SciELO Preprints

Estado da publicação: O preprint não foi publicado em outro meio.

Desigualdade educacional e fatores socioeconômicos uma análise com decomposição PARAFAC dos dados do ENEM 2022-2023

Lara Carla Freitas Rodrigues, Luiza Helena Félix de Andrade, Sílvio Roberto Fernandes

https://doi.org/10.1590/SciELOPreprints.18791

Submetido em: 2026-11-06
Postado em: 2026-11-10 (versão 1)
(AAAA-MM-DD)

SciELO Preprints - Este documento é um preprint e sua situação atual está disponível em: https://doi.org/10.1590/SciELOPreprints.18791

ARTIGO

DESIGUALDADE EDUCACIONAL E FATORES SOCIOECONÔMICOS: UMA ANÁLISE COM DECOMPOSIÇÃO PARAFAC DOS DADOS DO ENEM 2022-2023

LARA CARLA FREITAS RODRIGUES[1]
ORCID: https://orcid.org/0009-0000-6666-7519
lara.cfrodrigues@gmail.com
LUIZA HELENA FÉLIX DE ANDRADE[1]
ORCID: https://orcid.org/0000-0002-2255-2194
luizafelix@ufersa.edu.br
SÍLVIO ROBERTO FERNANDES[1]
ORCID: https://orcid.org/0000-0001-7300-8423
silvio@ufersa.edu.br

[1] Universidade Federal Rural do Semi-Árido. Mossoró, RN, Brasil.

RESUMO: Esta pesquisa buscou compreender quais fatores socioeconômicos mais impactaram o desempenho dos estudantes do ensino médio brasileiros em 2022 e 2023, a partir dos microdados do exame nacional do ensino médio (ENEM) disponibilizados pelo Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira (INEP). Foram considerados apenas participantes que realizaram todas as provas e não eram "treineiros". A amostra foi segmentada em dois grupos de proficiência, de acordo com a média anual: os acima ou iguais à média formaram o grupo de maior desempenho, e os abaixo da média, o de menor desempenho. Os dados foram estruturados em um tensor de ordem 3 (candidatos × questões × proficiência) e analisados por meio da decomposição em fatores paralelos (PARAFAC), com suporte de ferramentas computacionais em Python. Os resultados revelaram que fatores externos ao ambiente escolar, em especial a renda familiar e a escolaridade dos pais, foram os mais determinantes para a proficiência, fornecendo evidências para a formulação de políticas públicas que enfrentem desigualdades e favoreçam maior equidade social.

Palavras-chave: desempenho acadêmico, fatores socioeconômicos, políticas educacionais.

EDUCATIONAL INEQUALITY AND SOCIOECONOMIC FACTORS: AN ANALYSIS USING PARAFAC DECOMPOSITION OF ENEM 2022-2023 DATA

ABSTRACT: This research sought to understand which socioeconomic factors most impacted the performance of Brazilian high school students in 2022 and 2023, using microdata from the National Senior High School Exam (ENEM) provided by the Anísio Teixeira National Institute for Educational Studies and Research (INEP). Only participants who completed all tests and were not "mock candidates" (non-graduating students) were considered. The sample was segmented into two proficiency groups based on the annual average: those equal to or above the average formed the high-performance group, and those below the average, the low-performance group. The data were structured into a third-order tensor (candidates × questions × proficiency) and analyzed through Parallel Factor Analysis (PARAFAC) decomposition, supported by computational tools in Python. The results revealed that factors external to the school environment, especially family income and parental education level, were the most determinant for proficiency, providing evidence for the formulation of public policies that address inequalities and promote greater social equity.

Keywords: academic performance, socioeconomic factors, educational policies.

SciELO Preprints - Este documento é um preprint e sua situação atual está disponível em: https://doi.org/10.1590/SciELOPreprints.18791

DESIGUALDAD EDUCATIVA Y FACTORES SOCIOECONÓMICOS: UN ANÁLISIS CON DESCOMPOSICIÓN PARAFAC DE LOS DATOS DEL ENEM 2022-2023

RESUMEN: Esta investigación buscó comprender qué factores socioeconómicos impactaron más en el desempeño de los estudiantes de secundaria brasileños en 2022 y 2023, a partir de los microdatos del Examen Nacional de Enseñanza Media (ENEM) proporcionados por el Instituto Nacional de Estudios e Investigaciones Educativas Anísio Teixeira (INEP). Se consideraron únicamente los participantes que realizaron todas las pruebas y que no eran "treineiros" (estudiantes de práctica). La muestra se segmentó en dos grupos de competencia, de acuerdo con el promedio anual: aquellos iguales o por encima del promedio formaron el grupo de mayor desempeño, y los que estaban por debajo del promedio, el de menor desempeño. Los datos se estructuraron en un tensor de tercer orden (candidatos × preguntas × competencia) y se analizaron mediante la descomposición en factores paralelos (PARAFAC), con el apoyo de herramientas computacionales en Python. Los resultados revelaron que los factores externos al entorno escolar, en especial el ingreso familiar y el nivel de escolaridad de los padres, fueron los más determinantes para la competencia, aportando evidencias para la formulación de políticas públicas que enfrenten las desigualdades y favorezcan una mayor equidad social.

Palabras clave: desempeño académico, factores socioeconómicos, políticas educativas.

INTRODUÇÃO

A análise sobre a relação entre a educação e as desigualdades sociais é um tema bastante recorrente no meio científico e acadêmico. Ainda no século XX, diversas linhas de pensamento foram elaboradas com a finalidade de examinar a educação formal nas suas bases, sendo a reflexão sobre o papel da escolarização na reprodução de injustiças sociais na sociedade capitalista um dos principais pontos de discussão. Sobre isso, Pierre Bourdieu (1977) deu uma visão mais ampla sobre o papel dos sistemas educacionais no fomento das desigualdades.

Diferente do senso comum difundido na época, onde a escola se comportaria como uma instituição neutra com o único objetivo de difundir o conhecimento e selecionar os melhores a partir de critérios racionais e objetivos (Gonçalves et al., 2008), Bourdieu defendia que a escola não estaria isenta do papel de reprodução das desigualdades. Para ele, o desempenho individual de cada aluno é profundamente afetado por sua bagagem cultural e social. Sendo assim, sem o reconhecimento dessas diferenças, a escola continuaria a reproduzir as vantagens aos mais privilegiados.

Em paralelo com o viés de pensamento desenvolvido por Bourdieu, diversas pesquisas foram conduzidas entre os anos de 1950 e 1970, como o relatório de Coleman (1966). O mais interessante é que seus resultados entraram em concordância com Bourdieu, pois os fatores extraescolares se mostraram mais relevantes no desempenho dos estudantes do que propriamente os fatores intraescolares.

Na realidade brasileira, esses resultados não são diferentes. A apuração dos dados obtidos através do Índice de Desenvolvimento da Educação Básica (Ideb) de 2023 demonstra que os alunos mais pobres têm um atraso de cerca de 4 anos de aprendizado, quando comparado aos alunos com maior poder aquisitivo e pertencentes a escolas particulares (Brasil, 2023). Além disso, segundo o *Programme for International Student Assessment* (PISA) de 2022 (OCDE, 2023), o Brasil registrou médias abaixo das médias internacionais em todas as áreas do programa, que compreendem matemática, literatura e ciências, com uma variação de pelo menos 15% nas notas de matemática entre os mais favorecidos e menos favorecidos economicamente.

