IV SEMANA DA DEMOGRAFIA

DEMOGRAFIA DOS INVISIBILIZADOS: FORTALECENDO PRESENÇAS QUE ESCAPAM

UNICAMP – 06 a 10 de abril de 2026

FATORES ASSOCIADOS AO DESEMPENHO DOS ESTUDANTES NO ENSINO MÉDIO NO MARANHÃO: UMA ANÁLISE COM DADOS DO SAEB

Sarah Pestana Aroucha[1]
Luiza Helena Pinheiro Everton[2]
Pamella Damasceno Dantas de Souza[3]

RESUMO

A presente análise busca apontar os resultados do aprendizado de estudantes do terceiro ano do Ensino Médio no estado do Maranhão através dos dados do Saeb 2023. Por meio de modelos de regressão linear múltipla, a investigação apresenta como as variáveis individuais e institucionais, como raça/cor, sexo, tipo de rede de ensino e atraso escolar, impactam as notas em Língua Portuguesa e Matemática. Os resultados revelam que alunos de escolas privadas, composto por estudantes brancos e aqueles sem distorção idade-série, apresentam, em média, desempenhos superiores. Observou-se também uma disparidade de gênero, com mulheres se destacando na disciplina de Língua Portuguesa, enquanto homens obtiveram melhores resultados na disciplina de Matemática. Compreende-se que o rendimento dos estudantes de ensino médio no Maranhão está sujeito a reprodução de desigualdades sociais no ambiente escolar.

Palavras-chave: Desempenho; Educação; Maranhão; Saeb.

INTRODUÇÃO

A educação é fundamental para a construção de um país economicamente e socialmente mais desenvolvido, dada à sua relevância na formação de indivíduos capazes de interpretar a realidade, como na tomada de decisões, atuando de maneira ativa e benéfica na sociedade. Conforme Dewey (1979), a educação, de forma técnica e sucinta, pode ser definida como uma reconstrução ou reorganização da experiência que amplia e enriquece o seu significado, ao mesmo tempo em que fortalece a capacidade de orientar e regular experiências futuras.

[1] Instituto de Economia – IE/UNICAMP. E-mail: sarah.pestarou@gmail.com
[2] Instituto Maranhense de Estudos Socioeconômicos e Cartográficos – IMESC. E-mail: luizahelenapinheiroe@gmail.com
[3] Instituto de Geociências – IG/UNICAMP. E-mail: pamelladdsouza@gmail.com

No Brasil, apesar da educação ter apresentado uma série de avanços nos últimos anos, os indicadores de qualidade ainda mostram desafios, especialmente para o Ensino Médio. A ênfase na qualidade escolar é crucial, pois as melhorias no crescimento de longo prazo estão intimamente ligadas ao nível de habilidades cognitivas da população, e não meramente ao tempo de escolaridade, desviando a atenção para a importância da qualidade do ensino (Hanushek, 2013).

Nesse sentido, uma educação de qualidade não se limita à transmissão de conteúdos formais, mas envolve o êxito na assimilação desses conteúdos por meio de experiências significativas e transformadoras, capazes de estimular o pensamento reflexivo, a investigação e a participação democrática. Entretanto, compreende-se que o processo de aprendizagem não ocorre de forma homogênea, sendo este influenciado por características inerentes ao indivíduo, bem como pelo contexto socioeconômico e pela estrutura escolar na qual o estudante está inserido.

Dessa forma, ressalta-se a relevância dos estudos que abordam os fatores associados ao desempenho escolar, uma vez que diferenças nos resultados educacionais podem acarretar a constante reprodução de desigualdades e padrões sociais. No ensino médio, etapa da educação básica mais associada ao ingresso no mercado de trabalho, tais fatores assumem especial relevância, sobretudo em contextos marcados por desigualdades socioeconômicas, como o estado do Maranhão.

