multicolinearidade ou endogeneidade no modelo. Sobre a Estimação III, não foram constatados efeitos de multicolinearidade e o modelo convergiu tecnicamente, a partir do otimizador *nloptwrap*.

Tabela 1: Resultado das Estimações I, II e III

Coeficiente	Estimação I	Estimação II	Estimação III
Intercepto	7.759649991 *** (0.000380291)	7.5778472 *** (0.0065307)	7.832000 *** (0.005475)
mulher	-0.001713306 ***	0.0013093 ***	0.0009807 **

Coeficiente	Estimação I	Estimação II	Estimação III
	(0.000098376)	(0.0003799)	(0.0003642)
Preta	-0.021243951 ***	-0.0270515 ***	-0.01728 ***
	(0.000168563)	(0.0006853)	(0.0006906)
Parda	-0.017002079 ***	-0.0211928 ***	-0.01285 ***
	(0.000110083)	(0.0005447)	(0.0004466)
Amarela	-0.009406022 ***	-0.0116614 ***	0.006238 ***
	(0.000332890)	(0.0013380)	(0.001262)
Indigena	-0.046080050 ***	-0.0140432 ***	-0.01203 *
	(0.000633749)	(0.0009272)	(0.01627)
Publica	-0.076282126 ***		
	(0.000142164)		
Rural	0.002563565 ***		
	(0.000285356)		
Banheiro	0.018920686 ***	0.0022194 ***	0.0009374 ***
	(0.000237510)	(0.0000429)	(0.0004030)
Computador	0.032004738 ***	0.0021206 ***	0.003791 ***
	(0.000146302)	(0.0000410)	(0.0003791)
Escolaridade_Pai	0.019915012 ***	0.0016299 ***	0.008162 ***
	(0.000145091)	(0.0000524)	(0.0002295)
Escolaridade_Mae	0.027594269 ***	0.0025027 ***	0.01015 ***
	(0.000158641)	(0.0000513)	(0.0002340)
Renda_Familiar	0.035340267 ***	0.0049629 ***	0.1800 ***
	(0.000118004)	(0.0001425)	(0.005247)
Ocupacao_Pai	0.021480175 ***	0.0036816 ***	0.00066 **
	(0.000197791)	(0.0000911)	(0.0002356)
Ocupacao_Mae	0.012299594 ***	0.0009159 ***	0.001960 ***
	(0.000213165)	(0.0000819)	(0.0002515)
Internet_casa	-0.021017957 ***	0.0022282 ***	0.004863 ***
	(0.000136304)	(0.0000657)	(0.0003581)
log(Remuneracao)		0.0037817 ***	0.006421 ***
		(0.0000931)	(0.0004064)
log(IED)		-0.0023650 ***	(0.02014) ***
		(0.0001812)	(0.001946)
Biblioteca		0.0001082 **	0.001124

Coeficiente	Estimação I	Estimação II	Estimação III
		(0.0000419)	(0.001012)
Lab_Ciencias		0.0001065 .	-0.0007014
		(0.0000588)	(0.0007120)
Lab_Informatica		0.0000522	0.001309 .
		(0.0000536)	(0.0007156)
Refeitorio		0.0000113	0.003119 ***
		(0.0000515)	(0.0008187)
Internet_escola		-0.0000605	-0.001836
		(0.0000497)	(0.0001439)
log(Aluno_turma)		-0.0021067 ***	0.0005878
		(0.0001306)	(0.0005.567)
log(HORAS_AULA)		0.0022766 ***	0.0005813
		(0.0001525)	(0.0005784)
log(Prof_Turma)		0.0022666 ***	0.004916
		(0.0002187)	(0.003763)
log(ICG)		0.0030420 ***	0.001409 *
		(0.0003040)	(0.0006735)
log(PIB)		0.0015278 ***	0.0004598
		(0.0001379)	(0.0005045)
log(POPULACAO)		0.0003680 ***	0.001219 (0.001441)
		(0.0000626)	
COVID	0.015927523 ***	-0.0050000 ***	0.002005
	(0.000123141)	(0.0006244)	(0.002482)