De fato, a realidade demonstra que não é possível dissociar as questões socioeconômicas com o desenvolvimento da educação do país. Contudo, entre os diversos fatores que compõem essa dimensão, quais são mais impactantes na realidade dos estudantes brasileiros? O presente estudo busca responder essa questão, no contexto dos estudantes do ensino médio brasileiro para os anos de 2022 e

SciELO Preprints - Este documento é um preprint e sua situação atual está disponível em: https://doi.org/10.1590/SciELOPreprints.18791

2023. Para tanto, utilizou-se a análise do banco de dados do Exame Nacional do Ensino Médio (ENEM) dos respectivos anos, cuja ampla cobertura nacional, riqueza de detalhes e fácil acesso pelo portal do Instituto Nacional de Estudo e Pesquisas Educacionais Anísio Teixeira (INEP) o tornam uma fonte valiosa de informações para investigações dessa natureza. A análise foi feita a partir da aplicação do método *Parallel Factor Analysis* (PARAFAC), em português, Análise em Fatores Paralelos (Bro, 1997; Kolda e Bader, 2009).

A utilização do PARAFAC nesta pesquisa se justifica pela sua capacidade de lidar com dados multidimensionais, preservando a estrutura original das informações e permitindo a identificação de relações internas das variáveis que não seriam facilmente identificados por métodos tradicionais. Essa abordagem tem sido eficaz em estudos que buscam compreender os fatores associados ao rendimento escolar, como demonstrado por Hippolyto et al (2023). Além disso, é uma das formas, indicadas por Silva (2017), de compreender a complexidade dos fenômenos educacionais através do uso de ferramentas matemáticas e estatísticas robustas.

Também, sob condições apropriadas, a decomposição gerada pelo método é única, o que proporciona maior robustez e clareza na interpretação dos resultados. Essa característica é extremamente relevante quando se busca identificar quais fatores socioeconômicos impactam mais significativamente a proficiência dos participantes, pois os componentes extraídos têm significados diretos e não ambíguos. Assim, a aplicação do método se torna fundamental para lidar com a complexidade dos dados do ENEM, sintetizando um grande volume de variáveis em um número reduzido de componentes representativos, sem perda das informações mais relevantes para a análise.

Certamente, compreender os fatores socioeconômicos que mais influenciam o desempenho educacional possibilita a criação de políticas públicas que atuem nas raízes do problema e sejam eficazes no combate às desigualdades no país, colaborando não somente para a democratização do acesso à escola, mas a democratização do acesso ao ensino de qualidade, inclusive do ensino superior (Portela, 2007).

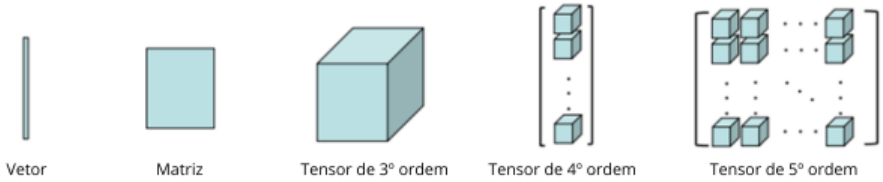
O presente trabalho está dividido em seis seções principais, após essa breve contextualização, a segunda seção tratará sobre conceitos básicos para o entendimento da técnica de análise utilizada, a terceira seção abordará os métodos utilizados para a construção do estudo, a quarta exporá os resultados, a quinta trará uma discussão detalhada sobre as questões mais relevantes e por fim, a sexta seção traz as considerações finais.

FERRAMENTAS MATEMÁTICAS PARA A ANÁLISE DE DADOS

Álgebra Tensorial

Os tensores são estruturas matemáticas multidimensionais que generalizam os conceitos de escalar (tensor de ordem 0), vetor (tensor de ordem 1), matriz (tensor de ordem 2) e arranjos com uma quantidade de dimensões ilimitadas. O número de dimensões é chamado de ordem do tensor e eles podem ser representados graficamente de acordo com a Figura 1.

Figura 1 - Representação dos tensores



Fonte: Adaptado de Yokota (2024).

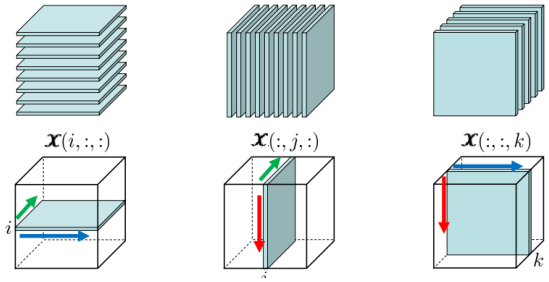
Para arranjos bidimensionais, como as matrizes, é comum se fazer uma diferenciação entre as partes do arranjo, como linhas (i) e colunas (j). Isso também é aplicado em modelos tridimensionais, que incorporam uma terceira dimensão, frequentemente denominada profundidade (k). Essa representação (i, j, k) pode ser rearranjada por fatias frontais, horizontais e verticais. A Figura 2 exemplifica como essa divisão pode ser visualizada, onde, dado o tensor $\boldsymbol{\mathcal{X}} \in \mathbb{R}^{IxJxK}$, temos a fatia horizontal $\boldsymbol{\mathcal{X}}(i,:,:) \in \mathbb{R}^{JxK}$, a fatia vertical $\boldsymbol{\mathcal{X}}(:,j,:) \in \mathbb{R}^{IxK}$ e a fatia frontal $\boldsymbol{\mathcal{X}}(:,:,k) \in \mathbb{R}^{IxJ}$. Nesse contexto, o tensor atua

SciELO Preprints - Este documento é um preprint e sua situação atual está disponível em: https://doi.org/10.1590/SciELOPreprints.18791

como um contêiner de dados, onde cada elemento $\boldsymbol{\mathcal{X}}(i, j, k)$ corresponde a um valor numérico referente ao objeto de estudo.

De forma geral, as técnicas que se utilizam do modelo multidimensional buscam capturar a variabilidade, reduzir a redundância das informações e descrever o banco de dados, geralmente muito grande e complexo, de forma mais simples. De forma semelhante ao que ocorre quando se utiliza a técnica da decomposição em valores singulares, em inglês, *Singular Value Decomposition* (SVD), para descrever uma matriz em um conjunto de três matrizes mais simples e estruturadas, facilitando sua análise e interpretação (Kolda e Bader, 2009).

Figura 2 - Representação das fatias de um tensor



Fonte: Adaptado de Yokota (2024)

Método PARAFAC

O PARAFAC, também conhecido como CANDECOMP ou CP, é uma técnica de análise de dados que utiliza diversos artifícios matemáticos para decompor um tensor em partes mais simples que capturem os padrões internos dos dados, os chamados fatores latentes, ou vetores de carregamento, ou componentes. Formalmente, dado um tensor $\boldsymbol{\mathcal{X}} \in \mathbb{R}^{IxJxK}$, a decomposição PARAFAC é pode ser descrita como:

$$\boldsymbol{\mathcal{X}} \approx \sum_{r=1}^{R} \lambda_r \cdot a_r \circ b_r \circ c_r ,$$

em que $a_r \in \mathbb{R}^I$, $b_r \in \mathbb{R}^J$ e $c_r \in \mathbb{R}^K$ são os vetores de carregamento para a componente r, $\circ$ denota o produto externo, e λ_r é um escalar que atua como peso de cada componente. Esses vetores são organizados em matrizes de carregamento $\mathbf{A} = [a_1, \ldots, a_R] \in \mathbb{R}^{IxR}$, $\mathbf{B} = [b_1, \ldots, b_R] \in \mathbb{R}^{JxR}$ e $\mathbf{C} = [c_1, \ldots, c_R] \in \mathbb{R}^{KxR}$. As colunas dessas matrizes correspondem aos vetores de carregamento, enquanto as linhas são todas as entradas da dimensão original.

Apesar de sua utilidade ao lidar com tensores, devido a possibilidade de otimizar grandes conjuntos de dados, o PARAFAC apresenta algumas limitações importantes: ele exige que os dados estejam completos, sem espaços ausentes, e de dimensões compatíveis entre todos os modos, não permitindo a análise de matrizes de tamanhos diferentes. Além de exigir um ajuste criterioso do número de componentes para garantir sua principal propriedade: a unicidade (Bro e Kiers, 2003), característica que garante um resultado único, independente de quantas vezes o método for aplicado na mesma base de dados.