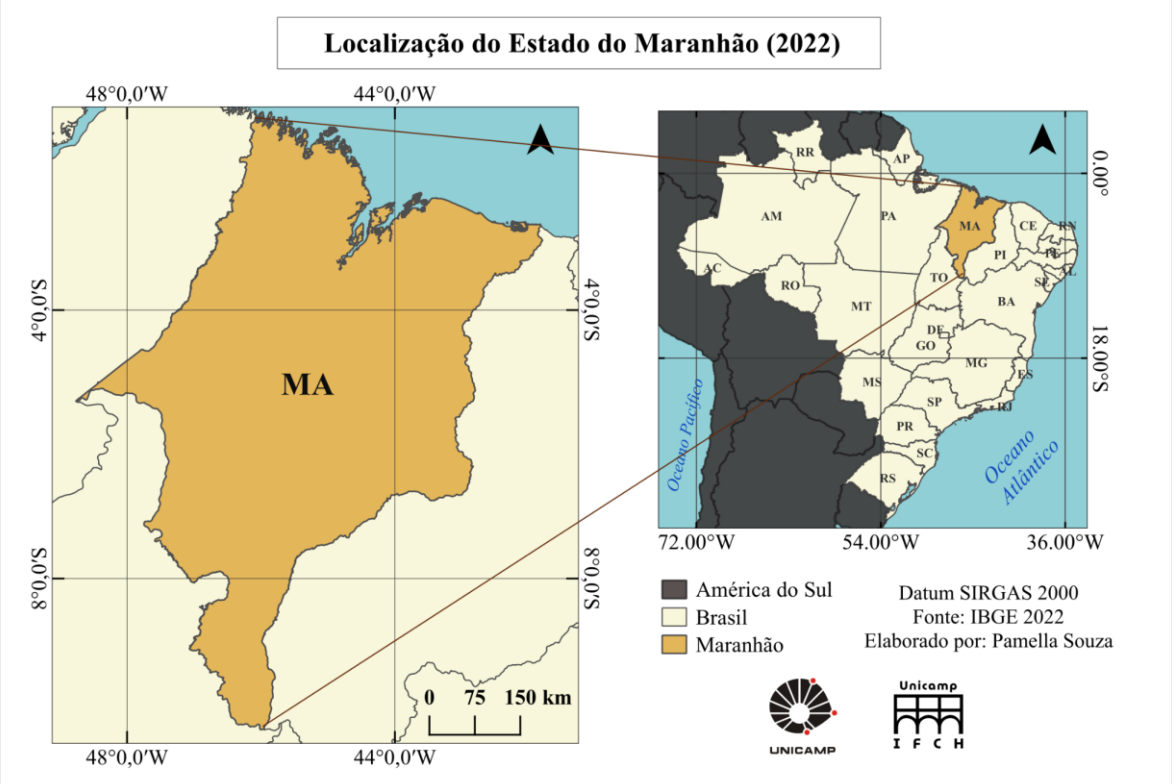
Para os resultados mais recentes do Sistema de Avaliação da Educação Básica (Saeb), referentes à edição de 2023 divulgados pelo Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira (Inep), verifica-se que o Maranhão apresenta desempenho inferior à média nacional no Ensino Médio da rede pública. Enquanto o Brasil registra nota média de 4,45, o Maranhão alcança 4,03, situando-se entre os estados com menores resultados no país e à frente apenas da Bahia.

Diante desse cenário, a identificação dos fatores associados ao desempenho dos estudantes constitui um passo fundamental para compreender as disparidades educacionais

existentes e subsidiar a formulação de políticas públicas voltadas à melhoria da qualidade do ensino. Nesse contexto, os dados do Saeb oferecem uma base empírica robusta para a análise do desempenho escolar e de seus determinantes.

O presente estudo tem como objetivo analisar os fatores associados ao desempenho dos estudantes do ensino médio em Língua Portuguesa e Matemática no estado do Maranhão (Imagem 1), a partir dos microdados do Saeb de 2023. Busca-se assim investigar de que forma variáveis relacionadas às características dos estudantes e a rede de ensino influenciam o desempenho dos estudantes no estado.

IMAGEM 1 – Representação do estado do Maranhão



Fonte: Elaboração por Pamella Souza, com base de dados IBGE (2022).

O trabalho aplicou o modelo de regressão linear múltipla, estimado separadamente para as proficiências em Língua Portuguesa e Matemática, nas quais as variáveis dependentes correspondem às respectivas proficiências nessas áreas e as variáveis explicativas incluem sexo, rede de ensino, raça/cor e distorção idade-série[4]. Além dessa introdução, o trabalho contou com uma seção que aborda os aspectos demográficos do Maranhão, seguida pela terceira seção que apresenta a metodologia utilizada. A quarta seção trouxe os resultados obtidos e a discussão à luz da literatura sobre os fatores associados ao desempenho escolar e à reprodução das desigualdades sociais no sistema de ensino. Finalmente, a quinta seção apresentou a conclusão, destacando as considerações finais.