Signif. codes: 0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
Modelo Hierárquico:
Efeitos aleatórios:

Grupos		Variância	Desvio-Padrão
Id	intercepto	1.167^{-02}	1.080^{-01}
Escola	intercepto	4.615^{-03}	6.793^{-02}
Ano	intercepto	1.401^{-04}	1.167^{-02}
Resíduos		1.168^{-10}	1.081^{-05}

Observações: 403.683; grupos: id: 381.686; Escola: 2801; ano: 6.

Os resultados da Estimação I permitem conceber efeitos individuais e familiares, revelando o caráter geracional da externalidade dos efeitos socioeconômicos sobre qualidades educacionais. Aspectos escolares são restritos à dependência administrativa da escola, se pública ou privada, e à localização se em espaço urbano ou rural. Todas as variáveis foram estatisticamente significativas com margem de erro inferior a 0,01%, o que revela a qualidade

estatística da proposta. O R^2 ajustado de 0,3545 revela o já esperado pela literatura. Aspectos individuais e familiares desempenham um papel parcial, apesar da suma importância, na determinação da aprendizagem. Os resultados no geral corroboram com a literatura no que diz respeito aos efeitos de aspectos individuais e do contexto familiar sobre o aprendizado e desempenho escolar dos indivíduos. Os dados apontam que há uma disparidade de gênero que implica em resultados aproximadamente 0,17% inferiores de mulheres em relação a homens. No que diz respeito a questões raciais, a questão é mais latente: pessoas pretas têm em média desempenho 2,12% inferior que brancas e para indígenas esta diferença equivale a aproximadamente 4,6%.

Estes resultados corroboram a hipótese da desigualdade estrutural brasileira no que cerne a aspectos raciais e de gênero. A condição individual, independente do contexto familiar e socioeconômico, é objeto de desigualdade ainda no ambiente da educação básica. Estes aspectos tendem a reforçar as disparidades no contexto pós-educacional simbolizadas em piores condições de trabalho e remuneração para indivíduos que não sejam homens e cuja característica racial prevalescente seja preta ou parda. Considerando que a renda média familiar de estudantes pretos e pardos é aproximadamente 50% inferior à de famílias brancas e que as ocupações de seus pais e mães são menos valorizadas é nítida a externalidade geracional entre aspectos familiares e características educacionais individuais dos estudantes.

Externalidades geracionais são visíveis pelos parâmetros referentes à escolaridade, ocupação e renda familiares, revelando desigualdades prévias ao próprio contexto escolar. Pais e mães com maior escolaridade contribuem para uma elevação média de 1,99% e 2,75% no aprendizado dos filhos, enquanto famílias mais ricas favorecem o desempenho dos filhos em aproximadamente 3,53% no Enem. Por sua vez, a qualidade do emprego dos pais e mães também afetam positivamente o desempenho dos filhos, implicando em melhoria média de 2,14% e 1,22% na nota, respectivamente. Banheiro foi utilizada como *proxy* de qualidade habitacional e refletiu positivamente sobre o aprendizado dos indivíduos. Melhores condições de habitação favorecem o aprendizado, o que também é reforçado pelo acesso a computadores em sua residência, o que aumenta a nota média em 1,89% e 3,20% respectivamente. A única variável que contrariou a literatura foi Internet. Os dados apontam que o acesso doméstico a internet reduz a nota média do estudante em aproximadamente 2,10%, o que exige um estudo mais aprofundado para compreensão deste fato destoante.