Construção do modelo PARAFAC

A unicidade depende da estimação correta do número de componentes, que pode ser obtido por métodos analíticos custosos. Por isso a decomposição é normalmente ajustada utilizando o algoritmo *Alternating Least Squares* (ALS) (Kossaifi, 2021) que, por meio de iterações, resolve problemas de mínimos quadrados mantendo duas matrizes fixas e atualizando a terceira.

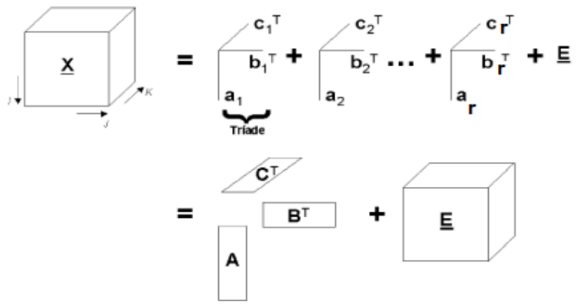
Esse processo é repetido até que a variação das iterações seja suficientemente pequena. O modelo resultante fica da forma descrita por:

SciELO Preprints - Este documento é um preprint e sua situação atual está disponível em: https://doi.org/10.1590/SciELOPreprints.18791

$$x_{ijk} = \sum_{r=1}^{R} a_{ir} b_{jr} c_{kr} + e_{ijk} .$$

sendo R o número de componentes do modelo, a_{ir}, b_{jr} e c_{kr} os valores pesos da r-ésima componente e e_{ijk} o ruído do modelo, ou seja, a informação não relevante para a análise (Silva et al., 2013). Uma representação pictórica do modelo pode ser vista na Figura 4.

Figura 4 - Representação pictórica do modelo PARAFAC utilizando o ALS



Fonte: Adaptado de Hippolyto et al (2023).

Como os modelos gerados pelo ALS apresentam algum erro residual, é prática comum a construção de múltiplos modelos com diferentes números de componentes, avaliando-se posteriormente o erro de cada modelo em comparação ao tensor original utilizando o *Sum of Squared Errors* (SSE) (Kolda e Bader, 2009). Também, a qualidade do modelo pode ser verificada pelo *Core Consistency Diagnostic* (CORCONDIA) (Bro e Kiers, 2003), método que verifica a organização interna das informações. Além disso, para verificar a unicidade da solução, pode-se utilizar o *Fit Mean Square* (FMS) (Kossaifi, 2021), que mede a similaridade entre os componentes fatoriais de dois modelos com o mesmo número de componentes, sendo fundamental devido à sensibilidade do PARAFAC à inicialização. Caso o número R de componentes não esteja bem estimado, múltiplas execuções com inicializações aleatórias podem gerar conjuntos distintos de fatores para o mesmo conjunto de dados, comprometendo a unicidade do modelo; o FMS permite avaliar se, independentemente do número de inicializações, será obtido um único resultado consistente. De modo que, em termos gerais, os critérios desejáveis ao analisar os resultados são: SSE muito baixo, FMS próximo ou igual a 1 e CORCONDIA acima de 60%.

Coeficiente de Spearman

O coeficiente de Spearman, (Zar, 2010) também conhecido como rho de Spearman (ρ), é uma medida estatística que avalia a correlação entre duas variáveis como base no posto dos dados. Ele avalia a relação em que as variáveis têm de crescer ou decrescer juntas, mesmo que ela não seja linear. O seu valor varia de -1 (correlação negativa perfeita) a 1 (correlação positiva perfeita), sendo 0 a ausência de correlação medida.

A principal vantagem do coeficiente de Spearman é sua capacidade de capturar relações crescentes ou decrescentes entre variáveis que não seguem uma relação linear exata. Esse fato é especialmente relevante em estudos socioeconômicos, como esse, onde as variáveis envolvidas muitas vezes se relacionam de maneira não linear.

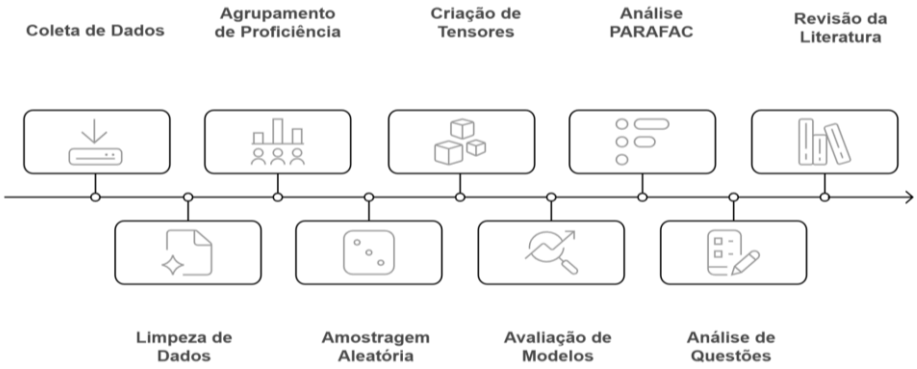
METODOLOGIA

O presente estudo adota uma abordagem metodológica mista, combinando procedimentos qualitativos e quantitativos a fim de proporcionar uma compreensão mais aprofundada e embasada sobre

SciELO Preprints - Este documento é um preprint e sua situação atual está disponível em: https://doi.org/10.1590/SciELOPreprints.18791

os fatores socioeconômicos associados ao desempenho dos participantes do ENEM nos anos de 2022 e 2023, a Figura 5 traz um breve resumo da metodologia aplicada e explanada ao longo do texto.

Figura 5 -Resumo metodológico



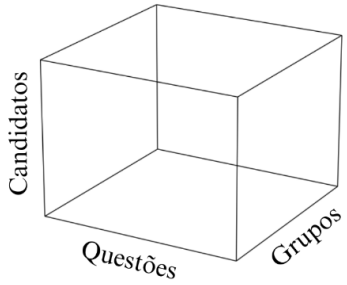
Fonte: Elaborado pelos autores com auxílio da ferramenta Napkin AI (2025).

Inicialmente, foi realizada uma pesquisa quantitativa utilizando os microdados do ENEM de 2022 e 2023, disponíveis no portal do INEP. A análise foi conduzida com bibliotecas ciências de dados padrão do Python e com TensorLy (Kossaifi et al., 2019) e TLViz (Kossaifi, 2024) envolvendo tratamento de dados, técnicas estatísticas, decomposição PARAFAC e análise exploratória das respostas relevantes. Foram aplicados filtros para excluir participantes "treineiros", aqueles que não haviam concluído o ensino médio, e os que faltaram a algum dia de prova, garantindo consistência nas médias. As respostas foram padronizadas em números inteiros, convertendo letras das alternativas do gabarito em números, seguindo a lógica A=1, B=2, etc.

Os participantes foram divididos em grupos de maior e menor proficiência, com base na média geral de cada ano. As matrizes resultantes têm colunas representando as questões e linhas as respostas dos candidatos. Para compatibilizar as dimensões exigidas pelo PARAFAC, as quantidades de amostras do conjunto de maior e menor proficiências precisam ser iguais. Assim, foram removidas aleatoriamente algumas linhas da matriz de menor desempenho, cerca de 13% em 2022 e 9% em 2023, sem comprometer a representatividade estatística, verificada por comparações de distribuições.

Com os dados organizados, foram construídos tensores para cada ano (dimensões: candidatos × questões × grupo de proficiência), conforme representação pictórica na Figura 6.