ASPECTOS DEMOGRÁFICOS

O Maranhão possui uma extensão territorial de 329.651,50 km², distribuída entre 217 municípios, com uma densidade demográfica de 20,56 hab./km². Dentre suas três regiões metropolitanas, a maior é a Região Metropolitana da Grande São Luís, composta por 13 municípios. No contexto demográfico, os dados do Censo Demográfico de 2022 indicam que o Brasil alcançou uma população de 203.080.756 habitantes, o que representa um crescimento de 6,5% em relação ao Censo de 2010. No mesmo período, a população maranhense também registrou expansão, passando de 6.574.789 habitantes, em 2010, para 6.776.699 habitantes, em 2022, correspondendo a um acréscimo de 3,1%, com taxa média de crescimento anual de 0,2% (IBGE, 2022; IMESC, 2024).

Em relação ao sexo, do total da população residente no estado em 2022, 50,9% eram mulheres, o que corresponde a 3.447.276 habitantes, enquanto os homens representavam 49,1% (3.329.423 habitantes). No que se refere à cor/raça, a maior parte da população maranhense se autodeclara parda (66,4%), seguida por pessoas que se declararam brancas (20,1%) e pretas (12,6%); a população indígena representa cerca de 0,8% do total (IBGE, 2022; IMESC, 2024).

[4] A distorção idade-série é a diferença entre a idade adequada para a série do estudante e a idade real do estudante (Portella; Bussmann; Oliveira, 2017).

Segundo o relatório do Instituto Maranhense de Estudos Socioeconômicos e Cartográficos (IMESC, 2024), o aumento da população potencialmente ativa que atingiu +4,3 p.p. em 2022, indicando que 68,5% encontra-se na faixa etária economicamente ativa, entre 15 a 64 anos. No entanto, o estado apresenta proporcionalmente o 4º menor percentual dessa população entre as unidades federativas, apesar de ter registrado um aumento em relação a 2010, o que elevou uma posição do Maranhão no ranking nacional. Ainda assim, a baixa proporção da PIA, associada ao processo de envelhecimento, pode comprometer e dificultar, no longo prazo, a elevação da renda e a redução dos níveis de pobreza no estado.

Outro tópico de maior atenção para a seguinte análise, relaciona-se aos níveis de alfabetização no Maranhão. Segundo a Secretaria de Educação do estado, "em 2010, a taxa de analfabetismo de pessoas de 15 anos ou mais chegou a 20,8%", reduzindo para 15,5% em 2022, sendo esta, a menor taxa nos últimos 12 anos (Maranhão, 2024). O Censo 2022 demonstra que o estado possui 84,95% da sua população alfabetizada (4.358.810 pessoas), ocupando o 24º lugar do ranking nacional, seguido de 772.309 habitantes que não possuem letramento e escrita em seu cotidiano (IBGE, 2022).

MÉTODOS

Essa seção apresenta a metodologia utilizada no trabalho. Todas as análises a seguir foram realizadas no software R. O código foi elaborado desde a etapa de preparação até a estimação, de forma a garantir transparência e replicabilidade dos resultados.

Esse trabalho baseou-se, empiricamente, nos microdados do Sistema de Avaliação da Educação Básica (Saeb), disponibilizado pelo Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira (Inep). O Saeb é um dos principais instrumentos de diagnóstico da educação básica. O sistema coleta dados sobre o desempenho dos estudantes e o contexto das escolas, por meio de dois tipos de instrumentos: testes cognitivos e questionários (Brasil, 2025).

Nesta pesquisa, a amostra foi composta por estudantes do Ensino Médio do Maranhão, abrangendo apenas aqueles que possuíam informações completas nas variáveis selecionadas para o estudo. Justifica-se o uso do Saeb em razão do alto grau de detalhamento, permitindo a avaliação do desempenho educacional dos alunos e a análise de fatores associados à proficiência em Língua Portuguesa e Matemática.

Foram utilizados os arquivos referentes ao questionário do aluno e ao desempenho do Saeb, filtrados para os estudantes da 3ª série do Ensino Médio do Maranhão. O tratamento dos dados iniciou-se com a etapa de limpeza e padronização das informações, na qual caracteres especiais ("*" e ".") foram convertidos em valores ausentes (NA). Em seguida, realizou-se a exclusão das observações com informações incompletas nas variáveis selecionadas para a análise. A base inicial continha 80.581 observações, das quais 16.105 foram removidas devido à presença de valores ausentes, correspondendo a aproximadamente 19,9% do total. Assim, a amostra final foi composta por 64.476 observações válidas, garantindo maior consistência para a estimação do modelo.

