



Revista de Administração Pública

ISSN: 0034-7612

ISSN: 1982-3134

Fundação Getulio Vargas

Melo, Rafael Oliveira; Freitas, Anne Caroline de; Francisco,
Eduardo de Rezende; Motokane, Marcelo Tadeu
Impacto das variáveis socioeconômicas no desempenho do Enem: uma análise espacial e sociológica
Revista de Administração Pública, vol. 55, núm. 6, 2021, Novembro-Dezembro, pp. 1271-1294
Fundação Getulio Vargas

DOI: <https://doi.org/10.1590/0034-761220200843>

Disponível em: <https://www.redalyc.org/articulo.oa?id=241077337002>

- ▶ Como citar este artigo
- ▶ Número completo
- ▶ Mais informações do artigo
- ▶ Site da revista em redalyc.org



Sistema de Informação Científica Redalyc
Rede de Revistas Científicas da América Latina e do Caribe, Espanha e Portugal
Sem fins lucrativos acadêmica projeto, desenvolvido no âmbito da iniciativa
acesso aberto

Impacto das variáveis socioeconômicas no desempenho do Enem: uma análise espacial e sociológica

Rafael Oliveira Melo ¹

Anne Caroline de Freitas ²

Eduardo de Rezende Francisco ³

Marcelo Tadeu Motokane ⁴

¹ Fundação Getúlio Vargas / Educação Executiva, São Paulo / SP – Brasil

² Universidade de São Paulo / Programa de Pós-graduação Interunidades em Ensino de Ciências
Instituto de Biociências, São Paulo/ SP - Brasil

³ Fundação Getúlio Vargas / Escola de Administração de Empresas de São Paulo, São Paulo / SP – Brasil

⁴ Universidade de São Paulo / Faculdade de Filosofia, Ciências e Letras de Ribeirão Preto, Departamento de Biologia,
Ribeirão Preto / SP – Brasil

O objetivo deste trabalho é identificar variáveis com maior impacto no desempenho dos municípios no Exame Nacional do Ensino Médio (Enem), tanto para a prova objetiva quanto para a redação, com base em técnicas de estatística espacial, que permite analisar a dinâmica da influência territorial, e da perspectiva da sociologia da educação. Pretende-se mensurar o grau de importância de cada categoria de variáveis para a média e a variância das notas do Enem. Para isso, utilizam-se modelos estatísticos e geoespaciais, como regressão linear múltipla e regressão espacial, a partir do Enem de 2018. Para a prova objetiva, variáveis como o percentual de estudantes com bolsa, renda, raça, escolaridade e nível instrucional da mãe são fatores relevantes para o desempenho e a dispersão das notas dos estudantes de cada município. Para a redação, as variáveis são similares às da prova objetiva, mas com menor impacto na média e na dispersão das notas. Esse fator explicativo aumenta quando introduzimos um componente espacial no modelo para as notas de redação, indicando que outros fatores regionais, diferentes dos socioeconômicos, impactam o desempenho e a dispersão dos municípios. Os resultados reforçam os estudos da sociologia da educação, ao destacarem as disparidades socioeconômicas refletidas no desempenho estudantil, fundamentos da reprodução escolar das desigualdades mencionadas por Bourdieu (1998, 2008, 2009). Estudos com essa perspectiva integrada são relevantes para o entendimento da influência no nível da localidade dos municípios, podendo evidenciar lacunas específicas, direcionando e influenciando ações públicas visando à superação das desigualdades.

Palavras-chave: Enem; sociologia; desempenho educacional; análise estatística; regressão espacial.

Impacto de las variables socioeconómicas en el desempeño del Enem: un análisis espacial y sociológico

El objetivo de este trabajo es identificar variables con mayor impacto en el desempeño de los municipios en el Examen Nacional de Enseñanza Media (Enem), tanto para la prueba objetiva como para la prueba de redacción, a partir de técnicas de estadística espacial que permiten analizar la dinámica de la influencia territorial, y de la perspectiva de la sociología educacional. Pretendemos medir el grado de importancia de cada categoría de variables para la media y varianza de la puntuación del Enem. A tal fin, utilizamos modelos estadísticos y análisis geoespaciales, como regresión lineal y espacial, a partir del Enem 2018. Los resultados muestran que, para la prueba objetiva del Enem, variables como el porcentaje de alumnos becarios, renta, raza, escolaridad y nivel de instrucción de las madres son factores relevantes para el desempeño y dispersión de las notas de los estudiantes en cada municipio. Para el desempeño en redacción, las variables son similares a las de la prueba objetiva, no obstante, con menor impacto en la media y dispersión de la puntuación. Cuando se introduce un componente espacial en el modelo de puntuación de redacción, el factor explicativo aumenta, señalando que existen otros criterios regionales, distintos a los socioeconómicos, que impactan sobre el desempeño y dispersión de los municipios. Los resultados refuerzan los estudios de la sociología de la educación, ya que subrayan las disparidades socioeconómicas reflejadas en el

DOI: <http://dx.doi.org/10.1590/0034-761220200843>

Artigo recebido em 08 dez. 2020 e aceito em 30 ago. 2021.

ISSN: 1982-3134



desempeño de los estudiantes, que son fundamentos de la reproducción escolar de las desigualdades analizadas por Bourdieu (1998, 2008, 2009). Las investigaciones con esta perspectiva integrada son relevantes para la comprensión de la influencia en la regionalidad de los municipios, pudiendo revelar brechas específicas y, en consecuencia, dirigir e incidir en acciones públicas orientadas a la superación de desigualdades.

Palabras clave: Enem; sociología; desempeño educativo; análisis estadístico; regresión.

Impact of socioeconomic variables on Enem performance: a spatial and sociologic analysis

This study aims to identify variables with the highest impact on students' performance in the National Exam of High School (Enem), per municipality. The research adopted the educational sociology perspective and used spatial statistics to analyze the dynamics of territorial influence. We analyzed Enem 2018, measuring the degree of importance of each category of variables for the average and variance of students' grades in the exam – which is separated in an essay and an objective test – and used statistical modeling and geospatial analysis, such as linear and spatial regressions. The results indicate that, for the objective test, variables like percentage of students with scholarship, income, race, schooling, and education level of students' mothers are relevant to students' performance and dispersion of grades in each municipality. For the essay, variables were similar to the objective test but with less impact on the average and variance of the grades. This explicative factor increases when a spatial component is introduced in the model for the essay grades, indicating that there are other regional factors, besides socioeconomics, impacting the performance and dispersion per municipality. The results reinforce the sociological studies on education since the socioeconomic disparities reflected in the students' performance stand out, which are also pointed out by Bourdieu (1998, 2008, 2009) in his studies of the fundamentals of school productivity of inequalities. Studies with this integrated perspective are relevant to understand the influence of municipalities' location, evidencing specific gaps and, consequently, directing and influencing public actions to overcome inequalities.

Keywords: Enem; sociology; educational performance; statistical analysis; spatial regression.

1. INTRODUÇÃO

De acordo com a Constituição Brasileira de 1988, “a educação, direito de todos e dever do Estado e da família, será promovida e incentivada com a colaboração da sociedade, visando ao pleno desenvolvimento da pessoa, seu preparo para o exercício da cidadania e sua qualificação para o trabalho” (Constituição da República Federativa do Brasil de 1988). Essa Constituição tem como metas a universalização do ensino e a erradicação do analfabetismo, proporcionando um novo panorama para a educação democrática e cidadã no Brasil (Cury, 2014). Tal perspectiva traz o entendimento de que a educação se propõe como uma das principais ferramentas para a superação de paradigmas socioeconômicos, oferecendo oportunidades para a construção de uma cidadania mais justa.

Para Freire (1967), a educação é um instrumento de libertação e de transformação de conjunturas sociais, sendo uma possibilidade para a reflexão e a problematização de contradições. No entanto, ao analisarmos a educação brasileira sob um viés histórico e social, deparamos com condições desiguais de ensino e de acesso à instituição escolar (Libâneo, 2012).

Tendo em vista a compreensão desses diferentes contextos, fazem-se necessárias práticas avaliativas que possam atuar como ferramentas importantes para o mapeamento e o acompanhamento do desempenho educacional (Menezes & Pazello, 2005). Com as políticas públicas educacionais, um conjunto de variáveis objetivas e fundamentadas para a criação de critérios de qualidade levou à proposição de indicadores de avaliação do desempenho de alunos em testes aplicados em larga escala (Hartman, 1999).

No Brasil, uma das avaliações de ensino em larga escala é o Exame Nacional do Ensino Médio (Enem), realizado anualmente pelo Instituto Nacional de Estudos e Pesquisas Educacionais Anísio

Teixeira (Inep). Essa avaliação tem, entre seus objetivos, analisar o desempenho escolar ao fim do ciclo básico de educação e colaborar para o ingresso no ensino superior, atuando como substituto ao vestibular para alguns cursos, instituições, e também como base para programas de financiamento estudantil (Inep, 2009).

O Enem é apontado na Lei de Diretrizes e Bases da Educação Brasileira (Lei nº 9.394, de 20 de dezembro de 1996), a qual determina, como responsabilidade da União, a implementação de um processo avaliativo de rendimento educacional, que tem como finalidade o mapeamento de prioridades e o aprimoramento da qualidade da educação. O surgimento do Enem se dá num contexto histórico e governamental marcante, no qual ocorre a implementação de testes em larga escala na América Latina, construindo uma nova configuração de papel do Estado e de acumulação de capital (Silva & Carvalho, 2010).

Testes em larga escala tem sua origem em meados do século VII, na China (Travitzki, 2013). Atualmente, existem diversos exames em larga escala em todo o mundo, como SAT e ACT (EUA), A-level (Inglaterra), Baccalauréate (França), PAU (Espanha), Abitur (Alemanha), Gaokao (China), PSU (Chile), HSC (Austrália) e USE (Rússia). De forma geral, um dos testes internacionais de maior relevância é o Programa Internacional de Avaliação de Estudantes (Pisa), realizado pela Organização para a Cooperação e o Desenvolvimento Econômico (OCDE). Esse exame considera o desempenho dos estudantes em competências genéricas de leitura, matemática e ciências. Em 2018, o Brasil apresentou os seguintes resultados: i) Leitura – OCDE 487, Brasil 413; faixa do Brasil no ranking: 55º e 59º; ii) Matemática – OCDE 489, Brasil 384; faixa do Brasil no ranking: 69º e 72º; iii) Ciências – OCDE 489, Brasil 404; faixa do Brasil no ranking: 64º e 67º (Inep, 2018c).