Este modelo revela que a busca pela melhoria na aprendizagem formal exige um olhar para fora da escola. Políticas educacionais, políticas econômicas e políticas sociais devem ser pensadas de forma síncrona, com objetivos similares em torno da melhoria das condições

familiares, minimização das desigualdades individuais de gênero e raça e melhoria dos ambientes escolares. Medidas econômicas que favoreçam condições de trabalho e habitação dos pais refletem significativamente sobre a capacidade de aprendizado dos filhos, que por sua vez desenvolvem formalmente habilidades cognitivas no ambiente escolar, reduzindo a tendência de manifestação posterior de desigualdades no mercado de trabalho. Por conseguinte, políticas públicas que objetivem a melhoria dos indicadores educacionais não devem se restringir à educação pública ou mesmo à educação, abrangendo instrumentos múltiplos em torno de objetivos comuns de médio e longo prazo.

No que diz respeito ao perfil da escola, estudantes de escolas públicas têm em média uma nota 7,62% inferior a estudantes de escolas particulares, enquanto estudantes rurais têm desempenho superior médio de 0,25% se comparados a estudantes de instituições urbanas, o que é em muito impulsionado pela qualidade relativa das instituições privadas nestas localidades, se comparadas às públicas. A qualidade da escola pode ser mais bem percebida na Estimação II que vincula os estudantes às escolas e permite um olhar mais profundo sobre os múltiplos aspectos que afetam a aprendizagem.

A Estimação II evidencia os efeitos que o aprendizado do indivíduo ao final do ciclo da educação básica, mensurado pelo desempenho na prova do Enem, sofre múltiplas influências. As únicas variáveis sem relevância estatística foram a presença de refeitórios, laboratórios de informática e acesso à internet em escolas. Por sua vez, presença de bibliotecas em escolas é muito pouco expressiva na amostra, implicando em uma elevação média 0,01% na nota em testes padronizados, sob um intervalo de confiança de 0,1%, o que pode ser reflexo de sua grande abrangência no recorte estabelecido de cidades pequenas, não incorporando variações estatisticamente relevantes para a análise.

Os resultados reforçam o que é evidenciado na Regressão 1 sobre as desigualdades raciais no que diz respeito ao desempenho escolar entre pessoas pretas, pardas, amarelas e indígenas se comparadas a pessoas brancas. Todavia, contraria a regressão anterior ao apontar que mulheres têm desempenhos 0,13% superiores a homens em cidades pequenas onde há somente uma escola cadastrada no censo escolar.

A sobrecarga de trabalho docente prejudica o aprendizado em aproximadamente 0,23%, o que apesar de pouco expressivo, é preocupante Este dado revela a necessidade da elevação da contratação de professores, especialmente em cidades pequenas, haja vista que a escassez de profissionais implica na sobrecarga de trabalho e consequente prejuízo ao desempenho escolar dos estudantes. Somado ao fato de que melhores remunerações de professores elevam em média o desempenho dos estudantes no Enem em 0,37%, faz-se latente pensar em melhores estratégias

para estimular e atrair profissionais na área, com melhores remunerações e condições de trabalho, beneficiando a comunidade escolar como um todo.

A Estimação III revela os resultados do modelo de painel hierárquico, considerando os efeitos aleatórios ligados aos indivíduos (id), às escolas (Escola) e aos anos (ano). Todos apresentaram uma baixa variância, o que significa que características não explicadas pelos efeitos fixos são bem captadas pelos efeitos aleatórios. Em outras palavras, há um padrão observável entre o comportamento de variáveis individuais, escolares e há um padrão de variação entre ano. Por não haver muita variabilidade, nota-se que a escolha destes níveis capta bem as explicações sobre o desempenho. Especificamente, é notável que a variância de id é menor que a variância de Escola, o que permite conceber que aspectos individuais e familiares têm um peso maior na explicação da aprendizagem e desempenho entre estudantes do ensino médio, que variáveis escolares. Ou seja, o peso familiar, geracional e intergeracional, é mais determinante que aspectos institucionais ou estruturais do ambiente escolar, dentro do espaço amostral analisado.