Para o objetivo deste estudo, as variáveis dependentes do modelo são as proficiências em Língua Portuguesa e Matemática, expressas pelas respectivas notas padronizadas do Saeb. As variáveis explicativas foram construídas a partir das respostas dos estudantes aos questionários do Saeb e recodificadas em variáveis binárias (dummies), de modo a garantir a consistência entre as categorias e a comparabilidade entre os grupos, com a exclusão de respostas não declaradas ou inválidas. A Tabela 1 descreve as variáveis explicativas utilizadas nos modelos e suas respectivas codificações, comuns às duas áreas de conhecimento analisadas.

TABELA 1 – Descrição e codificação das variáveis explicativas do modelo

Variável	Descrição	Codificação
IN_PUBLICA	Tipo de rede de ensino	1 = Pública; 0 = Privada
SEXO_M	Sexo do estudante	1 = Masculino; 0 = Feminino
RACA_COR	Raça/cor declarada	1 = Branco; 0 = Não branco
DISTORCAO_IS	Idade-série	1 = Até 18 anos; 0 = 19 anos ou mais

Fonte: Elaboração própria, a partir dos microdados do Saeb (Brasil, 2023).

As variáveis analisadas neste estudo abrangem características individuais e institucionais dos estudantes. No âmbito individual, consideram-se o sexo, a raça/cor e a idade, esta última operacionalizada por meio da variável de distorção idade-série. No que se refere às características institucionais, inclui-se o tipo de rede de ensino, distinguindo-se entre escolas públicas e privadas. Essas dimensões permitem captar diferenças no perfil dos alunos e no contexto escolar que podem se refletir no nível de proficiência dos alunos.

O desempenho acadêmico constitui a dimensão de resultado do estudo, sendo mensurado pelas proficiências individuais em Língua Portuguesa e Matemática, conforme os resultados do Sistema de Avaliação da Educação Básica (Saeb), expressos por meio das notas padronizadas nessas áreas.

Acerca dos procedimentos adotados, iniciou-se com as análises descritivas e exploratórias dos dados, a fim de examinar o comportamento da proficiência média segundo os grupos definidos por sexo, rede de ensino, raça/cor e idade. Optou-se, assim, pela elaboração de boxplots comparativos, com indicação gráfica das médias, permitindo identificar diferenças de desempenho entre os subgrupos.

A conexão entre características individuais e a proficiência dos estudantes foi estimada por meio de regressão linear múltipla *(Ordinary Least Squares* – OLS). Segundo Devore (2006), esse modelo probabilístico busca descrever a dependência de uma variável resposta Y em relação a um conjunto de variáveis explicativas $x_1, x_2, \ldots, x_k$, sendo representado pela equação:

$$Y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_k x_k + \varepsilon,$$

O termo ε_i é um erro aleatório com média zero e variância constante. Cada coeficiente β_i expressa a variação média esperada em Y quando x_i aumenta uma unidade, mantendo-se fixos os demais previsores (Devore, 2006).

A seguir apresenta-se o modelo empregado, seguido da discriminação de suas variáveis:

$$Y_i = \beta_0 + \beta_1 \text{IN_PUBLICA}_i + \beta_2 \text{SEXO_M}_i + \beta_3 \text{RACA_COR}_i + \beta_4 \text{DISTORCAO_IS} + \varepsilon_i$$

- Y_i representa a proficiência em Língua Portuguesa do aluno *i*;
- β_0 é o intercepto do modelo;
- β_1, β_2, β_3, β_4 são os coeficientes das variáveis explicativas;
- ε_i é o termo de erro aleatório.

Com o objetivo de avaliar a adequação estatística e a validade empírica do modelo estimado, foram realizados testes diagnósticos sobre os resíduos. A normalidade foi examinada por meio do teste de Jarque-Bera, cuja hipótese nula assume que os resíduos seguem distribuição normal, enquanto a homocedasticidade foi avaliada pelo teste de Breusch-Pagan, que testa a constância da variância dos erros.

Esses procedimentos possibilitam maior nível de confiança nas inferências realizadas a partir das estimativas, assegurando que sejam estatisticamente consistentes. Nos casos em que se identificou a presença de heterocedasticidade, as inferências passaram a ser conduzidas com erros-padrão robustos do tipo HC1, conforme proposto por White (1980), a fim de preservar a validade estatística dos resultados.

RESULTADOS E DISCUSSÃO

Antes de tudo, é de expressiva importância verificar a distribuição da amostra. Em 2023, a proficiência em Língua Portuguesa no ensino médio do Maranhão foi de 57,2 pontos, com desvio-padrão de 49,6, enquanto, em Matemática, a média foi de 253,6 pontos, com desvio-padrão de 44,5, conforme a Tabela 2. As medianas, próximas às médias em ambas as disciplinas, sugerem distribuições relativamente simétricas dos resultados.