O Enem, por sua vez, avalia competências específicas e mais gerais, sendo comparado com exames internacionais como o Gaokao chinês e o SAT estadunidense (Travitzki, 2013). O exame brasileiro conta com 180 questões objetivas divididas em áreas do conhecimento, incluindo também a avaliação de redação, na qual os alunos devem apresentar um texto dissertativo-argumentativo com base num assunto proposto (Inep, 2009). Esse teste é um indicador importante no país, já que é de abrangência nacional.

É preciso, no entanto, refletir criticamente sobre os fatores que influenciam o desempenho dos estudantes e as condições de acesso das diferentes realidades brasileiras. Isso se justifica em razão das desigualdades que perduram, em âmbito educacional, “entre os estudantes, entre as escolas, entre classes de uma dada escola e entre as regiões em que se localizam as escolas” (Torres, Bichir, Gomes & Carpin, 2006). Ao considerar tal panorama, é evidente a interferência de diversos fatores no desempenho dos estudantes, notadamente refletido em avaliações padronizadas de larga escala (Gremaud, Felício & Biondi, 2007; Lobo, Cassuce & Cirino, 2016).

Nesse contexto, o objetivo deste trabalho é identificar variáveis com maior impacto no desempenho dos municípios no Enem, tanto para a prova objetiva quanto para a redação. Pretende-se mensurar o grau de importância de cada categoria de variáveis para a média e a variância das notas. Para isso, utilizam-se modelos estatísticos e análises geoespaciais. Os dados referentes ao Enem 2018 (Microdados e Censo Escolar) em todo o território brasileiro são objeto deste estudo.

A relevância deste trabalho está no direcionamento de potenciais medidas de superação do cenário atual da educação brasileira baseado na compreensão das variáveis de maior impacto sobre as notas do exame. Isso pode induzir políticas e diretrizes para o contexto educacional de cada município.

2. REFERENCIAL TEÓRICO

É factível haver conjunturas sociais para além das questões organizacionais do âmbito escolar que influenciam no desempenho dos estudantes. Para compreender com mais profundidade a influência do contexto socioeconômico e cultural no processo formativo, serão usadas as contribuições da sociologia da educação. Neste trabalho, vão ser abordadas as contribuições de Pierre Bourdieu (1930-2002), visando ao entendimento mais profundo das questões analisadas.

Para Bourdieu (1998), a instituição escolar acredita em sua neutralidade, pois nela há um conjunto de regras absolutas e globais que garante as mesmas chances de desempenho a todos que nela ingressam. Essa ideia é reforçada por outra concepção da escola: a de que a avaliação é uma ferramenta objetiva, justa e democrática para todos. Nessa óptica tradicional, o fracasso escolar se pauta na inaptidão de certos indivíduos à vida escolar, resultando no mito do dom e da competência (Muzzeti, 1999). Segundo o autor, é preciso romper essa ideia, já que as chances de êxito educacional não são as mesmas. Para ele, alguns têm condições sociais mais propícias para responder às demandas educacionais. Logo, as desigualdades educacionais têm sua raiz no âmbito social, necessitando ultrapassar o ideário de desempenho respaldado nos dons individuais.

Assim, a derivação social tem papel preponderante no âmbito educacional (Bourdieu & Passeron, 2014). Nesse contexto, as contribuições sociológicas bourdianas possibilitam a análise do percurso dos estudantes. Aqueles pertencentes às classes hegemônicas dominantes têm tendência a alcançar maior êxito educacional, enquanto os de classes populares tendem a ser suprimidos do processo educacional (Lima, Ostermann & Rezende, 2013).

Para Bourdieu, o fator determinante do percurso educacional é o capital cultural, correlato à origem social dos diferentes grupos:

[...] a noção de capital cultural impôs-se, primeiramente, como uma hipótese indispensável para dar conta da desigualdade de desempenho escolar de crianças provenientes das diferentes classes sociais, relacionando o sucesso escolar, ou seja, os benefícios específicos que as crianças das diferentes classes e frações de classe podem obter no mercado escolar, à distribuição do capital cultural entre as classes e frações de classe (Bourdieu, 1998, p. 73).

As ideias de Bourdieu (2008) destacam que o sistema educacional se configura sob a exigência de uma “cultura aristocrática”, na qual as práticas educacionais utilizam uma linguagem que favorece a classe dominante, reproduzindo e conservando valores da ordem social. Logo, a escola contribui para a continuidade de uma cultura específica, privilegiando dons e méritos.

Essa visão se estende às ferramentas avaliativas, sobretudo aquelas em larga escala. Segundo Lima (2015), ao aplicarmos uma ferramenta avaliativa única a todos os alunos, advindos de diferentes realidades socioeconômicas, a tendência é que o resultado reflita as marcas sociais, visto que a avaliação incorporará a cultura dominante de seu tempo. As camadas socioeconômicas mais distantes da cultura hegemônica vigente terão desempenhos desfavorecidos nessas avaliações, servindo de base para uma escola conservadora e construindo uma ferramenta reguladora do neoliberalismo (Blini & Backes, 2019).

Segundo Blini e Backes (2019), essa perspectiva construída por avaliações em larga escala faz com que os alunos pobres, ou seja, aqueles com menor desempenho, sejam culpabilizados, secundarizando a responsabilidade do Estado em promover um ensino de qualidade.

Para Perrenoud, essas avaliações de sistema utilizam justificativas do ponto de vista mercadológico, como a demanda vigente de dados comparativos, a fim de neutralizar os impactos da localidade.

[...] as avaliações de sistema, que permitem comparar dados, não possuem tais restrições e podem “levar a sério os objetivos de formação”, o que supõe não somente construir um outro quadro das desigualdades em razão da padronização das provas, como também estimar de modo menos favorável a eficácia do sistema (Perrenoud, 2003, p. 11).

Assim, tais avaliações externas à realidade escolar se restringem em dados facilmente mensuráveis (conteudistas e tecnicistas), secundarizando a abordagem de outras habilidades, como “o raciocínio, a imaginação, a autonomia, a solidariedade, a cidadania, o equilíbrio corporal ou o ouvido musical” (Perrenoud, 2003, p. 12). Dessa forma, torna-se mais rápido e barato se ater a provas objetivas e/ou escritas, ou seja, testes que se restrinjam às questões cognitivas disciplinares e a aptidões técnicas.

Bourdieu destaca que o papel da avaliação escolar extrapola a mera aferição da aprendizagem, inculcando valores culturais e até morais dos estudantes (C. M. M. Nogueira & M. A. Nogueira, 2002). Observam-se na teoria bourdiana diversas contribuições, focalizando uma perspectiva macrossocial e destacando a não neutralidade das instituições escolares em reproduzir tais máculas sociais. No entanto, a literatura destaca limitações no argumento do autor, entre elas os efeitos que a escola pode imprimir em tais desigualdades, ou seja, ele negligência as contribuições do plano microssociológico. De fato, existem divergências no modo como cada escola e/ou professor intervém nesse processo de reprodução, como destacado por alguns autores que delineiam o efeito-escola (Soares, 2004).

Esses estudos, no entanto, precisam ser aprofundados e feitos sob um recorte longitudinal (Alves & Soares, 2007). Apesar dos dados elencados pelo efeito-escola, ainda permanecem evidentes as marcas externas do plano macrossocial, em especial num país onde a desigualdade étnica, econômica e social tem papel imperativo no sistema educacional (Louzano, 2013). Ademais, Arias (2009) destaca que, a despeito das contribuições dos estudos de efeito-escola, 65% a 95% da variação das notas das escolas são fatores contextuais, ou seja, variáveis contundentes que interferem no desempenho escolar e são externas à instituição, afetando significativamente seu desempenho.

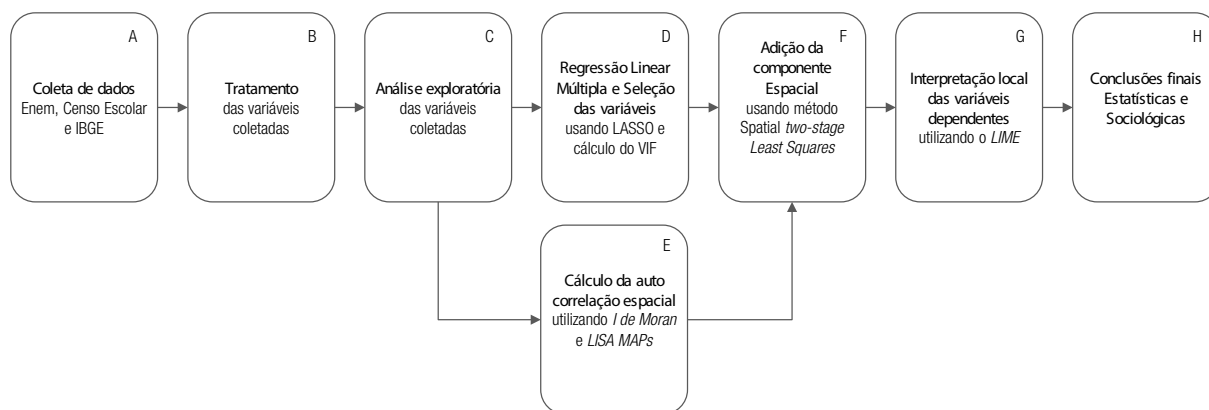
Dado tal panorama sociológico, faz-se necessário identificar e depreender os conceitos sociais que permeiam a sociedade e influenciam o contexto escolar, compreendendo o papel das instituições em reforçar ou transpor esse panorama (Silva, 2017).

3. METODOLOGIA

Inicialmente, procedeu-se à coleta de dados utilizando os dados do Enem, do censo escolar e do Instituto Brasileiro de Geografia e Estatística (IBGE, 2018, 2019; Inep, 2018), como visto na Figura 1, A. Em seguida, foi feito o tratamento das variáveis coletadas (Figura 1, B) para posterior análise exploratória (Figura 1, C). Realizaram-se, então, a regressão linear e a seleção das variáveis (Figura 1, D). Com base na análise exploratória, foi feito o cálculo da autocorrelação espacial (Figura 1, E). Em seguida, realizou-se o modelo de regressão linear com a componente espacial (Figura 1, F).

Na sequência, para uma análise específica por cidade, utilizou-se um modelo de interpretação local das variáveis dependentes mapeadas (Figura 1, G). Por fim, foram discutidos os dados do ponto de vista estatístico e sociológico (Figura 1, H). Tais etapas de análise e discussão de dados são detalhadas nas seções 3.1 a 3.5.