A interpretação do modelo hierárquico é sensivelmente distinta de um modelo de regressão convencional, porém os princípios são os mesmos. Pela análise dos efeitos fixos, é notável que dos 28 parâmetros, 19 são estatisticamente significativos. Destaca-se que, referente a estrutura das escolas, apenas o efeito da existência de internet e bibliotecas nas instituições escolares têm significância estatística, o que surpreende o fato de escolas com bibliotecas ou salas de leituras reduzirem em média a nota de estudantes em 0,002% enquanto a internet contribui para desempenhos aproximadamente 0,004% melhores.

Ademais, os resultados reforçam as informações da Estimação II, a qual considera que mulheres têm notas sensivelmente maiores que dos homens, enquanto indivíduos pretos, pardos, amarelos e indígenas têm desempenho médio inferior ao de estudantes de cor branca. Aspectos do contexto familiar como escolaridade, ocupação e renda dos pais são altamente significativos estatisticamente, o que reforça que a desigualdade é prévia ao contexto escolar e que estratégias de melhoria do aprendizado devem ser pensadas em escala intergeracional: é necessário melhorar as condições socioeconômicas de emprego e renda de indivíduos para que seus filhos tenham um maior potencial de aprendizado.

Estes resultados isoladamente parecem não ser relevantes. Todavia, ao se considerar os efeitos cumulativos e padrões de características estruturais e organizacionais entre instituições públicas, privadas, rurais e urbanas, assim como ao se atentar às disparidades socioeconômicas entre as famílias brasileiras, é notável a complexidade de aspectos que afetam o desempenho de estudantes nas escolas do país. Neste sentido, faz-se mister se pensar multidimensionalmente

o problema da educação brasileira, para além de questões internas ao ambiente escolar, enquadrando questões relativas ao mercado de trabalho e políticas sociais amplas que não se restrinjam ao estudante ou profissionais de instituições educacionais.

Considerações finais

A precariedade educacional é uma realidade no país. A baixa condição de aprendizado dos estudantes no ciclo de educação básica que implica em desigualdades no acesso e na qualidade ao mercado de trabalho são resultantes, por sua vez, de desigualdades prévias ao próprio ambiente escolar. O estudo dos dados censitários da educação básica e dos microdados do Enem, ambos fornecidos pelo Inep, revelam que apesar da melhoria geral do aprendizado escolar nos últimos anos, ainda persiste uma desigualdade racial e socioeconômica, marcada pelo melhor desempenho e condições familiares de estudantes brancos, se comparados a sujeitos de outras cores/raças. Além disso, é notável que os mesmos estudantes que apresentam melhores desempenhos são filhos de pais com melhores níveis de escolaridade, ocupação e renda, o que reforça o caráter intergeracional de manutenção e perpetuação da desigualdade.

A Estimação I considerou a amostragem de todos os estudantes que realizaram o Enem, vinculados a escolas em funcionamento, pela primeira e única vez, não zerando ou abandonando nenhuma prova. Todos os parâmetros foram altamente significativos e indicam que características individuais e familiares são determinantes nas condições de aprendizado de estudantes, independente das instituições educacionais onde concluíram o ensino médio. O recorte institucional nesta amostra permitiu discriminar tão somente estudantes de escolas públicas, privadas, rurais e urbanas, revelando que escolas públicas incluem estudantes com desempenhos relativamente piores que escolas privadas e surpreende ao indicar que estudantes de escolas rurais têm desempenhos melhores que de escolas urbanas.

Os dados censitários não permitem relacionar diretamente um estudante à instituição educacional onde estava matriculado ao concluir o ensino médio. A estratégia adotada para solucionar esta lacuna de dados foi considerar um estudo específico para estudantes de municípios onde somente há uma escola, permitindo conceber a escola onde estudou e assim abarcar informações amplas de cunho individual, familiar, escolar e municipais, mesmo que para uma amostra reduzida, mas significativa, de mais de 400 mil estudantes em mais de 2800 municípios. A regressão realizada nesta amostra reforçou o defendido pela literatura e os resultados empíricos da Estimação I, apontando que o desempenho é influenciado por questões familiares e escolares.