TABELA 2 – Estatísticas descritivas das variáveis do modelo e desempenho médio por grupo

Variável	Nº de observações	Contínua			Dummy		
		Média (Y)	DP (Y)	Mediana (Y)	% do grupo (=1)	Proficiência média (=1)	Proficiência média (=0)
Proficiência LP	64.476	257,2	49,6	255,5	-	-	-
Rede pública (=1)	64.476	-	-	-	96,6	255,5	304,2
Sexo masculino (=1)	64.011	-	-	-	46,2	251,0	262,5
Branco (=1)	63.462	-	-	-	17,9	270,0	254,6
≤ 18 anos (=1)	64.476	-	-	-	84,4	261,8	232,2
Proficiência Matemática	64.476	253,6	44,5	248,2	-	-	-
Rede pública (=1)	64.476	-	-	-	96,6	251,8	303,8
Sexo masculino (=1)	64.011	-	-	-	46,2	256,5	251,1
Branco (=1)	63.462	-	-	-	17,9	263,5	251,6
≤ 18 anos (=1)	64.476	-	-	-	84,4	256,9	235,7

Fonte: Elaboração própria, a partir dos microdados do Saeb (Brasil, 2023).
Nota: As estatísticas da variável contínua referem-se à proficiência média geral. Para variáveis dummies, apresentam-se a proporção de casos (=1) e as médias condicionais de proficiência para cada grupo.

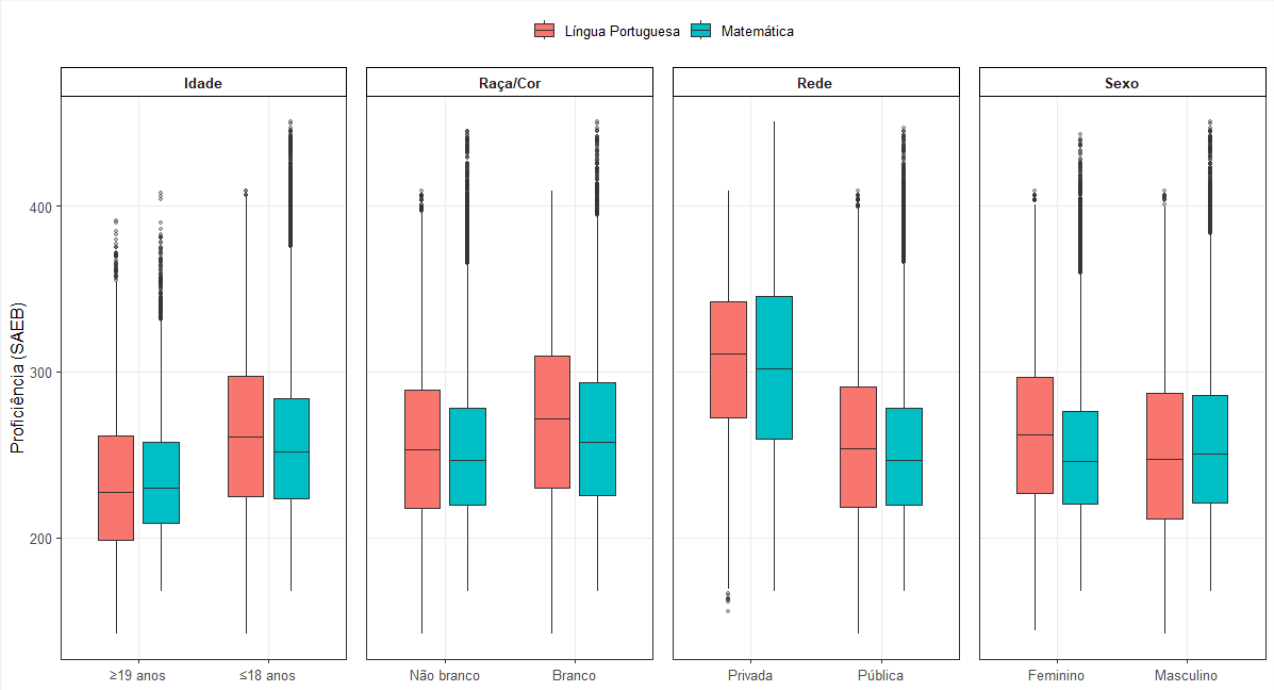
As estatísticas descritivas indicam que a maioria dos estudantes é da rede pública de ensino, que concentra 96,6% de toda a amostra. Ressalta-se que, nas escolas públicas, a avaliação é censitária, ao passo que, na rede privada, a participação ocorre por amostragem. No que se refere ao perfil dos alunos, 46,2% são do sexo masculino, 17,9% se autodeclaram brancos e 84,4% possuem até 18 anos de idade, indicando que apenas uma parcela reduzida apresenta distorção idade-série.