FIGURA 1 ETAPAS METODOLÓGICAS DO ESTUDO (A-H: PERCURSO METODOLÓGICO E FERRAMENTAS ANALÍTICAS UTILIZADAS)



Fonte: Elaborada pelos autores.

3.1. Coleta e Amostra

Foram usados dados de três fontes: microdados do Enem 2018 (Inep, 2018b), microdados do Censo Escolar 2018 (Inep, 2019a) e estatísticas do IBGE (2018) (Figura 1, A).

Nos microdados do Enem, são apresentadas as notas de cada área do conhecimento e da redação por aluno, além das variáveis socioeconômicas oriundas do questionário preenchido pelo candidato no momento de inscrição (Inep, 2018b).

Nos microdados do Censo Escolar, constam informações sobre infraestrutura e disciplinas oferecidas pelas escolas, como disponibilidade de salas, laboratórios, refeitório, equipamentos eletrônicos, turmas de 3º ano com aulas de cada disciplina escolar – por exemplo, português, matemática, espanhol e inglês –, quantidade de profissionais administrativos por escola etc. (Inep, 2018a).

Os dados do IBGE foram utilizados para aprofundar a compreensão dos fatores socioeconômicos do país (IBGE, 2018). São usadas as seguintes variáveis: Índice de Gini, Produto Interno Bruto *per capita* (PIB *per capita*) e proporção da população abaixo da linha da pobreza.

3.2. Seleção e Tratamento das Variáveis

Os dados do Enem foram coletados na granularidade do aluno; os do Censo Escolar, na da escola; e os do IBGE, na do município. A primeira parte do tratamento da base de dados foi o entendimento das variáveis presentes em cada uma das fontes (Figura 1, B).

A etapa seguinte consistiu no tratamento e na transformação dos dados por meio de *feature engineering*, com criação de *dummies* e de novas variáveis derivadas das informações originais das bases de dados descritas (Seção 3.1). A base final resultante apresenta as informações de 79 variáveis coletadas no nível de município. Foram encontradas informações para 5.548 dos 5.570 municípios do Brasil (IBGE, 2019), o que representa 99,6% do total.

As variáveis *dependentes* se baseiam na média e na variância das notas da prova objetiva e da redação. Assim, essas serão: a) nota média da prova objetiva; b) variância da prova objetiva; c) nota média da prova de redação; d) variância da prova de redação (Quadro 1). As variáveis *independentes* – ou seja, que serão utilizadas para prever as notas médias e as variâncias do Enem – foram classificadas em cinco tipos (Quadro 1): econômicas, raciais, de perfil instrucional da mãe, de incentivo escolar e de infraestrutura e ensino escolar. Cada uma dessas variáveis foi representada com diferentes tons de cores nos gráficos deste trabalho (Quadro 1), a fim de facilitar a leitura e a identificação dos mapas.

QUADRO 1 DESCRIÇÃO DAS CATEGORIAS DE VARIÁVEIS E DAS CORES COM QUE CADA UMA FOI REPRESENTADA

Categoria de variável	Tonalidade usada nos gráficos	Exemplo
Média e variância municipal das notas do Enem	Marrom	Média da prova objetiva
Variáveis econômicas	Vermelho	Percentual de estudantes com renda domiciliar alta
Variáveis raciais	Roxo	Percentual de estudantes negros
Variáveis de perfil instrucional da mãe	Azul	Percentual de estudantes com mães que concluíram o ensino fundamental 2
Variáveis de incentivo escolar	Laranja	Percentual de estudantes que fizeram o ensino médio com bolsa
Variáveis de infraestrutura e ensino escolar	Verde	Percentual de escolas com refeitório

Fonte: Elaborado pelos autores.

A média representa o desempenho de cada município no Enem, enquanto a variância, por ser uma medida de dispersão da média, representa a amplitude de distanciamento das notas dos estudantes com relação à média do município. As variáveis dependentes serão apresentadas em escala de tons marrom.

3.3. Análise Exploratória

Com o intuito de entender o comportamento das variáveis e levantar hipóteses a serem comprovadas por meio de técnicas estatísticas, fez-se a análise exploratória das variáveis dependentes e independentes (Figura 1, C). O método usado para essa análise exploratória foi o desenvolvimento de mapas cloropléticos em nível municipal, com a fórmula de otimização Jenks Natural Breaks (Jenks, 1967), com 6 faixas (nomeadas de 0 a 5), as quais foram diferenciadas nos gráficos por diferentes tonalidades em *dégradé* – respeitando as tonalidades usadas para cada variável (Quadro 1) –, com base na categorização das variáveis contínuas.

Além disso, para a exploração da localidade municipal, foi feita uma análise do município de Itapipoca, localizado na região norte do Ceará. Valemo-nos desse município para caracterizar em quais faixas ele se posiciona numa das variáveis acima. Depois, será feita a interpretação local do modelo de regressão utilizando essa cidade como exemplo.

3.4. Técnicas Estatísticas

O presente estudo fez uso de cinco técnicas estatísticas: regressão linear múltipla, *least absolute shrinkage and selection operator* (Lasso), *variance inflation factor* (VIF), *spatial two-stage least squares* (STLS) e *local interpretable model-agnostic – explanations* (L), conforme Figura 1 (caixas D a G).

O trabalho seguiu com a aplicação do modelo de regressão linear múltipla, tanto para média quanto para a variância das notas da prova objetiva e de redação referentes ao município: nota média da prova objetiva, nota média de redação, variância da nota da prova objetiva e variância da nota de redação (Figura 1, D).

Aplicou-se o modelo de regressão linear múltipla utilizando o método Lasso (Tibshirani, 2017) (Figura 1, D), responsável por selecionar as variáveis independentes que realmente têm alguma influência nas variáveis de média e na variância das provas de redação e objetiva. Para evitar o *overfitting*, usou-se o método K-FOLD (com $k=5$) (Stone, 1974).

Em seguida, fizeram-se o cálculo do VIF (Fox & Monette, 1992), para avaliação e eliminação de multicolinearidade, e o cálculo para as variáveis selecionadas pelo Lasso, eliminando aquelas que apresentavam VIF maior que 10. O Quadro 3 apresenta o R^2 ajustado dos modelos de regressão finais de todas as variáveis dependentes, ao passo que o Gráfico 1 (itens A a C) apresenta a importância de cada uma das variáveis independentes.

Depois, o *I* de Moran (1948) foi usado para uma análise descritiva da componente espacial das variáveis dependentes (Silva, Rebouças, Abreu & Ribeiro, 2018) (Figura 1, E). Para o caso em questão, a matriz de vizinhança escolhida foi a de contiguidade *Queen* de ordem 1. Em seguida, valeu-se da técnica estatística *Lisa Maps* (*local indicator of spatial association*), que apresenta os indicadores locais de autocorrelação espacial e permite observar a formação de *clusters* espaciais onde existe uma relação entre as variáveis de municípios vizinhos (Silva, Rebouças, Abreu & Ribeiro, 2018) (Figura 1, E).

Com as variáveis selecionadas pelos métodos descritos, foi implementado o modelo de regressão espacial, por meio do método STLS (Anselin, 2012), com a mesma matriz de vizinhança do cálculo do *I* de Moran (Figura 1, F).

Por fim, usou-se o método Lime (Ribeiro, Singh & Guestrin, 2016) (Figura 1, G) para entender as variáveis que impactam positiva e negativamente na média e na variância de cada um dos municípios. Como exemplo, temos o município de Itapipoca-CE.

3.5. Ferramentas Utilizadas

A etapa de manipulação de base de dados (Figura 1, B) utilizou o *software* R 4.0.2 (R Development Core Team, 2020). Com o auxílio dessa ferramenta, foram feitos os cruzamentos entre as três bases de dados de origem: Enem, Censo Escolar e IBGE (Seleção e Tratamento das Variáveis – tópico 3.2).

A análise exploratória e os modelos regressão linear, espacial e de explicação local (Figura 1, C a 1, G) foram realizados no *software* Python (Python Software Foundation, 2020), outro de livre

acesso. Além do pacote Scikit Learn, para a regressão linear, foram usados *pysal*, *geopandas* e *esda*, do Python, para as análises de estatística espacial, além do *matplotlib*, para a construção dos gráficos.

4. RESULTADOS

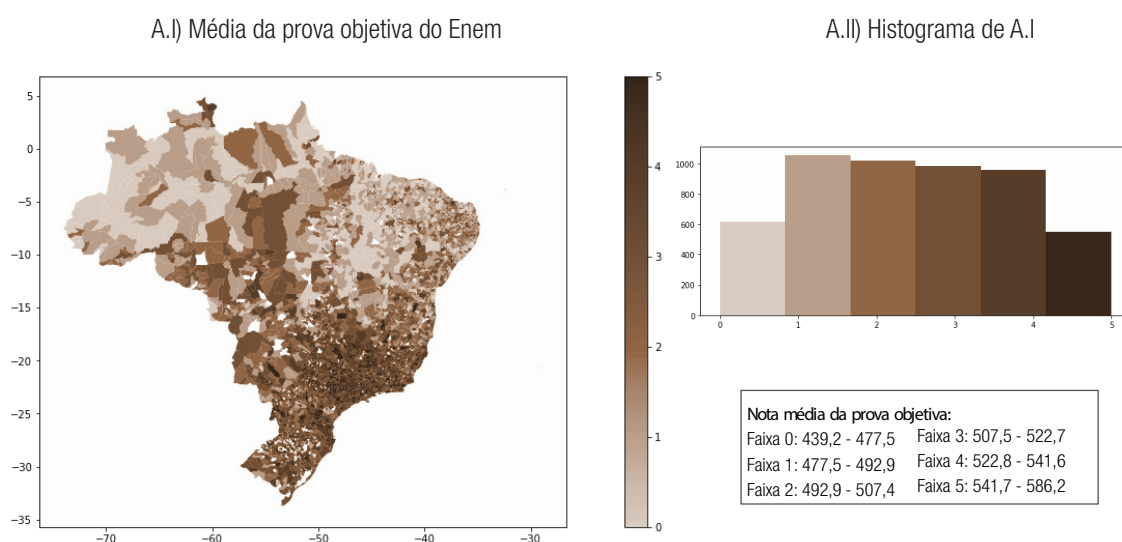
4.1. Análise Exploratória

As Figuras 2 e 3 mostram os mapas coropléticos desenhados com o método Jenks Natural Breaks (Jenks, 1967) (Seção 3.3). As figuras também apresentam histogramas das quantidades de municípios por cada uma das faixas definidas pelo algoritmo.