Estes resultados sugerem que pensar a melhoria da educação brasileira implica em pensar estratégias de melhoria socioeconômica dos pais, proporcionando melhores condições de emprego e renda. O contraditório é que a precariedade nas condições de trabalho dos mais pobres que implica em fragilidades educacionais, já é resultado de fragilidades educacionais prévias, o que realça o caráter complexo e multidimensional da educação brasileira. Políticas públicas que objetivem a melhoria educacional devem englobar aspectos amplos da realidade

O estudo abrangente dos aspectos multidimensionais em torno do aprendizado dos estudantes brasileiros foi restrito a uma amostra de pequenos municípios. A tentativa de superar um viés, pode ter implicado em outro. Sugere-se para pesquisas futuras que se investigue para uma amostra mais ampla, incluindo uma maior representação dos estudantes, escolas e municípios brasileiros. Todavia, isto demanda esforços institucionais para acesso aos dados, haja vista a restrição garantida pela Lei Geral de Proteção de Dados (LGPD) restringe o acesso a tais informações.

No curto prazo, portanto, é plausível supor que se supere problemas relacionados ao baixo nível de escolaridade dos jovens brasileiros. Porém, para que isto se reflita em uma melhora substancial no acúmulo de capital humano e aprendizado na população, é necessário tempo e mudanças estruturais na sociedade. O entendimento é simples. Políticas educacionais devem ser compreendidas em passos múltiplos retroalimentados. Os jovens recém egressos do ensino médio tendem a conseguir empregos menos precários que àqueles com pior escolaridade. A tendência é que a próxima geração, filha destes jovens com um grau mínimo de escolaridade e acesso a empregos decentes, eleve a probabilidade de aprendizado, podendo galgar maiores sonhos no que cerne à escolarização e emprego.

Em síntese, pode ser notado que a qualidade da educação deve ser mensurada por uma ótica multidimensional, apreendendo não tão somente aspectos institucionais da escola, mas também aspectos individuais e socioeconômicos ligados às famílias dos educandos, como as condições de moradia, ocupação e escolaridade dos pais. É percebido que o principal potencializador do aprendizado é decorrente do acúmulo intergeracional de competências. Ou seja, há uma herança de oportunidades ou de desafios, adquirida por externalidades intergeracionais entre estudantes e seu seio familiar. A escola deve buscar estratégias para, junto com outras frentes de políticas, garantir de fato a inclusão social e minimizar os gargalos históricos expressos na juventude brasileira.

REFERÊNCIAS

Arellano, M. Computing Robust Standard Errors For Within-Groups Estimators. **Oxford Bulletin Of Economics & Statistics**, V. 49, N. 4, 1987.

Balogh, T.; Streeten, P. P. The Coefficient Of Ignorance. **Bulletin Of The Oxford University Institute Of Economics & Statistics**, 25(2), Pp.99-107, 1963.

Baltagi, B. **Econometric Analysis Of Panel Data**. John Wiley & Sons, 2008.

Bierens, H. J. The Logit Model: Estimation, Testing And Interpretation. **Pennstate, Department Of Economics**, 2008.

Breusch, T.; A. Pagan. The Lm Test And Its Applications To Model Specification In Econometrics. **Review Of Economic Studies**, 47, Pp. 239–254, 1980.

Card, D.; Krueger, A. B. School Resources And Student Outcomes: An Overview Of The Literature And New Evidence From North And South Carolina. **Journal Of Economic Perspectives**, V. 10, No. 4, P. 31-50, 1996.

Coleman, J. S. **Equality of Educational Opportunity [Summary Report]**. US Department of Health, Education, and Welfare, Office of Education; 1966.