Figura 6 - Representação pictórica do tensor

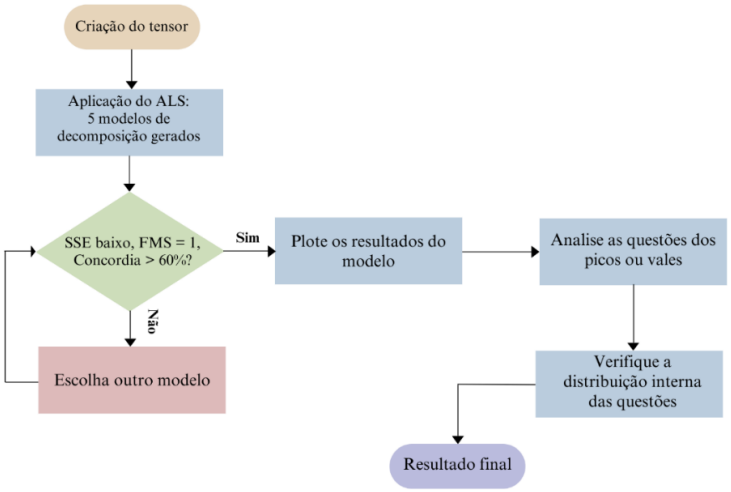


Fonte: Elaborado pelos autores (2025).

O PARAFAC gerou cinco modelos, com 1 a 5 componentes, avaliados por SSE, FMS e CORCONDIA para determinar a melhor configuração que preservasse a unicidade dos resultados. As questões indicadas pelo método como mais relevantes foram analisadas internamente e em cruzamentos entre variáveis, permitindo a construção de figuras explicativas. Além disso, foi utilizado o cálculo do coeficiente de Spearman para avaliar a correlação das variáveis de estudo com a proficiência dos estudantes. O fluxograma que resume o procedimento, da aplicação do PARAFAC até a análise das questões, está apresentado na Figura 7. Por último, foi realizada a segunda etapa da pesquisa, a análise qualitativa, interpretando os resultados em consonância com a literatura.

SciELO Preprints - Este documento é um preprint e sua situação atual está disponível em: https://doi.org/10.1590/SciELOPreprints.18791

Figura 7 - Resumo do processo PARAFAC e avaliação dos modelos



Fonte: Elaborado pelos autores (2025).

RESULTADOS

Para criação dos grupos, calculamos a média geral de proficiência de cada ano. Considerando os de menor proficiência os que estão abaixo desta média, e os de maior proficiência quando estão iguais ou superiores à média geral. Para o ano de 2022 a média foi de 551,35 pontos, com desvio padrão de 101,36 pontos e coeficiente de variação de 18,38%, o que sugere uma variação moderada entre as notas. Já no ano de 2023 a média foi de 547,16 pontos, com desvio padrão de 111,12 pontos e coeficiente de variação de 20,30%, tendo o ano de 2023 uma variação de notas maior que o do ano anterior.

Em ambos os casos, foi observado que o grupo de menor desempenho representou uma maior parcela dos candidatos, sendo, em 2022, 53,42% dos participantes e, em 2023, 52,23% dos que realizaram a prova. A Figura 8 explicita como ficou organizado cada grupo, onde as colunas da tabela, ou matriz, representam as questões descritas no Apêndice A e as linhas as respostas de cada participante. Para 2022, o tensor tem dimensões 858507 x 37 x 2. Já em 2023, o tensor tem dimensões 996662 x 37 x 2.

Figura 8 - Representação da organização final das matrizes

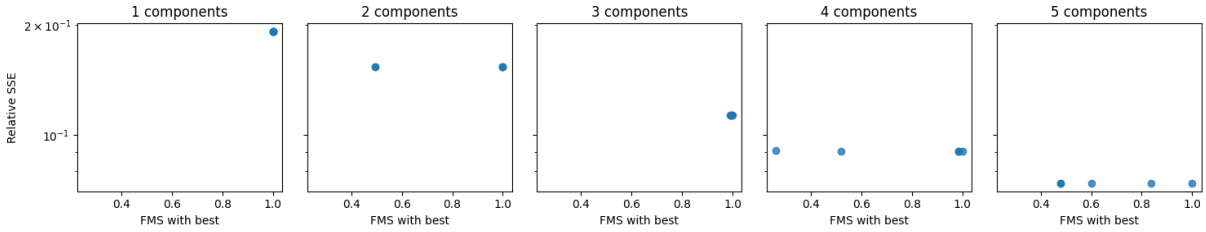
	TP_FAIXA_ETARIA	TP_SEXO	TP_ESTADO_CIVIL	TP_COR_RACA	TP_NACIONALIDADE	TP_ST_CONCLUSAO	TP_ANO_CONCLUIU	TP_ESCOLA	TP_ENSINO	TP_DEPENDENCIA_ADM_ESC	...
3	6	13	1	3	1	1	2	1	0.0	0.0	...
7	6	6	1	2	1	1	0	1	0.0	0.0	...
9	4	6	1	3	1	1	1	1	0.0	0.0	...
22	4	6	1	1	1	1	0	1	0.0	0.0	...
25	5	13	1	3	1	1	4	1	0.0	0.0	...
...	...	...	...	...	...	...	...	...	...	...	...
3476091	3	13	0	2	1	1	1	1	0.0	0.0	...
3476093	3	6	1	1	1	2	0	2	1.0	2.0	...
3476095	12	13	2	1	1	1	7	1	0.0	0.0	...
3476098	2	13	0	3	1	2	0	2	1.0	2.0	...
3476099	11	6	1	1	1	1	11	1	0.0	0.0	...

Fonte: Elaborado pelos autores (2025).

A Figura 9 explicita os resultados do SSE e FMS para o ano de 2022, demonstrando que a quantidade de componentes que apresenta menor erro de ajuste e maior estabilidade são três. Além disso, a mesma figura, também deixa bem evidente que com três componentes nós temos um resultado único, independente de quantas vezes o algoritmo for executado. Apesar disso, o teste de consistência de núcleo, demonstrou que a melhor organização das informações estaria no modelo de até 2 componentes, como descrito na Figura 10, mas como vimos pelo teste de unicidade, o modelo de 2 componentes não garante um resultado único. Dessa forma, como o modelo com 3 componentes ainda tem uma consistência de núcleo acima de 60%, esta é a quantidade componentes selecionada para análise.

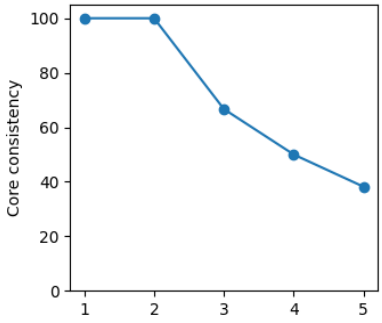
SciELO Preprints - Este documento é um preprint e sua situação atual está disponível em: https://doi.org/10.1590/SciELOPreprints.18791

Figura 9 - Teste de unicidade com SSE e FMS para o ano de 2022



Fonte: Elaborado pelos autores (2025).

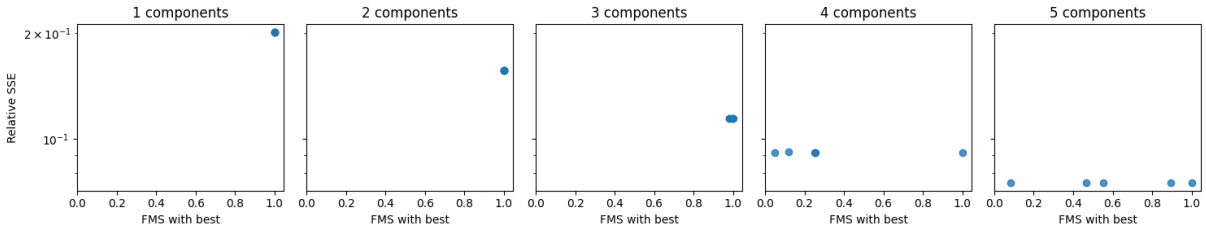
Figura 10 - Resultado CORCONDIA para 2022



Fonte: Elaborado pelos autores (2025).