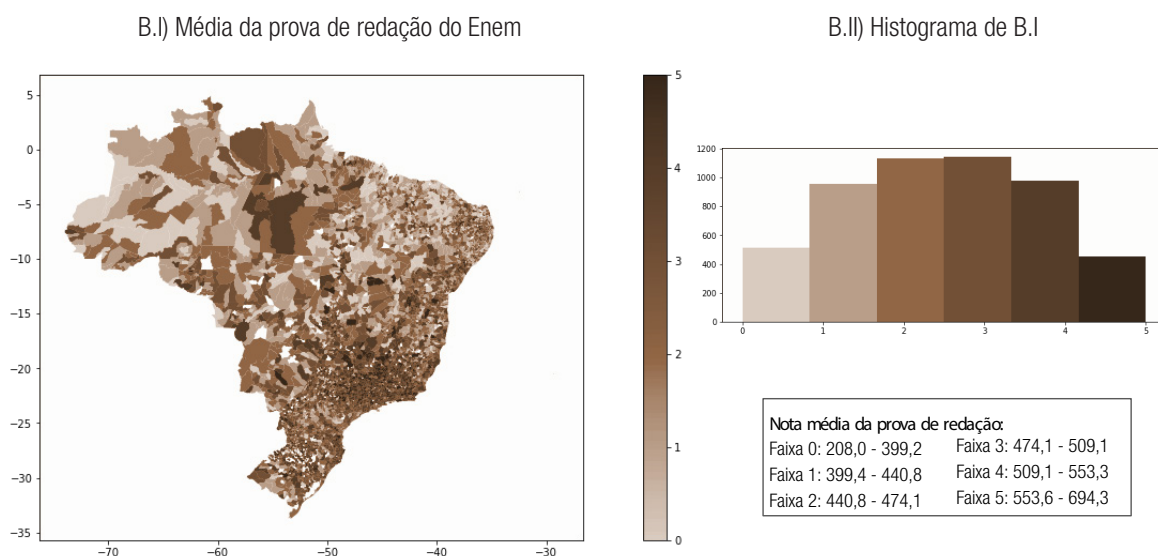
A Figura 2 representa duas das quatro variáveis dependentes analisadas neste estudo: nota média da prova objetiva e nota média da prova de redação. A Figura 3 apresenta três das 79 variáveis independentes utilizadas neste estudo: percentual de renda alta, representando as variáveis econômicas; percentual de mães com alta especialização, representando variáveis de perfil instrucional das mães; e percentual de escolas com refeitório, representando as variáveis de infraestrutura e ensino escolar.

Para variáveis dependentes, há uma grande diferença regional para a média da prova objetiva (Figura 2 A.I e A.II). Na média da prova de redação (Figura B.I e B.II), observa-se a mesma regionalidade, mas a concentração de municípios das faixas 4 e 5 nas regiões Sul e Sudeste não é tão grande quanto a média da prova objetiva.

FIGURA 2 MAPAS (À ESQUERDA) E HISTOGRAMAS DAS DISTRIBUIÇÕES (À DIREITA) DAS VARIÁVEIS DEPENDENTES (FAIXA 0, DE MENOR INTENSIDADE, ATÉ FAIXA 5, DE MAIOR INTENSIDADE)



Continua



Nota: Mapas construídos segundo critério de classificação Jenks Natural Breaks (Jenks, 1967).

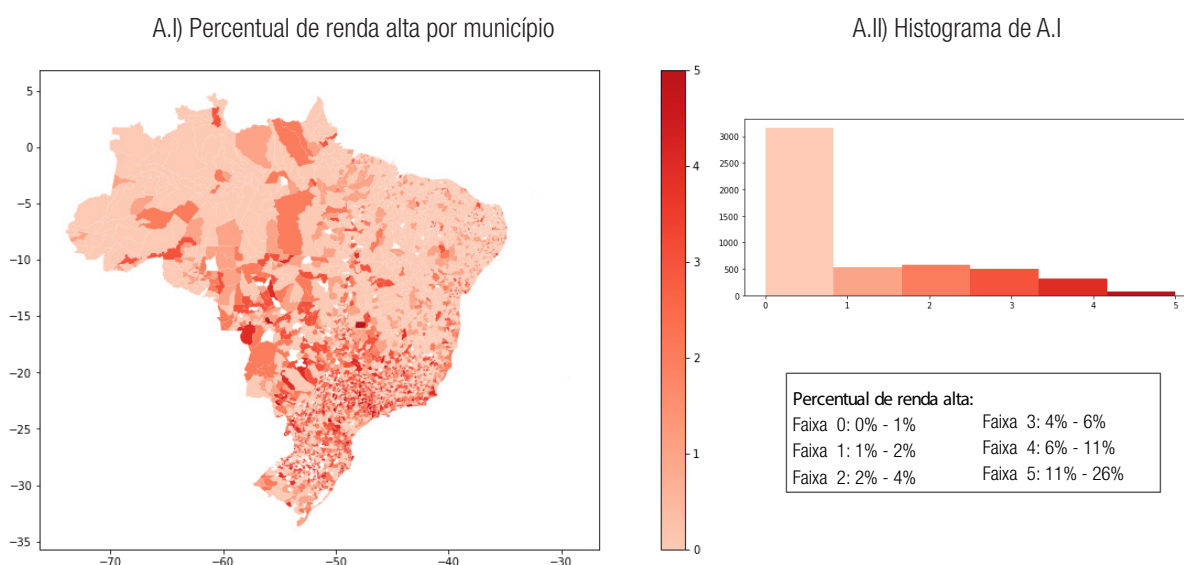
Fonte: Elaborada pelos autores.

Para as variáveis independentes, observa-se, na Figura 3, que o percentual de renda alta dos municípios tem concentração nas regiões Sul, Sudeste e parte sul do Centro-Oeste (Figura 3, A.I e A.II), indicando uma possível correlação com as médias das provas objetivas e de redação, a ser comprovada no cálculo da regressão na seção 4.2.

A variável “mães com alta especialização” (Figura 3, B.I e B.II) faz um contraponto com os dados até então descritos, destacando-se também na região Norte, com alta quantidade de municípios nas faixas 4 e 5.

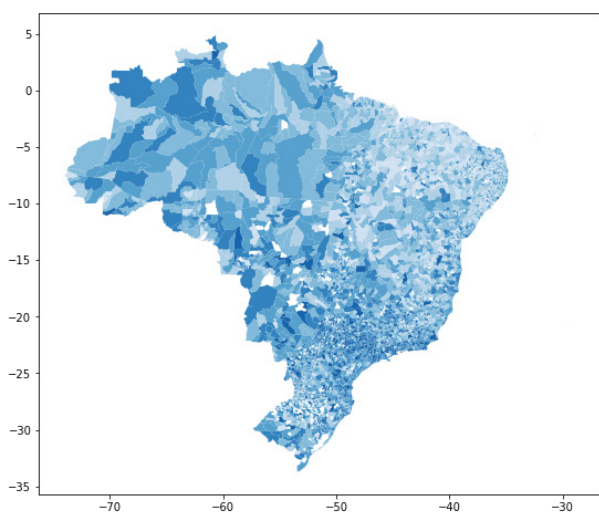
Para a variável “percentual de escolas com refeitório” (Figura 3, C.I e C.II), constata-se que as regiões Norte e Sudeste se destacam, apresentando grande parte dos municípios com alto índice de refeitórios em suas escolas.

FIGURA 3 MAPAS (À ESQUERDA) E HISTOGRAMAS DAS DISTRIBUIÇÕES (À DIREITA) DAS VARIÁVEIS INDEPENDENTES (FAIXA 0, DE MENOR INTENSIDADE, ATÉ FAIXA 5, DE MAIOR INTENSIDADE)

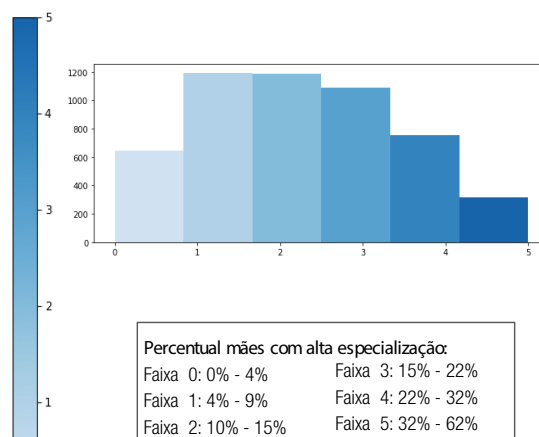


Continua

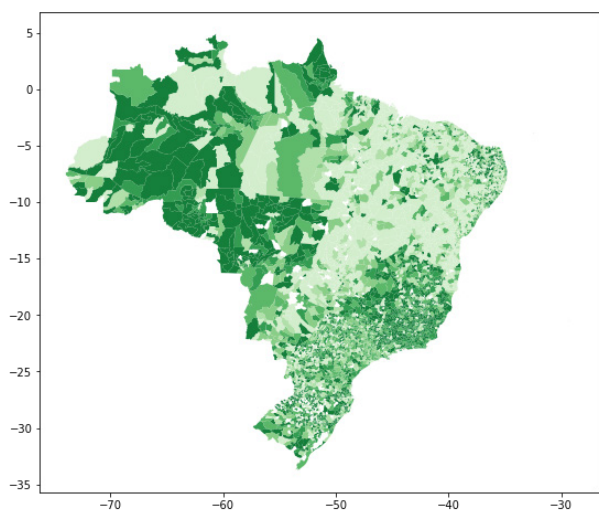
B.I) Percentual de mães com alta especialização por município



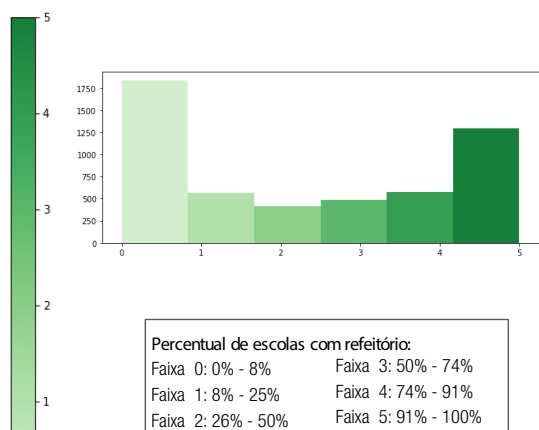
B.II) Histograma de B.I



C.I) Percentual de escolas com refeitório por município



C.II) Histograma de C.I



Nota: Mapas construídos segundo critério de classificação Jenks Natural Breaks (Jenks, 1967).