Cramer, J. S. The Origins And Development Of The Logit Model. **Logit Models From Economics And Other Fields**, V. 2003, P. 1-19, 2003.

Curi, A. Z.; Filho, N. A. M. Mensalidade Escolar, Background Familiar E Os Resultados Do Exame Nacional Do Ensino Médio (Enem). **Pesquisa E Planejamento Econômico (Ipea)**, V. 43, N. 2, Pp. 223-254, 2013.

Dedrick, R. F., Ferron, J. M., Hess, M. R., Hogarty, K. Y., Kromrey, J. D., Lang, T. R., ... & Lee, R. S. (2009). Multilevel modeling: A review of methodological issues and applications. *Review of educational research*, *79*(1), 69-102.

Duflo, E.; Dupas, P.; Kremer, M. Peer effects, teacher incentives, and the impact of tracking: Evidence from a randomized evaluation in Kenya. **American economic review** 101, no. 5: 1739-1774, 2011.

Figueirêdo, E.; Nogueira, L. Leite, F. S. Igualdade De Oportunidades: Analisando O Papel Das Circunstâncias No Desempenho Do Enem. **Revista Brasileira De Economia**, N. 68, Pp. 373-392, 2014.

Fryer, R. G., Steven D. Levitt, J. L; Sadoff, S. **Enhancing The Efficacy Of Teacher Incentives Through Loss Aversion**: A Field Experiment. No. W18237. National Bureau Of Economic Research, 2012.

Fryer, R. G., Steven D. Levitt, J. L; Sadoff, S. **Enhancing The Efficacy of Teacher Incentives Through Loss Aversion**: A Field Experiment. No. w18237. National Bureau of Economic Research, 2012.

Greene, W. H. **Limdep**: Version 8.0: Econometric Modeling Guide. Econometric Software, 2002.

Gujarati, D. N.; Porter, D. C. **Econometria Básica - 5ª Edição**. Porto Alegre: Amgh, 2011.

HANUSHEK, E. A. Assessing the effects of school resources on student performance: An update. **Educational Evaluation and Policy Analysis**, v. 19, no. 2, p. 141-164, 1997.

Hanushek, E. A. The Failure Of Input-Based Schooling Policies. **The Economic Journal**, V. 113, N. 485, P. F64-F98, 2003.

Hanushek, E. A.; Kain, J.; O'brien, D.; Rivkin, S.G. The Market For Teacher Quality. No. W11154. **National Bureau Of Economic Research**, 2005.

Hanushek, E. A.; Woessmann, L. Education And Economic Growth. In **Oxford Research Encyclopedia Of Economics And Finance**. 2021.

Hanushek, E. A.; Woessmann, L. The Role Of Cognitive Skills In Economic Development. **Journal Of Economic Literature**, *46*(3), Pp.607-668, 2008.

Moraes, C. P.; Peres, R. T.; Barbosa, T. S; Pedreira, C. E. Efeito Escola A Partir De Indicadores Educacionais: Análise Entre Escolas Públicas E Privadas No Enem. **Revista Meta: Avaliação**, 14(42), Pp.67-93, 2022.

Pires, A. Renda Familiar E Escolaridade Dos Pais: Reflexões A Partir Dos Microdados Do Enem 2012 Do Estado De São Paulo. **Etd Educação Temática Digital**, V. 17, No. 3, Pp. 523-541, 2015.

Schultz, T. W. Investment in human capital. **The American Economic Review**, 51(1), pp.1-17, 1961.

Shaffer, H. G. Investment in human capital: Comment. **The American Economic Review**, *51*(5), p. 1026-1035, 1961.

Wooldridge, J. M. **Introdução À Econometria**: Uma Abordagem Moderna. São Paulo: Cengage Learning, 2010.

Wooldridge, J. M. **Introductory Econometrics**: A Modern Approach 3rd Ed. 1996.