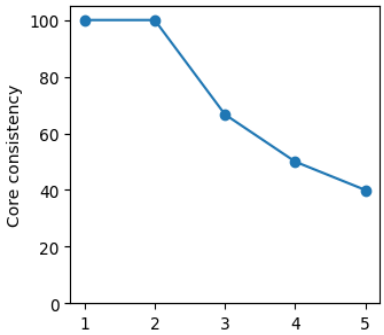
Os testes SSE e FMS para o ano de 2023 são apresentados na Figura 11. Neste caso, o modelo com duas componentes é o mais adequado, pois para modelos a partir de três componentes há mais de uma solução possível, ferindo a qualidade da unicidade do método. Para além disso, o teste CORCONDIA, representado na Figura 12, confirma que a melhor organização dos dados está no modelo com duas componentes. Logo, para o ano de 2023, o modelo que garante unicidade e menos ruído é o de duas componentes.

Figura 11 - Teste de unicidade 2023: resultado dos testes SSE e FMS



Fonte: Elaborado pelos autores (2025).

Figura 12 - Resultado CORCONDIA 2023



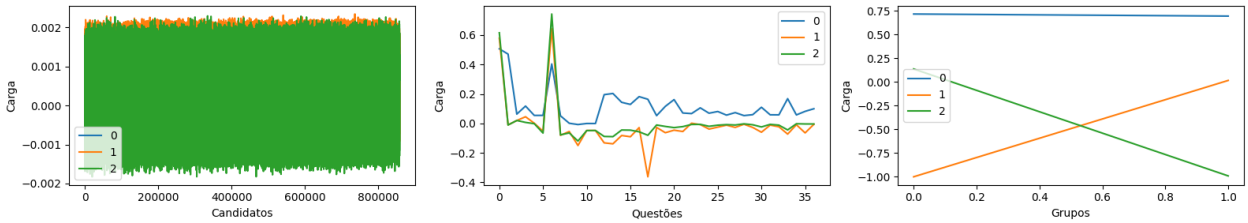
Fonte: Elaborado pelos autores (2025).

A Figura 13 representa os resultados do PARAFAC para o ano de 2022 para as três matrizes de carregamento e a Figura 14 mostra os resultados do ano de 2023 com duas componentes. Estes resul-

SciELO Preprints - Este documento é um preprint e sua situação atual está disponível em: https://doi.org/10.1590/SciELOPreprints.18791

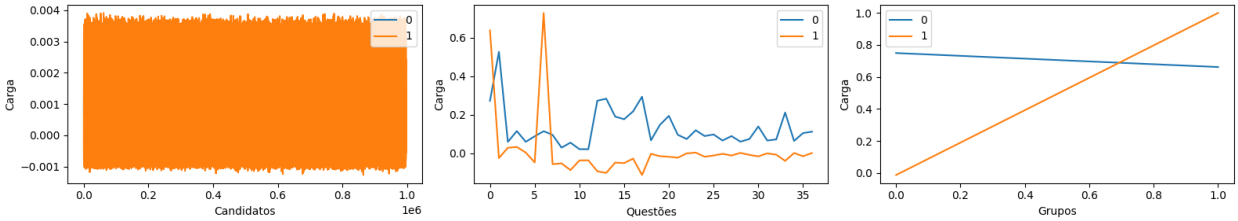
tados ajudam a descobrir quais as principais questões socioeconômicas foram mais impactantes no desempenho dos participantes de cada ano. Por isso, a análise ficará restrita ao gráfico que analisa o comportamento das questões, o que está localizado no meio das figuras, aqui chamado de módulo B.

Figura 13 -Resultados PARAFAC 2022



Fonte: Elaborado pelos autores (2025).

Figura 14 -Resultados PARAFAC 2023



Fonte: Elaborado pelos autores (2025).

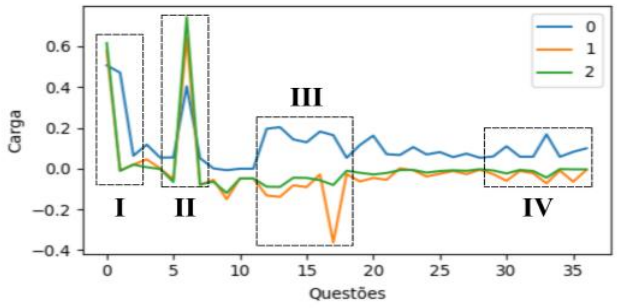
As Figuras 15 e 16 representam os resultados do módulo B para o ano de 2022, já as Figuras 17 e 18 são os resultados do mesmo módulo para o ano de 2023. É importante ressaltar que a interligação dos pontos foi feita para facilitar a visualização dos valores de picos, mas a linha que os interliga não implica em uma relação de crescimento ou decrescimento entre as questões, sendo os valores das variáveis independentes entre si.

Além disso, para que seja possível interpretar a solução dada pelo PARAFAC, é importante lembrar que ele é um método que capta os padrões do banco de dados (Silva, 2017) e oferece para o usuário uma direção de onde a pesquisa deve seguir, sendo assim, é necessário analisar as regiões onde os gráficos se comportam de maneira diferente, com picos ou vales, por isso, quatro áreas das Figuras 15 e 17 estão destacadas e serão estudadas, a fim de entender sua relevância para o tema de estudo.

As partes I e II, de ambos os anos, mostram uma variação significativa entre as questões 0 (faixa etária),1 (sexo) e 6 (ano de conclusão do ensino médio), esse fato demonstra que as respostas para essas questões tiveram padrões muito distintos, mas não necessariamente que elas tiveram relação direta com o desempenho dos estudantes, isso porque o método capta padrões, mas a interpretação é feita pelo pesquisador, como atesta Silva (2017).

Na parte III, também para os dois anos, é possível observar uma variação expressiva especialmente da componente 1 na questão 17, referente à renda mensal familiar dos participantes. Além disso, principalmente na componente 0, é observado um comportamento diferente nas questões de 12 e 13, elas se referem à escolaridade do pai e da mãe dos participantes.

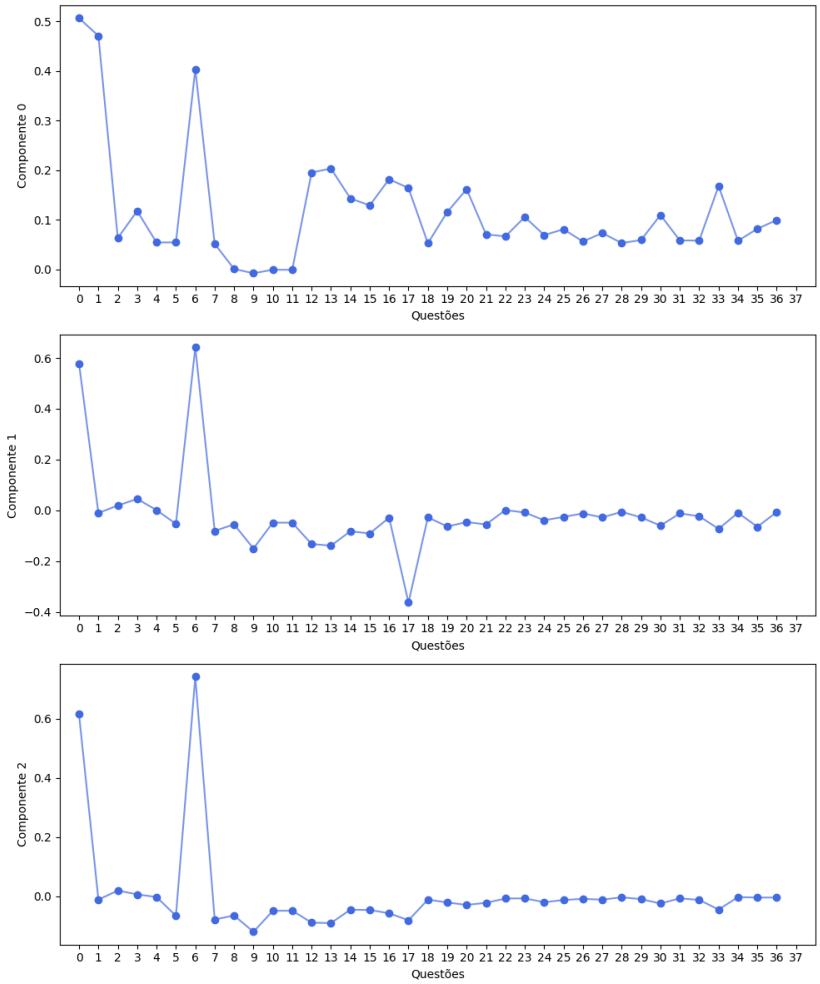
Figura 15 -Carregamento do módulo B de 2022