Fonte: Elaborada pelos autores.

Buscando destacar um exemplo local e entender como essas variáveis se comportam em nível de município, destacamos a cidade de Itapipoca-CE. O Quadro 2 apresenta como esse município se posiciona em cada uma das variáveis descritas nas Figuras 2 e 3. Destaca-se que a cidade se enquadra na Faixa 1 para a “média da prova objetiva e redação”, ou seja, entre as notas médias mais baixas do país.

Para as variáveis independentes, a cidade se situa na faixa 0 para “percentual de famílias de renda alta”; e na faixa 1, para “percentual de mães com alta especialização” e “percentual de escolas com refeitório”.

QUADRO 2 VALORES E FAIXAS DAS VARIÁVEIS APRESENTADAS NOS MAPAS DAS FIGURAS 2 E 3 PARA A CIDADE DE ITAPIPOCA-CE

Variável	Valor	Faixa
Nota média da prova objetiva	481,9	1
Nota média da prova de redação	417,8	1
% de estudantes com família de renda alta	0,4%	0
% de estudantes com mães com trabalho de alta especialização	8,5%	1
% de escolas com refeitório	24,3%	1

Fonte: Elaborado pelos autores.

4.2. Regressão Linear Múltipla e Métodos de Seleção de Variáveis

O Quadro 3 apresenta o R^2 ajustado, obtido em cada um dos quatro modelos de regressão linear múltipla – um para cada uma das variáveis dependentes –, após a utilização dos métodos de seleção de variáveis explicadas acima.

QUADRO 3 R^2 DA REGRESSÃO LINEAR MÚLTIPLA APÓS APLICAÇÃO DO LASSO E EXCLUSÃO DE VARIÁVEIS MULTICOLINEARES VIA VIF

Modelo	R^2 ajustado
Média nota da prova objetiva	0,68
Variância nota da prova objetiva	0,37
Média nota redação	0,37
Variância nota redação	0,10

Fonte: Elaborado pelos autores.

Observa-se que a média da prova objetiva é altamente explicada pelas variáveis independentes selecionadas (68%), enquanto sua variância é moderadamente explicada (37%). Em outras palavras, as variáveis socioeconômicas selecionadas pelo modelo impactam bastante o desempenho dos municípios, representado pela média, e um pouco a dispersão das notas dos alunos por município, representada pela variância.

Para a prova de redação, a média é em parte explicada pelas variáveis selecionadas (37%), enquanto a variância é pouco explicada (10%). Dessa forma, as variáveis econômicas selecionadas têm um relativo impacto no desempenho dos municípios, representado pela média, e pouco impacto na dispersão das notas dos estudantes por município, representado pela variância. Assim, existe o indicativo de que

outros fatores também influenciam na variância da nota da redação. Logo, nenhuma análise foi feita sobre a importância das variáveis independentes para a dispersão das notas de redação, em razão do baixo R^2 encontrado em seu modelo (Quadro 3).

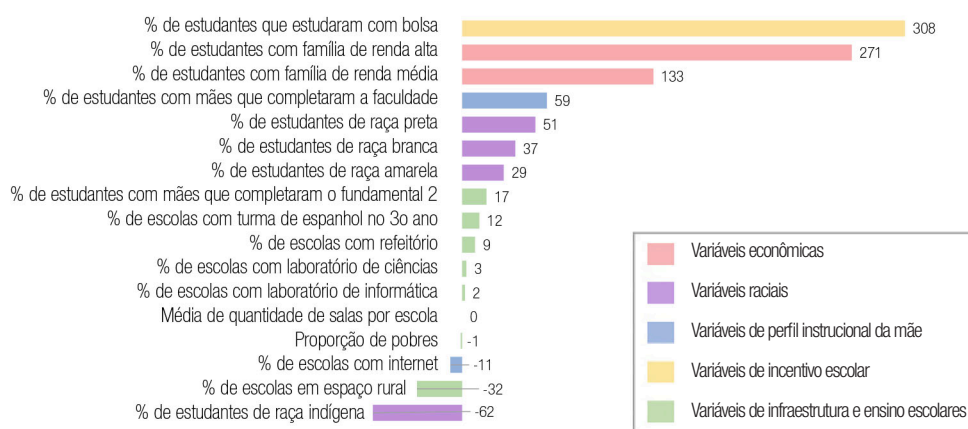
Com base nesses resultados, faz-se a avaliação de como cada uma das variáveis independentes influenciam no desempenho e na dispersão das notas dos municípios. O Gráfico 1 (A a C) representa a importância de cada *feature* para cada uma das variáveis dependentes analisadas neste trabalho, exceto a da prova de redação, graças ao baixo R^2 .

GRÁFICO 1 IMPORTÂNCIA DAS VARIÁVEIS SELECIONADAS PARA MÉDIA E VARIÂNCIA DA NOTA OBJETIVA E MÉDIA DA PROVA DE REDAÇÃO



Continua

Média da prova de redação



Fonte: Elaborado pelos autores.

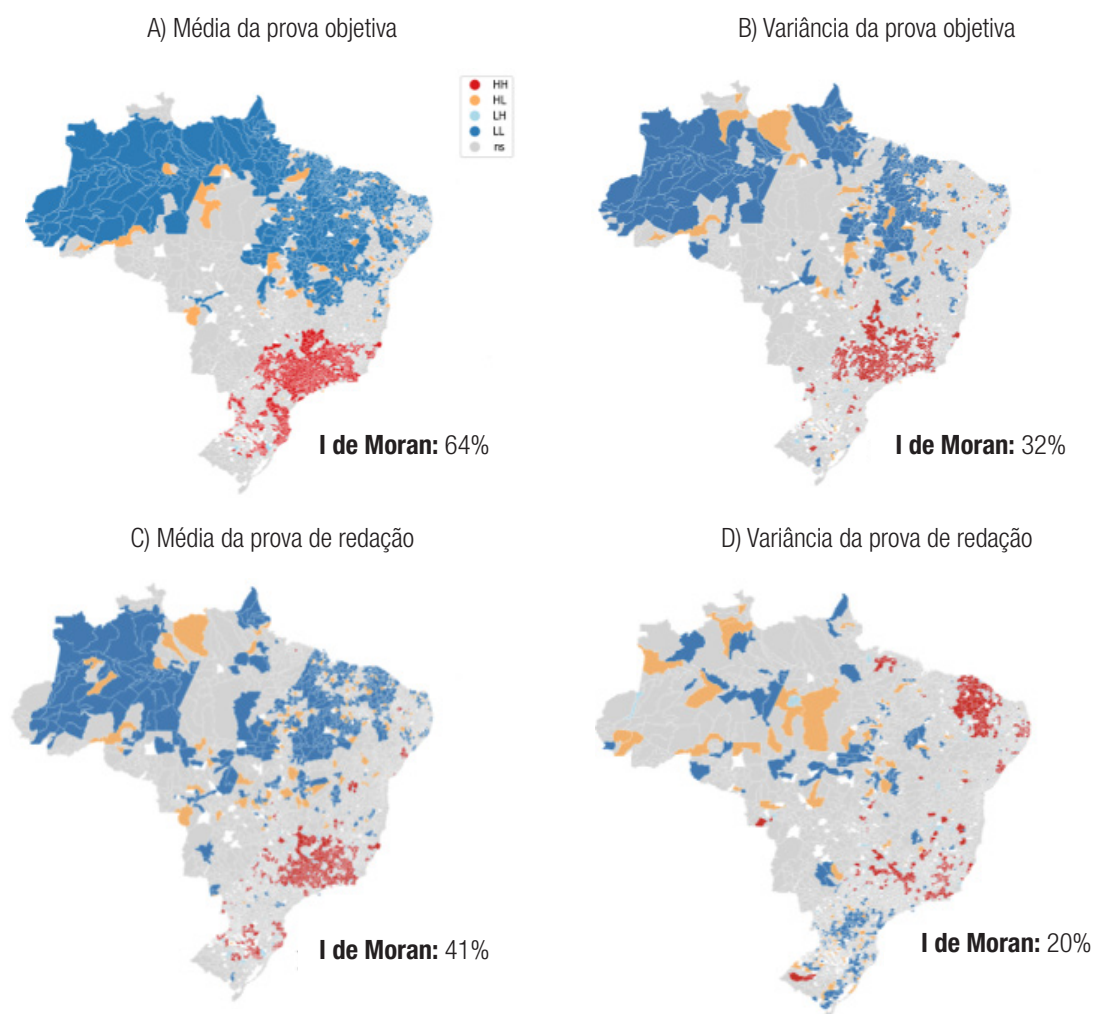
Para a prova objetiva, o “percentual de estudantes com bolsa em escola privada” é uma variável relevante para a média do desempenho e da dispersão das notas do município (Gráfico 1, A e B). Essa porcentagem influencia positivamente a nota dos estudantes, ao passo que aumenta a dispersão. Nota-se também que algumas variáveis socioeconômicas têm alto impacto para a prova objetiva, seguindo a respectiva ordem de importância: variáveis econômicas – municípios com maior percentual de famílias com renda alta têm maior desempenho e dispersão, enquanto aqueles com maior proporção de pobres têm menor desempenho e menor dispersão –; perfil instrucional da mãe – maiores índices de escolaridade e especialização das mães do município traduzem em maior desempenho e dispersão –; iii) variáveis raciais – municípios com maior percentual de estudantes de raça branca têm melhor desempenho, e aqueles com maior percentual de raça negra têm maior dispersão.

Para a prova de redação, o “percentual de estudantes com bolsa em escola privada” é também uma variável muito relevante para o desempenho do município no Enem (Gráfico 1, C). Além disso, algumas variáveis socioeconômicas apresentam alto impacto para a prova de redação, seguindo a respectiva ordem de importância: variáveis econômicas – municípios com maior percentual de famílias com renda alta têm maior desempenho –; variáveis raciais – municípios com maior percentual de estudantes de raça indígena têm pior desempenho –; e perfil instrucional da mãe – maiores índices de escolaridade e especialização das mães do município se traduzem em maior desempenho.

Tanto para a prova objetiva quanto para a redação, outras variáveis interferem na nota, porém em menor grau. Elas são referentes à *infraestrutura e ao ensino das escolas* do município, como percentual de escolas com refeitório e laboratório de informática – quanto maior o percentual, maiores desempenho e dispersão – e percentual de escolas com turmas de espanhol no terceiro ano do ensino médio – quanto maior o percentual, maiores desempenho e dispersão.