Ao analisar detalhadamente os resultados por grupos, observa-se que estudantes da rede privada apresentam médias de proficiência superiores tanto em Língua Portuguesa quanto em Matemática. De modo semelhante, alunos autodeclarados brancos e aqueles com idade adequada à série possuem desempenho médio mais elevado nas duas disciplinas. No que se refere ao sexo, as meninas apresentam médias superiores em Língua Portuguesa. Em Matemática as diferenças entre os sexos são menos acentuadas, ainda assim, nota-se notas maiores para os meninos (Tabela 2).

Esses padrões foram reforçados pelos *boxplots* (Figura 1), que evidenciam distribuição de proficiências sistematicamente mais baixas em Língua Portuguesa e Matemática entre estudantes da rede pública e daqueles pertencentes a grupos socialmente mais vulneráveis.

Adicionalmente, há maior dispersão das proficiências na rede pública, especialmente em Matemática, o que sugere elevada heterogeneidade interna e reflete desafios estruturais associados à oferta educacional nesse segmento.

FIGURA 1 – *Boxplots* da proficiência em Língua Portuguesa e Matemática por variáveis explicativas



Fonte: Elaboração própria, a partir dos microdados do Saeb (Brasil, 2023).

De acordo com a Tabela 3, a análise dos resultados obtidos pelo modelo de regressão linear múltipla com erros robustos, apontaram significância em ambos os modelos, o teste F indica significância conjunta das variáveis explicativas ($p < 0{,}001$), rejeitando a hipótese nula de que todos os coeficientes são conjuntamente iguais a zero, indicando que o conjunto das variáveis independentes possui um efeito estatisticamente relevante para explicar as médias das notas nos exames.

Todavia, o poder explicativo do modelo avaliado pelo R^2 ajustado, é reduzido, sendo de 0,0926 para Língua Portuguesa e 0,0813 para Matemática, isto é, cerca de 9,3% e 8,1%,

respectivamente, da variabilidade total do desempenho dos estudantes são explicados pelas quatro variáveis individuais e institucionais incluídas (Tabela 3). Vale ressaltar que preditores binários, por sua própria natureza, possuem capacidade explicativa limitada, o que contribui para contextualizar os percentuais observados.

Em contrapartida a essa significância, os modelos propostos também deixam a maior parte da variação sem explicação, o que remete que outros fatores além das características contempladas podem exercer influência sob ambas as proficiências, a saber: infraestrutura da escola, aspectos pedagógicos e nível socioeconômico familiar.

TABELA 3 – Resultados da regressão linear múltipla para Língua Portuguesa e Matemática, com erros-padrão robustos (HC1)

Disciplinas	Variável	Coeficiente	Erro-padrão robusto	Valor t	Sig.
Língua Portuguesa	Intercepto	278,64	1,17	238,8	***
	Rede pública	-43,07	1,06	-40,7	***
	Sexo masculino	-9,97	0,38	-26,2	***
	Branco	12,37	0,52	24	***
	≤18 anos	26,83	0,49	55,3	***
	R^2 ajustado = 0.0926; F (4, 63046) = 1.61×10^3; p < 0.001				
Matemática	Intercepto	277,57	1,3	212,8	***
	Rede pública	-47,2	1,24	-38,1	***
	Sexo masculino	6,53	0,35	18,9	***
	Branco	8,7	0,48	18,1	***
	≤18 anos	20,34	0,42	48,8	***
	R^2 ajustado = 0.0813; F (4, 63046) = 1.40×10^3; p < 0.001				

Fonte: Elaboração própria, a partir dos microdados do Saeb (Brasil, 2023).

Embora haja limitada capacidade explicativa, os coeficientes estimados demonstram padrões consistentes. O intercepto de Língua Portuguesa para o ensino médio maranhense indicou uma proficiência média de aproximadamente 278,6 pontos para o grupo de referência. Já o de Matemática mostrou uma proficiência média de 277,6 pontos.

O coeficiente associado à rede pública é negativo em ambos os modelos, indicando desempenho inferior dos alunos dessa rede em comparação aos da rede privada. Esse efeito é particularmente mais intenso em Matemática, com um coeficiente médio de −47,2 pontos, o que reforça não apenas a existência, mas também a magnitude das desigualdades entre as redes de ensino, especialmente nessa disciplina (Tabela 3).