Fonte: Elaborado pelos autores (2025).

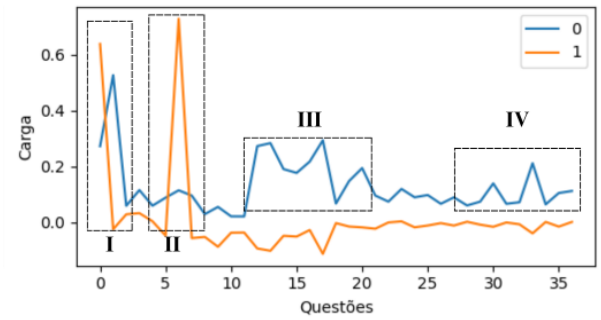
SciELO Preprints - Este documento é um preprint e sua situação atual está disponível em: https://doi.org/10.1590/SciELOPreprints.18791

Figura 16 -Gráfico de dispersão do Módulo B, separado por componentes 2022



Fonte: Elaborado pelos autores (2025).

Figura 17 -Carregamento do módulo B de 2023



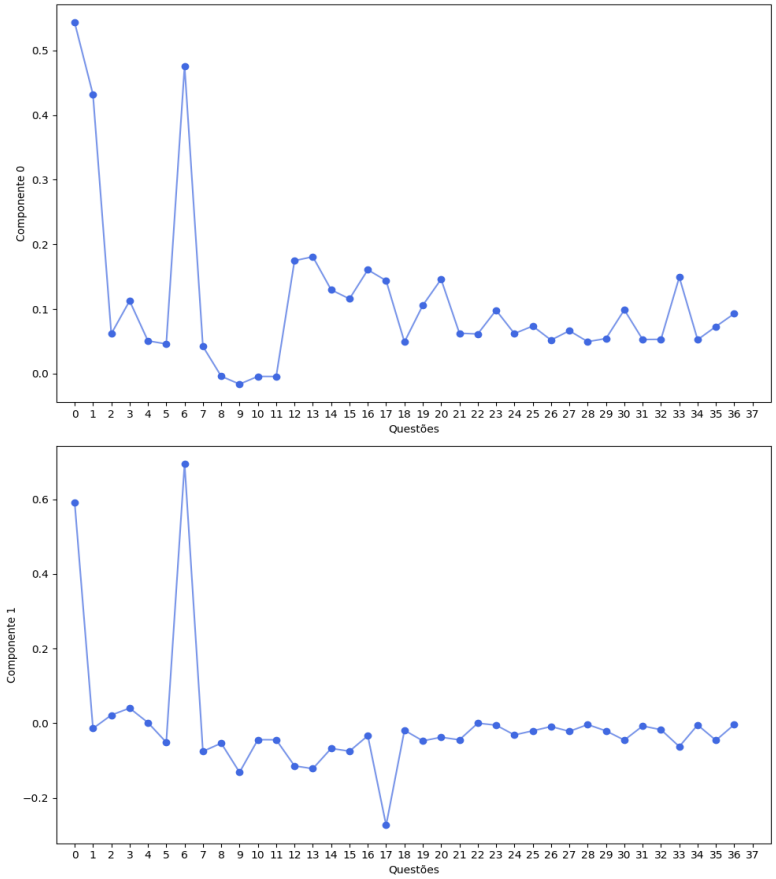
Fonte: Elaborado pelos autores (2025).

Por fim, a parte IV indica algumas questões que são vistas como picos menores, em comparação aos que já foram citados anteriormente: as questões 23, 30 e 33. A questão 23 se refere à quantas geladeiras existem na casa do participante; a questão 30, é sobre a quantidade de televisões em cores da residência; e a 33, se relaciona com a quantidade de telefones da residência.

A Tabela 1 apresenta os resultados mais relevantes encontrados a partir da análise de cada questão apontada pelo PARAFAC nas quatro partes destacadas para cada ano, mostrando a proporção das respostas em cada grupo de proficiência. Para algumas questões com mais de 2 possibilidades de respostas, a porcentagem destacada foi a da alternativa, ou intervalo de alternativas, mais predominante.

SciELO Preprints - Este documento é um preprint e sua situação atual está disponível em: https://doi.org/10.1590/SciELOPreprints.18791

Figura 18 -Gráfico de dispersão do Módulo B, separado por componentes 2023



Fonte: Elaborado pelos autores (2025).

Tabela 1: Resumo da distribuição de respostas das questões que se apresentaram como picos em 2022 e 2023

Questões	**Categoria**	**2022**		**2023**	
		Maior desempenho	**Menor desempenho**	**Maior desempenho**	**Menor desempenho**
Q00 - Faixa etária	17-19 anos	68%	63%	66,8%	59,4%
	Fora da faixa predominante	32%	37%	33,2%	40,6%
Q01 - Sexo	Feminino	57,6%	63,16%	58,07%	63,39%
	Masculino	42,4%	36,84%	41,93%	36,61%
Q06 - Ano de conclusão do ensino médio	Não informado	49,8%	56,6%	50,3%	55%
	Fora da faixa predominante	50,2%	43,4%	49,7%	45%
Q12 - Escolaridade do pai	Ensino fundamental incompleto	32,3%	53,8%	33,6%	54,8%
	Faculdade completa	15,7%	4,8%	14,5%	4,2%

SciELO Preprints - Este documento é um preprint e sua situação atual está disponível em: https://doi.org/10.1590/SciELOPreprints.18791

Questões	Categoria	2022		2023	
		Maior desempenho	Menor desempenho	Maior desempenho	Menor desempenho
Q13 - Escolaridade da mãe	Ensino fundamental incompleto	23,5%	46,1%	24,6%	47,5%
	Faculdade completa	19,2%	8%	18,1%	7,2%
Q17 - Renda	Baixa renda	37,22%	70,91%	39,66%	72,5%
	Não baixa renda	62,78%	29,09%	60,34%	27,5%
Q23 - Quantidade de geladeiras	1 geladeira	91,60%	94,62%	91,97%	94,46%
	Fora da faixa predominante	8,4%	5,38%	8,03%	5,54%
Q30 - Quantidade de televisão em cores	1 televisão	54,02%	72,56%	57,14%	73,87%
	Fora da faixa predominante	45,98%	27,44%	42,86%	26,13%
Q33 - Quantidade de celulares	2 a 3 celulares	58,56%	58,12%	58,74%	56,61%
	Fora da faixa predominante	41,44%	41,88%	41,26%	43,39%

Fonte: Elaborado pelos autores (2025).

DISCUSSÃO

Os resultados apresentados, assim como a composição dos grupos para cada questão abordada, trazem pontos de extrema relevância para o presente estudo. Entretanto, é digno de nota que a discussão será focada nas questões 12,13 e 17, referentes, respectivamente, à escolaridade da mãe, do pai e a renda familiar.