4.3. Análise de Vizinhança e I de Moran

Por meio da metodologia *Queen* de ordem 1, foram avaliadas as correlações de vizinhança para a média e a variância, tanto para a nota da prova objetiva quanto para a de redação. A Figura 4 (A-D) apresenta os Lisa Maps das variáveis e o I de Moran de cada uma delas.

FIGURA 4 LISA MAPS DE CADA UMA DAS VARIÁVEIS DEPENDENTES

Nota: A (mesma) legenda para todos os mapas apresenta a seguinte descrição: HH – Alto-Alto; HL – Alto-Baixo; LH – Baixo-Alto; LL – Baixo-Baixo; ns – não significante.

Fonte: Elaborada pelos autores.

Os Lisa Maps (Figura 4, A a D) demonstram os locais onde há concentração de notas/variâncias altas (HH – cor vermelha), de notas/variâncias baixas (LL – cor azul) e transição de regiões de concentração de notas/variâncias alta e baixas (HL e LH – cor laranja e cor azul claro, respectivamente). Os Lisa Maps da Figura 4 (A e C) indicam que há evidente concentração de maiores notas nas regiões Sul e Sudeste (HH – cor vermelha) e das menores notas nas regiões Norte e Nordeste (LL – cor azul), tanto para a nota objetiva quanto para a de redação. A Figura 4 (B e D) aponta que a dispersão das notas dos alunos para a prova objetiva é maior no Sudeste (HH – cor vermelha). Já para redação, a dispersão das notas é maior no Sudeste e no Nordeste (HH – cor vermelha).

As mesmas figuras indicam que há autocorrelação espacial para as médias e também para as variâncias, sendo muito alta para a média (64% de explicação espacial) e alta para a variância (32% de explicação espacial) da nota objetiva. Para a redação, há uma autocorrelação espacial alta para

a média da nota de redação (41% de explicação espacial) e moderada para variância desta (20% de explicação espacial).

4.4. Regressão Espacial versus Regressão Linear

A etapa seguinte consiste em testar os modelos de regressão espacial por meio do método STLS descrito na seção 3.3.1. Para isso, foram utilizadas as mesmas variáveis selecionadas pelo método de regressão linear múltipla. O Quadro 4 apresenta um comparativo entre o R^2 ajustado da regressão linear múltipla e o R^2 ajustado da regressão espacial.

QUADRO 4 R^2 DA REGRESSÃO LINEAR MÚLTIPLA *VERSUS* REGRESSÃO ESPACIAL

Modelo	Regressão linear múltipla	Regressão espacial	Incremento absoluto da componente espacial	Incremento relativo da componente espacial
Média nota objetiva	0,68	0,70	0,02	3,0%
Variância nota objetiva	0,37	0,39	0,02	5,2%
Média nota redação	0,37	0,42	0,05	13,8%
Variância nota redação	0,10	0,19	0,09	92,6%

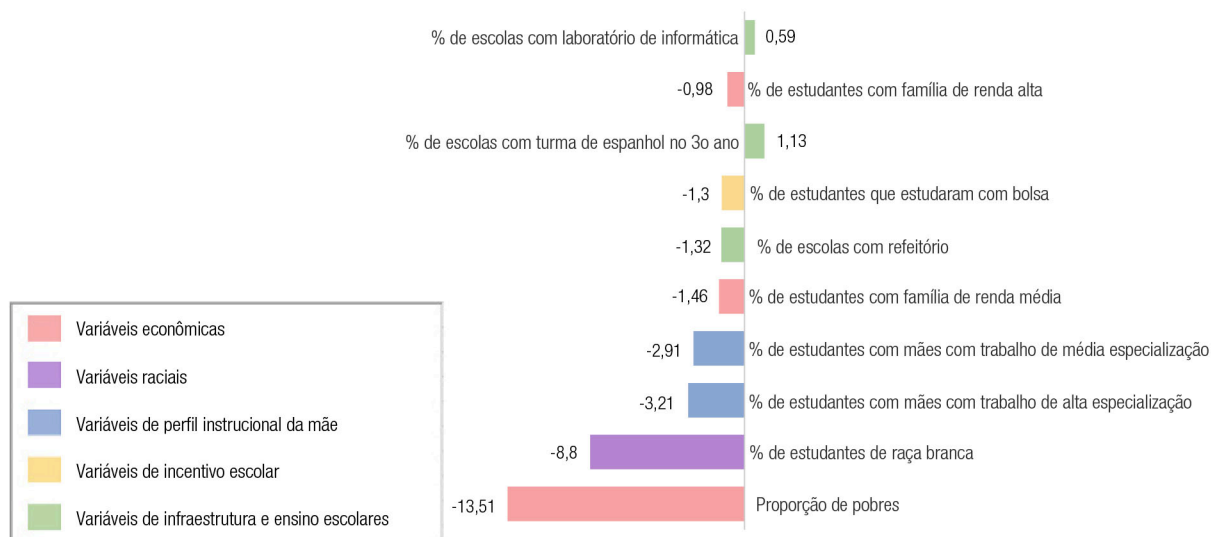
Fonte: Elaborado pelos autores.

Pode-se inferir que a adição da componente espacial tem um incremento marginal na correlação para a média e a variância da nota da prova objetiva. Tal resultado indica que as variáveis definidas no modelo linear representam forte relação espacial. Para a média e a variância da nota de redação, a adição da componente espacial tem um incremento relativamente maior (13,8% para a média e 92,6% para a variância). Assim, alguns fatores regionais, diferentes dos socioeconômicos, influenciam no desempenho e na dispersão das notas dos municípios.

4.5. Explicação Local

Por fim, usa-se o método Lime para obter um entendimento mais diretivo da análise das variáveis que apresentam maior poder explicativo por município.

GRÁFICO 2 IMPACTO DE CADA VARIÁVEL DO MODELO PARA A CIDADE DE ITAPIPOCA-CE NO DESEMPENHO DA PROVA OBJETIVA DO ENEM, SEGUNDO O MÉTODO LIME



Fonte: Elaborado pelos autores.

No Gráfico 2, tem-se a saída gerada via Lime quanto à nota média da prova para a cidade de Itapipoca-CE, que foi de 481.9, em comparação com 507.7 da média municipal brasileira. Percebe-se, como sugerido pela análise exploratória (Seção 4.1), que em Itapipoca os fatores *econômicos*, *raciais* e de *perfil instrucional das mães* dos estudantes fazem com que a nota fique abaixo da média nacional. Além disso, o baixo *incentivo de bolsas de estudo* e a *infraestrutura das escolas* – percentual de escolas com refeitório e com internet – são fatores que reduzem a média.

5. DISCUSSÃO

Os dados apresentados corroboram estudos que apresentam as questões socioculturais e econômicas como o cerne das desigualdades de desempenho dos estudantes (Peis, 2011). Dessa forma, a depender da maior condição socioeconômica, maior será o desempenho nas avaliações escolares (Travitski, 2013).

Observamos, pelos dados obtidos, que questões socioeconômicas, como rendimento familiar e formação das mães (Gráfico 1, A-C), têm forte relação com o desempenho dos estudantes, refletindo a dependência entre origem socioeconômica e resultados educacionais (Bonamino, Alves & Franco, 2010). Esses aspectos vão de encontro às perspectivas de Bourdieu, para quem há grupos sociais que, graças às condições de existência (*capital*, *habitus* e *campo*), apresentam uma trajetória marcada por pressões materiais e urgência temporal, distanciando-se do êxito educacional (1998). As famílias com maior capital econômico e cultural tendem a investir nas experiências escolares dos filhos, consequentemente reproduzindo a ordem estrutural de classes e a manutenção do poder (Lima, 2015; Peis, 2011).

Um dado controverso é o fato de a porcentagem de alunos com bolsas em escolas particulares aumentar a média do município. Para Soares, Batista, Alves e Teixeira (2002), a heterogeneidade social poderia propiciar o aprimoramento do desempenho estudantil de camadas sociais mais pobres. Logo, por essa perspectiva, seria interessante para o município investir em bolsas escolares. No entanto, é preciso refletir cautelosamente para essas propostas, visto que podem ser fadadas a programas

compensatórios e, por consequência, reforçar a teoria do déficit social (Ratz, 2019). Além disso, essa porcentagem de alunos com bolsa pode acirrar as desigualdades do desempenho escolar, refletidas na dispersão das notas apontado pelas análises estatísticas aqui realizadas.

Outra variável relevante para o desempenho no Enem foi a raça, fator de forte influência no capital cultural e econômico do grupo, já que é sabido que o Brasil é marcado por um contexto histórico de exclusão racial, no qual sujeitos pardos e negros têm rendimento financeiro inferior à cultura hegemônica branca. Explica-se, então, o desempenho superior dos estudantes brancos em comparação com os de outros grupos raciais (Heringer, 2002). Para os estudos científicos da área, o processo educacional, por muitas vezes, reforça a desigualdade de desempenho entre brancos e pardos/negros, crescendo em favor dos primeiros (Basso et al., 2012), assim como também apontam nossos dados.

A variável geográfica/espacial é outro fator importante, sendo a região Sul/Sudeste aquela que tem melhor desempenho. Nossos dados corroboram os resultados encontrados por Santos (2018), para quem o desempenho reflete as desigualdades socioeconômicas das diferentes regiões do país. Assim, o Enem se torna uma ferramenta controversa, já que, ao passo que amplia a competitividade, também incorpora a ela as diferenças socioeconômicas de cada região, sendo um problema evidente num país de notada desigualdade regional (Mello, Medeiros, Paiva & Simões, 2014).

Essa variável espacial é um componente relevante ao analisar as notas dos estudantes em redação. Os resultados indicam que questões regionais, aparentemente para além das variáveis socioeconômicas, influenciam o desempenho em redação (Quadro 4). Essa influência se torna evidente no baixo desempenho de populações indígenas e rurais (Gráfico 1, C). Esses dados podem estar relacionados com as influências culturais próprias das diferentes regiões do país, em particular na linguagem.

É sabido que há uma dinâmica relacional entre o aprimoramento linguístico e o meio sociocultural: o uso das diferentes funções da língua e dos diferentes níveis de fala, assim como a transferência da língua oral para a escrita – do sistema fonológico para o ortográfico. Assim, é preciso considerar que os fatores socioculturais estão intimamente ligados à oralidade, a qual influencia muito na escrita (Batista-Santos & Pereira, 2017). Segundo os mesmos autores, o repertório linguístico se refere à escrita, implicando uma maior ou menor incidência de inadequações e marcas da oralidade, ou seja, a origem cultural. Consequentemente, o acesso ao capital linguístico é um fator determinante para o desempenho na redação.