A variável sexo apresenta resultados consistentes com a literatura educacional. Conforme evidenciado por Soares e Alves (2003), observa-se vantagem feminina em testes de leitura e Língua Portuguesa, ao passo que o desempenho masculino tende a ser superior em Matemática. Nesse sentido, o coeficiente associado ao sexo masculino é negativo em Língua Portuguesa (−9,97), indicando que os meninos obtêm, em média, notas inferiores às das meninas nessa disciplina no Maranhão. Em contraste, em Matemática, o sexo masculino apresenta efeito positivo (+6,53), refletindo melhor desempenho médio dos meninos.

Por sua vez, a variável raça/cor apresenta coeficiente positivo em ambos os modelos, com estudantes brancos obtendo, em média, 12,37 pontos adicionais em Língua Portuguesa e 8,7 pontos em Matemática. Esse resultado indica desempenho superior desse grupo e reforça a evidência apontada por Soares e Alves (2003) de que as disparidades raciais no rendimento escolar persistem mesmo quando os alunos compartilham um mesmo ambiente de ensino.

Somado a isso, os estudantes sem distorção idade-série apresentam desempenho superior em ambas as disciplinas, com ganhos médios de 26,83 pontos em Língua Portuguesa e 20,34 pontos em Matemática em relação àqueles em situação de atraso escolar. Essa diferença pode ser associada ao efeito cumulativo da defasagem idade-série, que, conforme apontam Portella; Bussmann e Oliveira (2017), impacta não apenas a probabilidade de reprovação, mas também eleva o risco de abandono escolar.

A avaliação dos modelos contou com a aplicação de testes diagnósticos sobre os resíduos. No caso do modelo de Língua Portuguesa, o teste de Jarque–Bera indicou desvio de normalidade ($\chi^2 = 833$; $p < 0{,}001$), levando à rejeição da hipótese nula, resultado comum em amostras extensas, dado o elevado poder dos testes formais em detectar desvios mesmo pouco relevantes do ponto de vista prático. Resultado análogo foi observado no modelo de Matemática, cujo teste de Jarque–Bera também apontou ausência de normalidade dos resíduos ($\chi^2 = 3.876$; $p < 0{,}001$).

Quanto à homocedasticidade, o teste de Breusch–Pagan revelou a presença de heterocedasticidade em ambos os modelos, tanto em Língua Portuguesa (BP = 460; $p < 0{,}001$) quanto em Matemática (BP = 1.446; $p < 0{,}001$), o que fundamenta a utilização de erros-padrão robustos do tipo HC1 nas estimativas. Em conjunto, esses diagnósticos indicam que, apesar das violações às hipóteses clássicas, os modelos estão adequadamente especificados para a análise dos determinantes do desempenho escolar, assegurando a robustez das inferências estatísticas apresentadas.

CONSIDERAÇÕES FINAIS

Este estudo buscou analisar alguns fatores associados à atuação dos estudantes do ensino médio em Língua Portuguesa e Matemática no Maranhão, com base nos microdados do Saeb 2023, considerando características individuais e institucionais dos alunos. O trabalho, realizado por meio de um modelo de regressão linear múltipla, mostrou que o desempenho escolar pode ser afetado por esses fatores, mas que eles não são suficientes para explicar, de forma isolada, toda a variabilidade observada nos resultados educacionais.

Entre as variáveis examinadas, os resultados indicaram que os estudantes da rede pública, em comparação aos da rede privada, apresentaram desempenho inferior em ambas as disciplinas, com maior intensidade em Matemática. Ademais, identificaram-se diferenças associadas ao sexo, à raça/cor e à distorção idade-série, o que reforça o papel das características individuais e institucionais na determinação dos níveis de proficiência. Em análise conjunta, as

evidências sugerem que as desigualdades relacionadas a esses atributos tendem a se reproduzir no interior do sistema educacional.

O recorte por sexo, raça/cor e distorção idade-série evidenciou tendências relevantes no desempenho escolar dos estudantes. As alunas apresentaram desempenho superior em Língua Portuguesa, resultado que corrobora a literatura sobre diferenças de proficiência entre áreas do conhecimento segundo o sexo. Por sua vez, estudantes autodeclarados brancos registraram médias mais elevadas em ambas as disciplinas, em comparação aos não brancos, sinalizando a persistência de desigualdades raciais no desempenho educacional, inclusive em um contexto regional específico como o Maranhão. Ademais, a ausência de distorção idade-série mostrou-se associada a níveis mais elevados de proficiência, reforçando a compreensão de que o atraso escolar produz efeitos cumulativos e negativos sobre a trajetória educacional dos estudantes.