Isso porque, embora outras questões tenham sido destacadas pelo PARAFAC, como a faixa etária (questão 0), o sexo dos participantes (questão 1) e o ano de conclusão do ensino médio (questão 6), sua influência nos resultados apresentados pelo método se deve à alta dispersão nas respostas, e não necessariamente a uma associação direta com a proficiência dos estudantes, é possível notar, inclusive, que as variações da composição de cada grupo seguem a mesma tendência da distribuição geral. Esses elementos geram variações estruturais no banco de dados que o PARAFAC é condicionado matematicamente para captar (Bro, 1997), mas não implicam, por si só, em uma contribuição relevante para o objetivo deste estudo.

Além disso, variáveis destacadas pelas questões 23, 30 e 33 atreladas ao número de geladeiras, televisões e celulares disponíveis na família dos participantes, apesar de serem ligadas aos aspectos econômicos da realidade de cada estudante, aparecem como picos de menor magnitude, com uma distribuição similar às observadas nas questões 0,1 e 6. Isso indica que sua influência, embora existente, não se mostrou tão expressiva quanto às principais questões escolhidas.

SciELO Preprints - Este documento é um preprint e sua situação atual está disponível em: https://doi.org/10.1590/SciELOPreprints.18791

As questões 12 e 13, sobre a escolaridade do pai e da mãe, evidenciam uma relação entre a escolaridade dos pais e a proficiência dos estudantes. A intensidade desta relação foi calculada pelo coeficiente de Spearman, os resultados podem ser vistos na Tabela 3. Nos dois anos, os valores indicam que há uma correlação positiva moderada entre o aumento da escolaridade dos pais e a proficiência dos estudantes, sendo o pai, no contexto da presente análise, o que exerce maior influência no desempenho do participante.

Tabela 3 - Resultados do coeficiente de Spearman para escolaridade dos pais

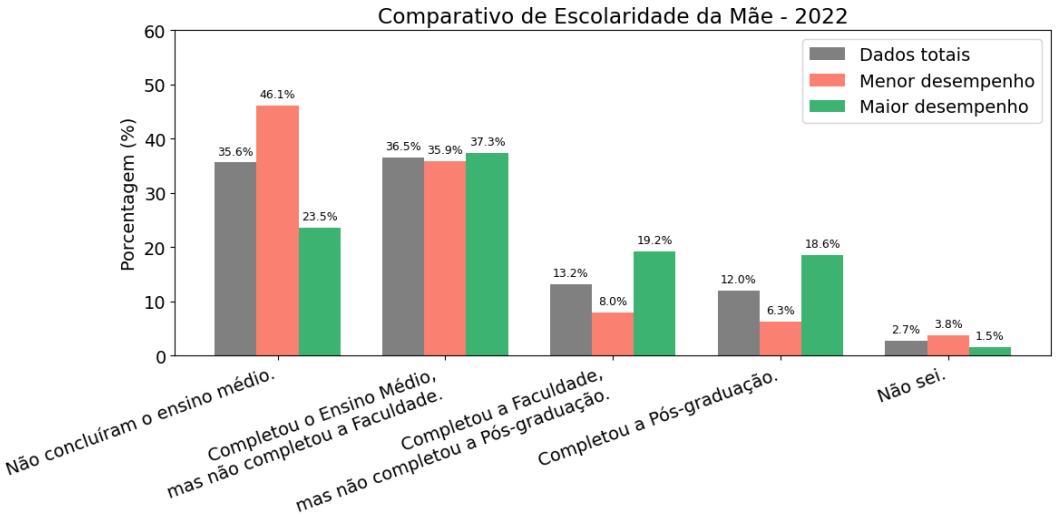
Ano	Mãe	Pai
2022	0,368	0,377
2023	0,373	0,378

Fonte: Elaborado pelos autores (2025).

Além disso, como explicita as Figuras 19, 20, 21 e 22, em ambos os anos de estudo, o grupo de menor desempenho apresentou cerca de 20% a mais de participantes cujo pai e/ou mãe não concluíram o ensino médio, em comparação ao grupo de maior desempenho. Por outro lado, com o aumento do nível da escolaridade da mãe e/ou do pai, observa-se um aumento de participantes do grupo de maior desempenho e uma diminuição de estudantes do de menor desempenho.

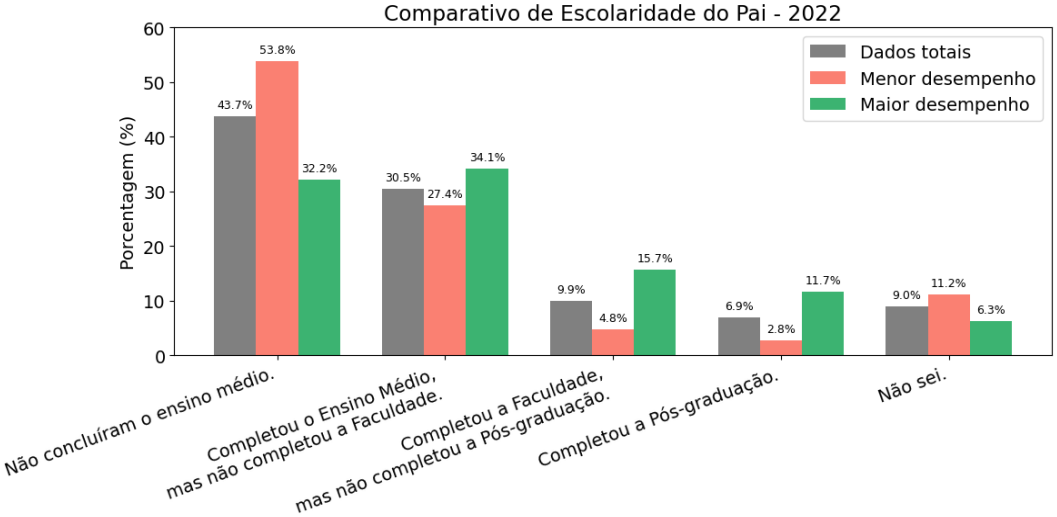
De fato, não há como questionar que o investimento em capital humano durante a vida do indivíduo exerce grandes consequências nos próximos passos de sua jornada acadêmica, profissional e socioeconômica.

Figura 19 - Relação da escolaridade da mãe - ENEM 2022



Fonte: Elaborado pelos autores (2025).

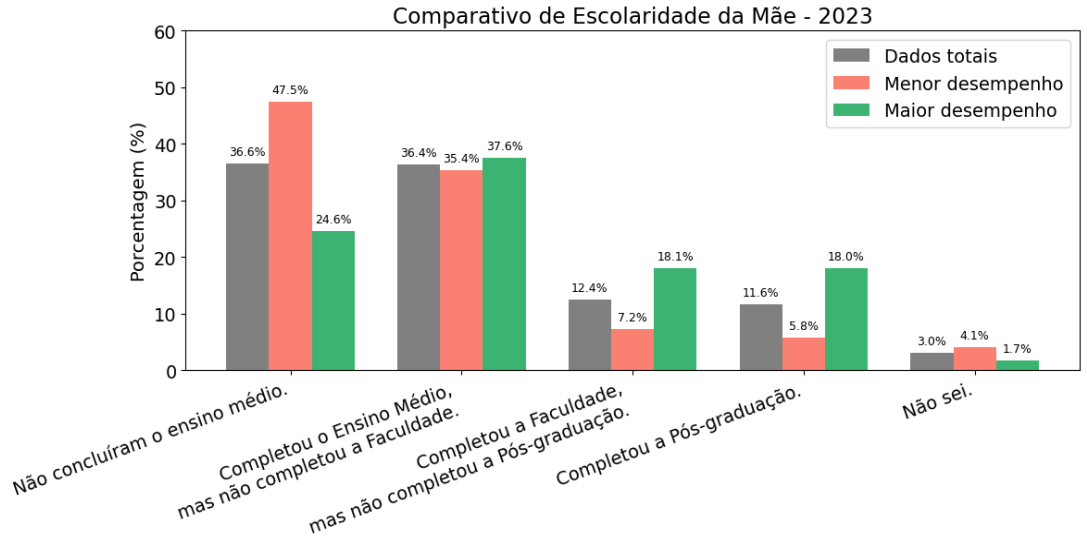
Figura 20 - Relação da escolaridade do pai - ENEM 2022



Fonte: Elaborado pelos autores (2025).