Para Bourdieu, a prática da escrita é fundamento do capital cultural dos alunos, a qual “permite a acumulação da cultura até então conservada no estado incorporado e, correlativamente, a acumulação primitiva do capital cultural” (Bourdieu, 2009, p. 209). Logo, aqueles que se encontram mais distantes da linguagem hegemônica têm dificuldade para apreender a linguagem escolar, oral ou escrita. Assim, uma avaliação na qual a linguagem hegemônica seja o componente principal tende a acirrar as desigualdades.

O que observamos é que a influência da regionalidade no desempenho em redação reflete a teoria da violência escolar simbólica, desenvolvida por Bourdieu (1992), pois a avaliação educacional privilegia a cultura dominante – norma padrão da língua, parte do capital linguístico de regiões privilegiadas socioeconomicamente, como Sul e Sudeste.

Esses dados refletem asserções de outros trabalhos que evidenciam o conceito de fracasso escolar com base no contexto brasileiro de desigualdades econômicas e sociais, como apontado por Louzano. Segundo a autora, apesar do progresso recente, a pobreza e a desigualdade continuam sendo um dos problemas centrais do Brasil: “As oportunidades educacionais não são igualmente distribuídas; o fracasso escolar afeta de maneira diferente alunos de distintos grupos econômicos, sociais e étnico-raciais” (Louzano, 2013, p. 112).

Nesse contexto, é sabido que as questões socioeconômicas no Brasil têm origem histórica e cumulativa, assim como questões raciais, étnicas e culturais. Elas são superadas pelas lutas sociais e civis constantes. Ainda se faz necessário um olhar profundo para a superação dessas marcas no contexto escolar. Uma das formas de romper o estigma de escola como reprodutora das desigualdades é o empoderamento dos estudantes de classes desfavorecidas com o “conhecimento poderoso” (Young, 2007), criando oportunidades didáticas para que os estudantes se apropriem de conhecimentos considerados refinados para a atuação em sociedade. Seria a conquista do conhecimento chamado por Bourdieu (1998) de praxiológico, o qual se torna a fonte para a conscientização do sujeito/agente social na transformação de sua sociedade.

Logo, é preciso que os profissionais da educação, incluindo diretrizes e políticas da área, superem a visão determinista e finalista das condições de diferentes classes, assim como refletir sobre “quais condições de estrutura social e de transmissão dos saberes na educação escolar atuam na superação/manutenção das desigualdades de acesso ao conhecimento científico pelos alunos da classe trabalhadora” (Ratz, 2019, p. 17).

Tendo em vista os dados obtidos, faz-se necessário refletir acerca das possibilidades de utilização do Enem como ferramenta fomentadora de políticas públicas, tanto para o ensino básico quanto para o superior. Para isso, é preciso reconhecer o contexto de desigualdade brasileiro, oitavo país mais desigual do mundo, de acordo com o índice de Gini (Costa & Gartner, 2017; Indexmundi, 2017).

A economia inclusiva, destacada nos estudos de mensuração da desigualdade de Cowell (2011), aponta como possibilidade a redução ou a eliminação da perspectiva excludente na sociedade, causada por diferenças da renda ou algum outro fator. Logo, nossos resultados enfatizam a necessidade de aplicações de políticas públicas de ações reparatórias focalizadas (Mello et al., 2014) que permitam a democratização de acesso e a permanência nas instituições de ensino. Identificadas as necessidades de implementação, é preciso também um olhar mais específico para o âmbito escolar, desde secretarias e diretorias de ensino até ações pedagógicas do professor (Louzano, 2019).

Além disso, este estudo aponta que existem fatores determinantes na nota, como subsídios infraestruturais das instituições escolares. Como apresentado nos modelos deste estudo, aspectos físicos da instituição escolar, como refeitório e laboratório, são profícuos para o êxito educacional. Esses dados corroboram estudos sobre a influência e a importância desses insumos escolares no desempenho escolar (Araújo, 2019). Assim, os indicadores infraestruturais abrem espaço para refletir sobre administração e políticas públicas, já que o Enem é uma avaliação nacional, demandando uma análise de seus dados para o desenvolvimento da educação, como apontado pela LDB (Lei nº 9.394, de 20 de dezembro de 1996).

O presente estudo é relevante para construir perspectivas de políticas públicas em nível municipal, sendo factível mapear demandas que possam influenciar na diminuição das desigualdades e associá-las aos progressos educacionais. Em investigação feita em municípios de Minas Gerais, Caetano, Ávila e Tavares (2017) não encontraram uma relação empírica entre gastos com educação e qualidade do ensino, o que poderia ser um indicativo de recursos mal geridos.

O fato de não ter sido encontrada uma relação empírica entre os gastos com educação e a qualidade do ensino não exige a importância do financiamento dos serviços educacionais municipais. Os recursos podem estar sendo mal geridos devido a uma possível ineficiência do setor público. Nesse caso, os administradores municipais devem estar mais atentos à gestão dos fundos disponíveis e preocuparem-se com o retorno e a eficiência dos serviços ofertados.

Este estudo corrobora essa óptica, indicando a importância de atentar nas ações que podem ser tomadas para a melhoria do desempenho educacional em nível municipal, visando a uma maior eficiência dos gastos públicos.

6. CONCLUSÃO

O objetivo deste trabalho foi identificar as variáveis com maior impacto no desempenho dos municípios no Enem, tanto para a prova objetiva quanto para a redação. Os dados referentes ao Enem 2018 (microdados e Censo Escolar) em todo o território brasileiro foram objeto deste estudo. Buscou-se avaliar diferenças regionais nos resultados e explorar as variáveis – do contexto social dos estudantes e da infraestrutura das escolas do município – com maior impacto sobre o desempenho escolar no exame. Para isso, foram usados modelos estatísticos e análises geoespaciais: microdados do Enem de 2018 (Inep, 2018b); microdados do Censo Escolar de 2018 (Inep, 2018a) e estatísticas do IBGE (2018). Foram aplicadas algumas técnicas estatísticas, como mapas temáticos, Lasso, *I* de Moran, STLS e LIME.

Os resultados obtidos apontam para um conjunto de variáveis com significativo poder explicativo para a prova objetiva. Há forte indicação de que o nível de escolaridade e profissionalização da mãe, a raça do estudante e a renda média da família são variáveis relevantes para o desempenho do município e a dispersão das notas na prova objetiva, bem como o percentual de estudantes de escolas particulares com bolsa. Com menor impacto, a proporção de escolas com refeitório e laboratório de ciências, assim como a quantidade de turmas do terceiro ano do ensino médio com espanhol, também impacta o desempenho e a dispersão da prova objetiva.

Para a redação, as variáveis que mais explicam o desempenho são as mesmas da prova objetiva, porém em menor proporção. Com a adição da componente espacial, o percentual de explicação do modelo cresce, tornando-se claro que existem outros fatores regionais, além dos socioeconômicos. Podemos notar que, para ambas as provas, as variáveis relacionadas com o contexto macrosocioeconômico (econômicas, raciais e perfil instrucional da mãe) têm maior poder explicativo do que as variáveis referentes à infraestrutura e ao ensino do contexto escolar.

Os resultados corroboram os estudos da sociologia da educação, já que refletem a desigualdade educacional quanto ao desempenho estudantil. Eles estão atrelados aos conceitos de capital cultural de Bourdieu (1998), pois o rendimento dos estudantes depende de sua origem social.

Estudos que integram o viés da sociologia da educação com análises estatística com dependência espacial são importantes para a compreensão mais ampla das desigualdades de desempenho escolar e possibilitam uma reflexão crítica e profunda para a promoção de diretrizes, políticas e ações educacionais que rompam a violência simbólica. Dessa forma, passaríamos de reprodução à transformação (Pies, 2011). Assim, este estudo contribuiu para a construção de pesquisas sobre reflexões das desigualdades educacionais, principalmente reproduzidas pelo Enem.

Este artigo também se mostrou relevante para fins gerenciais, visto que, uma vez compreendidas as variáveis de maior impacto sobre essas notas em nível municipal, se torna possível entender para qual aspecto devem ser direcionadas as políticas de redução das desigualdades e relacioná-las com os avanços no desenvolvimento educacional.

REFERÊNCIAS

- Alves, M. T. G., & Soares, J. F. (2007). Efeito-escola e estratificação escolar: o impacto da composição de turmas por nível de habilidade dos alunos. *Educação em Revista*, 45, 25-59.
- Anselin, L. (2013). *Spatial econometrics: methods and models*. Santa Barbara, California: Springer Science & Business Media.
- Araújo, D. L. (2019). *Determinantes do desempenho no Enem dos concluintes do ensino médio no município de Viçosa-MG* (Dissertação de Mestrado). Universidade Federal de Viçosa, MG.
- Arias, D. R. M. (2009). Usos, aplicaciones y problemas de los modelos de valor añadido en educación. *Revista de Educación*, 348, 217-252.
- Basso, A. C. M., Rodrigues, C. F., Brooke, D. A. L., Vignoli, D. A., Candian, J. F., ... Rezende, W. S. (2012). Desigualdade de desempenho e raça: uma análise a partir do Paebs 2009. *Estudos em Avaliação Educacional*, 23(51), 40-56.
- Batista-Santos, D. O., & Pereira, E. B. (2017). A influência da oralidade na escrita: reflexões e desafios no ensino de língua materna. *Desafios – Revista Interdisciplinar da Universidade Federal do Tocantins*, 4(2), 167-184.
- Blini, B. A. Q., & Backes, J. L. (2019). As diferenças culturais no contexto da avaliação em larga escala e de desigualdade social. *Atos de Pesquisa em Educação*, 14(2), 589-617.
- Bonamino, A., Alves, F., Franco, C., & Cazelli, S. (2010). Os efeitos das diferentes formas de capital no desempenho escolar: um estudo à luz de Bourdieu e de Coleman. *Revista Brasileira de Educação*, 15(45), 487-499.
- Bourdieu, P. (1992). *A reprodução*. Rio de Janeiro, RJ: Francisco Alves.
- Bourdieu, P. (1998). *Escritos de educação*. Petrópolis, RJ: Vozes.
- Bourdieu, P. (2008). *Razões práticas: sobre a teoria da ação*. Campinas, SP: Papirus.
- Bourdieu, P. (2009). *O senso prático*. Petrópolis, RJ: Vozes.
- Bourdieu, P., Passeron, J. C., & Silva, C. P. G. (1975). *A reprodução: elementos para uma teoria do sistema de ensino*. Petrópolis, RJ: Vozes.
- Caetano, C. C. R., Ávila, L. A. C. D., & Tavares, M. (2017). A relação entre as transferências governamentais, a arrecadação tributária própria e o índice de educação dos municípios do estado de Minas Gerais. *Revista de Administração Pública*, 51(5), 897-916.
- Constituição da República Federativa do Brasil de 1988*. (1988). Brasília, DF. Recuperado de http://www.planalto.gov.br/ccivil_03/constituicao/constituicao.htm
- Costa, G. P. C. L. D., & Gartner, I. R. (2017). O efeito da função orçamentária alocativa na redução da desigualdade de renda no Brasil: uma análise dos gastos em educação e saúde no período de 1995 a 2012. *Revista de Administração Pública*, 51(2), 264-293.
- Cowell, F. A. (2011). *Measuring inequality* (3a ed.). Oxford, UK: Oxford University Press.
- Cury, C. R. J. (2013). Sistema Nacional de Educação: uma reflexão provocativa ao debate. *Sistema Nacional de Educação*, 2, 34.
- Fox, J., & Monette, G. (1992). Generalized collinearity diagnostics. *Journal of the American Statistical Association*, 87(417), 178-183.
- Freire, P. (1994). *Pedagogia do oprimido* (23a ed.). São Paulo, SP: Paz e Terra.
- Gremaud, A. P., Felício, F., & Biondi, R. L. (2007). *Indicador de efeito escola: uma metodologia para a identificação dos sucessos escolares a partir dos dados da Prova Brasil*. Brasília, DF: Inep.
- Hartman, J. E. (1999). Accountability, testing, and politics. *Profession*, 125-136. Recuperado de <http://www.jstor.org/stable/25595678>
- Heringer, R. (2002). Desigualdades raciais no Brasil: síntese de indicadores e desafios no campo das políticas públicas. *Cadernos de Saúde pública*, 18, S57-S65.
- Index Mundi. (2017). *Gini Index (World Bank Estimate): country ranking*. Recuperado de www.indexmundi.com/facts/indicators/SI.POV.GINI/rankings