Embora os modelos estimados apresentam baixo poder explicativo, os coeficientes obtidos são estatisticamente significativos e revelam padrões robustos, indicando que fatores não observados exercem influência relevante sobre a determinação do desempenho e não foram captados neste estudo. Tal limitação não invalida os resultados, mas evidencia a necessidade de uma abordagem multifatorial na análise do desempenho escolar.

Em síntese, este trabalho contribui para a compreensão dos fatores associados ao desempenho dos estudantes do ensino médio no Maranhão, ao evidenciar como características individuais e institucionais se articulam às desigualdades educacionais observadas. Os resultados indicam que a melhoria da qualidade do ensino médio no estado requer abordagens multidimensionais, capazes de considerar simultaneamente as diferenças de sexo, raça/cor e distorção idade-série, bem como as desigualdades entre as redes de ensino. Nessa perspectiva, o aprimoramento dos resultados educacionais depende não apenas de investimentos em infraestrutura e formação docente, mas também da formulação de políticas públicas integradas que enfrentam as disparidades estruturais que condicionam as trajetórias escolares.

REFERÊNCIAS

BRASIL. Ministério da Educação. Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira. **Inep iniciou a aplicação do Saeb 2025.** Brasília, DF, 2025. Disponível em: https://www.gov.br/inep/pt-br/centrais-de-conteudo/noticias/saeb/inep-iniciou-a-aplicacao-do-saeb-2025. Acesso em: 08 nov. 2025.

BRASIL. Ministério da Educação. Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira. **Sistema de Avaliação da Educação Básica (Saeb).** Brasília, DF, 2023. Disponível em: https://www.gov.br/inep/pt-br/areas-de-atuacao/avaliacao-e-exames-educacionais/saeb. Acesso em: 25 out. 2025.

DEVORE, Jay L. **Probabilidade e estatística:** para engenharia e ciências. Tradução de: Joaquim Pinheiro Nunes da Silva. São Paulo, SP: Cengage Learning, 2006.

DEWEY, John. **Democracia e educação:** introdução à filosofia da educação. 4. ed. São Paulo, SP: Companhia Editora Nacional, 1979.

HANUSHEK, Eric A. Economic growth in developing countries: the role of human capital. **Economics of Education Review,** Oxford, v. 37, p. 204-212, 2013. DOI: http://dx.doi.org/10.1016/j.econedurev.2013.04.005

IBGE – INSTITUTO BRASILEIRO DE GEOGRAFIA E ESTATÍSTICA. **Panorama Censo 2022.** Rio de Janeiro, RJ, 2022. Disponível em: https://censo2022.ibge.gov.br/panorama/indicadores. Acesso em: 04 fev. 2026.

IMESC – INSTITUTO MARANHENSE DE ESTUDOS SOCIOECONÔMICOS E CARTOGRÁFICOS. **Boletim Social no Maranhão,** São Luís, MA, v. 6, n. 1, 2024. Disponível em: https://imesc.ma.gov.br/wp-content/uploads/2024/04/V-6-N-1-2024.pdf. Acesso em: 05 fev. 2026.

MARANHÃO (Estado). Secretaria da Educação. **Analfabetismo cai no Maranhão e tem a menor taxa dos últimos 12 anos, segundo dados do IBGE 2022.** São Luís, MA, 2024. Disponível em: https://www.educacao.ma.gov.br/analfabetismo-cai-no-maranhao-e-tem-a-menor-taxa-dos-ultimos-12-anos-segundo-dados-do-ibge-2022/. Acesso em: 07 fev. 2026.

PORTELLA, Alysson Lorenzon; BUSSMANN, Tanise Brandão; OLIVEIRA, Ana Maria Hermeto. A relação de fatores individuais, familiares e escolares com a distorção idade-série no ensino público brasileiro. **Nova Economia,** Belo Horizonte, MG, v. 27, n. 3, p. 477-509, 2017. DOI: http://dx.doi.org/10.1590/0103-6351/3138

SOARES, José Francisco; ALVES, Maria Teresa Gonzaga. Desigualdades raciais no sistema brasileiro de educação básica. **Educação e Pesquisa,** São Paulo, SP, v. 29, n. 1, p. 147-165, 2003. DOI: https://doi.org/10.1590/S1517-97022003000100011

WHITE, Halbert. A heteroskedasticity-consistent covariance matrix estimator and a direct test for heteroskedasticity. **Econometrica,** Chichester, v. 48, n. 4, p. 817-838, 1980. DOI: https://doi.org/10.2307/1912934