- Instituto Brasileiro de Geografia e Estatística. (2018). *Estatísticas*. Recuperado de www.ibge.gov.br/estatisticas/downloads-estatisticas.html
- Instituto Brasileiro de Geografia e Estatística. (2019). *Estrutura territorial do território brasileiro*. Recuperado de https://geofp.ibge.gov.br/organizacao_do_territorio/estrutura_territorial/divisao_territorial/
- Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira. (2009). *Apresentação do Exame Nacional do Ensino Médio (Enem)*. Recuperado de www.gov.br/inep/pt-br/areas-de-atuacao/avaliacao-e-exames-educacionais/enem
- Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira. (2018a). *Microdados do Censo Escolar*. Recuperado de <http://inep.gov.br/web/guest/microdados>.
- Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira. (2018b). *Microdados do Enem*. Recuperado de <http://inep.gov.br/web/guest/microdados>
- Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira. (2018c). *Pisa 2018 revela baixo desempenho escolar em leitura, matemática e ciências no Brasil*. Recuperado de http://portal.inep.gov.br/artigo/-/asset_publisher/B4AQV9zFY7Bv/content/pisa-2018-revela-baixo-desempenho-escolar-em-leitura-matematica-e-ciencias-no-brasil/21206
- Jenks, G. F. (1967). The data model concept in statistical mapping. *International Yearbook of Cartography*, 7, 186-190.
- Kerstenetzky, C. L. (2006). Políticas sociais: focalização ou universalização? *Brazilian Journal of Political Economy*, 26(4), 564-574.
- Lei nº 9.394, de 20 de dezembro de 1996. (1996). Estabelece as diretrizes e bases da educação nacional. Brasília, DF. Recuperado de http://www.planalto.gov.br/ccivil_03/leis/19394.htm
- Libâneo, J. C. (2012). O dualismo perverso da escola pública brasileira: escola do conhecimento para os ricos, escola do acolhimento social para os pobres. *Educação e Pesquisa*, 38(1), 13-28.
- Lima, P. J. (2015). Crítica sociológica do Exame Nacional do Ensino Médio: uma análise bourdiana. *Encontro Nacional de Pesquisa em Educação em Ciências*, 10, 1-8.
- Lima, P. J., Ostermann, F., & Rezende, F. (2013). Análise dos condicionantes sociais do sucesso acadêmico em cursos de graduação em física à luz da sociologia de Bourdieu. *Ensaio – Pesquisa em Educação em Ciências*, 15(1), 113-129.
- Lobo, G. D., Cassuce, F. C. C., & Cirino, J. F. (2017). Avaliação do desempenho escolar dos estudantes da região nordeste que realizaram o Enem: uma análise com modelos hierárquicos. *Revista Espacios*, 38(5), 12.
- Louzano, P. (2013). Fracasso escolar: evolução das oportunidades educacionais de estudantes de diferentes grupos raciais. *Cadernos Cenpec*, 3(1), 111-133.
- Louzano, P., Freitas, P. F., Santos, A. F., Ribeiro, V. M., & Gusmão, J. B. (2019). Implementação de políticas educacionais: elementos para o debate e contribuições para o campo. *Cadernos Cenpec*, 8(2).
- Mello, R. D. N., Medeiros, H. A. V., Paiva, F. S., & Simões, J. L. (2014). O impacto do Enem nas políticas de democratização do acesso ao ensino superior brasileiro. *Comunicações*, 21(3), 109-123.
- Menezes, N. F., & Pazello, E. (2007). Do teachers' wages matter for proficiency? Evidence from a funding reform in Brazil. *Economics of Education Review*, 26(6), 660-672.
- Moran, P. A. P. (1948). The interpretation of statistical maps. *Journal of the Royal Statistical Society – Series B (Methodological)*, 10(2), 243-251.
- Muzzeti, L. R. (1999). Consenso ou conflito: contribuições das teorias sociológicas de Émile Durkheim e de Pierre Bourdieu. *Boletim do Departamento de Didática*, 15, 43-62.
- Nogueira, C. M. M., & Nogueira, M. A. (2002). A sociologia da educação de Pierre Bourdieu: limites e contribuições. *Educação & Sociedade*, 23(78), 15-35.
- Perrenoud, P. (2003). Sucesso na escola: só o currículo, nada mais que o currículo! *Cadernos de Pesquisa*, 119, 9-27.
- Pies, N. G. (2011). *Capital cultural e educação em Bourdieu* (Tese de Doutorado). Universidade de Passo Fundo, Passo Fundo, RS.
- Python Software Foundation. (2020). *Python Language Reference: version 3.7.4*. Recuperado de www.python.org

R Development Core Team. (2020). *R: a language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing. Recuperado de <http://www.R-project.org/>

Ratz, S. V. S. (2019). *Influência do nível de exigência conceitual da prática pedagógica de uma professora no desempenho de alunos socialmente diferenciados em uma sequência didática investigativa sobre ecologia de morcegos* (Tese de Doutorado). Universidade de São Paulo, São Paulo, SP.

Ribeiro, M. T., Singh, S., & Guestrin, C. (2016). “Why should I trust you?” Explaining the predictions of any classifier. In *Proceedings of the 22nd ACM SIGKDD international conference on knowledge discovery and data mining*, San Francisco California.

Silva, F. G. M. (2017). *Desempenho dos estudantes no Enade: influenciadores ambientais e organizacionais sob a luz da sociologia da educação de Bourdieu* (Dissertação de Mestrado). Universidade Salvador, Salvador, BA.

Silva, J. F. B. A., Rebouças, S. M. D. P., Abreu, M. C. S. D., & Ribeiro, M. D. C. R. (2018). Construção de um índice de desenvolvimento sustentável e análise espacial das desigualdades nos municípios cearenses. *Revista de Administração Pública*, 52(1), 149-168.

Soares, J. F. (2004). O efeito da escola no desempenho cognitivo de seus alunos. *Reice – Revista*

Iberoamericana sobre Calidad, Eficacia y Cambio en Educación, 2(2), 83-104.

Soares, J. F., Batista, J. R., Alves, M. T. G., & Teixeira, M. C. A. (2002). *Fatores explicativos do desempenho em língua portuguesa e matemática: a evidência do Saeb-99* (Relatório Técnico). Belo Horizonte, MG: Game/LME/Proav.

Stone, M. (1974). Cross-validatory choice and assessment of statistical predictions. *Journal of the Royal Statistical Society – Series B (Methodological)*, 36(2), 111-133.

Tibshirani, R. (1996). Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society – Series B (Methodological)*, 58(1), 267-288.

Torres, H. G., Bichir, R. M., Gomes, S., & Carpim, T. R. P. (2008). Educação na periferia de São Paulo: ou como pensar as desigualdades educacionais? In L. C. Q. Ribeiro, & R. Kaztman (Orgs.), *A Cidade contra a Escola? Segregação urbana e desigualdades educacionais em grandes cidades da América Latina* (pp. 59-90). Rio de Janeiro, RJ: Letra Capital.

Travitzki, R. (2013). *Enem: limites e possibilidades do Exame Nacional do Ensino Médio enquanto indicador de qualidade escolar* (Tese de Doutorado). Universidade de São Paulo, São Paulo, SP.

Young, M. F. D. (2007). *Bringing knowledge back in: from social constructivism to social realism in the sociology of education*. London, UK: Routledge.

Rafael Oliveira Melo



<https://orcid.org/0000-0002-4570-1375>

Pós-Graduado em Business Analytics & Big Data na Fundação Getulio Vargas (FGV Educação Executiva).
E-mail: rafaelmelo353@gmail.com

Anne Caroline de Freitas



<https://orcid.org/0000-0002-0898-074X>

Doutoranda em ensino de ciências no Instituto de Biociências da Universidade de São Paulo (USP).
E-mail: annefreitas23@gmail.com

Eduardo de Rezende Francisco



<https://orcid.org/0000-0001-8895-2089>

Doutor em Administração de Empresas pela Fundação Getulio Vargas (FGV EAESP); Vice-Coordenador do Curso de Graduação em Administração da Fundação Getulio Vargas (FGV EAESP).
E-mail: eduardo.francisco@fgv.br

Marcelo Tadeu Motokane



<https://orcid.org/0000-0002-8597-6832>

Doutor em Educação pela Universidade de São Paulo (USP); Professor do departamento de Biologia da Faculdade de Filosofia Ciências e Letras de Ribeirão Preto da Universidade de São Paulo (USP).
E-mail: mtmotokane@ffclrp.usp.br