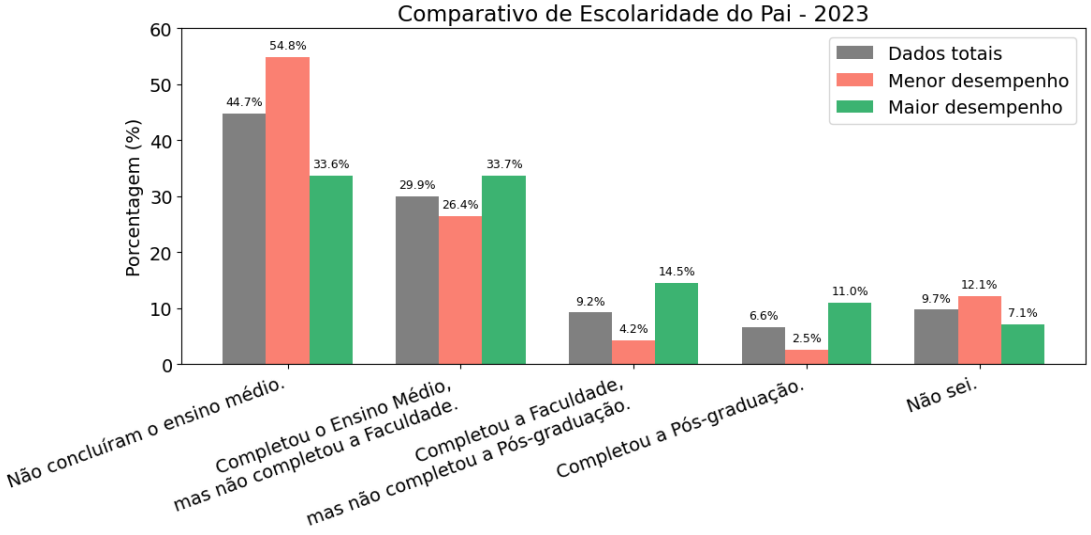
SciELO Preprints - Este documento é um preprint e sua situação atual está disponível em: https://doi.org/10.1590/SciELOPreprints.18791

Figura 21 - Relação da escolaridade da mãe- ENEM 2023



Fonte: Elaborado pelos autores (2025).

Figura 22 - Relação da escolaridade do pai - ENEM 2023



Fonte: Elaborado pelos autores (2025).

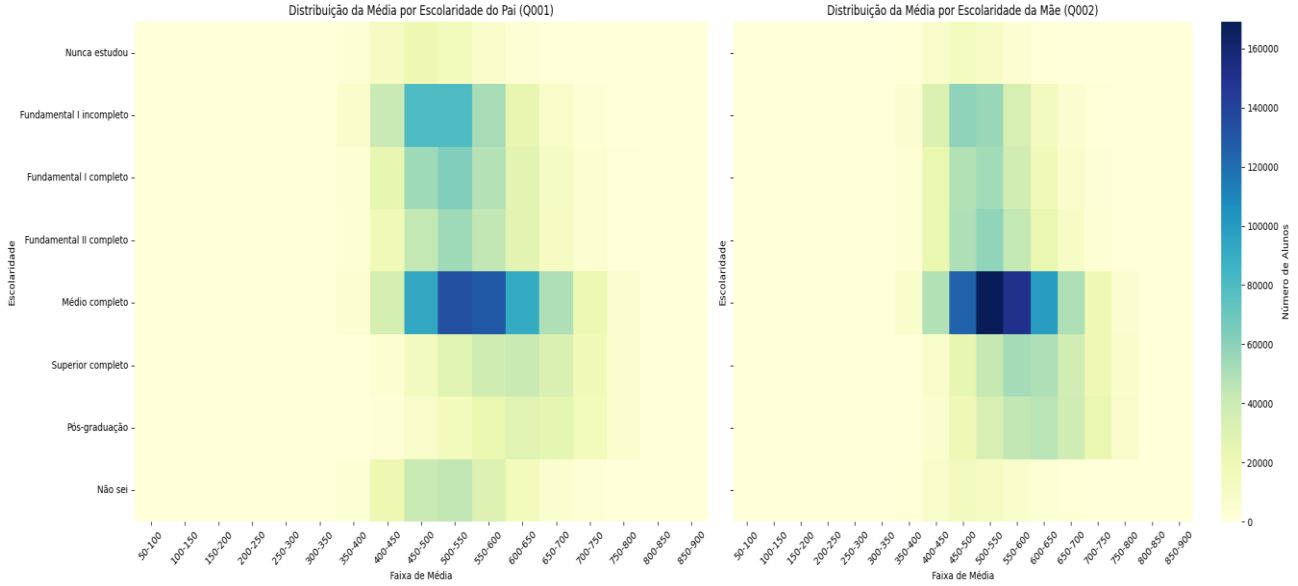
As Figuras 23 e 24 são mapas de calor que foram criados a partir do cruzamento do nível de escolaridade do pai ou da mãe em função da média de cada participante. No eixo vertical, estão dispostos os diferentes níveis de escolaridade, enquanto no eixo horizontal encontram-se as faixas das médias dos candidatos. A intensidade da cor indica a concentração de candidatos em cada combinação: quanto mais escuro o tom, maior é a quantidade de candidatos naquela faixa específica.

Percebe-se, nas Figuras 23 e 24, que apesar de haver uma grande concentração de estudantes com pais na faixa "Ensino médio completo", é evidente que, à medida que a escolaridade dos pais aumenta, há uma maior concentração de estudantes nas faixas com médias mais altas, posicionadas mais à direita no eixo horizontal.

Tal fato reforça a grande influência que a origem familiar exerce no desempenho de cada estudante (Bourdieu, 1977) e pesquisas ao longo dos anos confirmam esse resultado (Alencar et al., 2023; Barros et al., 2001; Basseto,2021). Seguindo o que foi defendido por Feijó et al. (2022), isso acontece porque, pais com uma maior escolaridade tendem a estimular mais o aprendizado dos seus filhos, mesmo que com ações indiretas, como a disponibilidade de uma melhor infraestrutura domiciliar, com espaços adequados de estudo, acesso à internet, computador e livros.

SciELO Preprints - Este documento é um preprint e sua situação atual está disponível em: https://doi.org/10.1590/SciELOPreprints.18791

Figura 23 - Relação entre a escolaridade dos pais e média de desempenho dos candidatos ENEM 2022



Fonte: Elaborado pelos autores (2025).

Figura 24 - Relação entre a escolaridade dos pais e média de desempenho dos candidatos ENEM 2023



Fonte: Elaborado pelos autores (2025).

Para além de condições de infraestrutura, estudos demonstram que pais com mais instrução têm mais interesse nos resultados dos seus filhos na escola, tendo uma melhor visão de seu desenvolvimento e uma noção mais realista quanto às melhores escolas para matricular seus filhos (Diambo e Branco, 2021). Sendo assim, eles acabam sendo responsáveis por fomentar o desenvolvimento das capacidades cognitivas dos filhos e exercem uma influência, em alguns casos, até maior que a renda domiciliar per capita da família na proficiência (Barros et al, 2001).

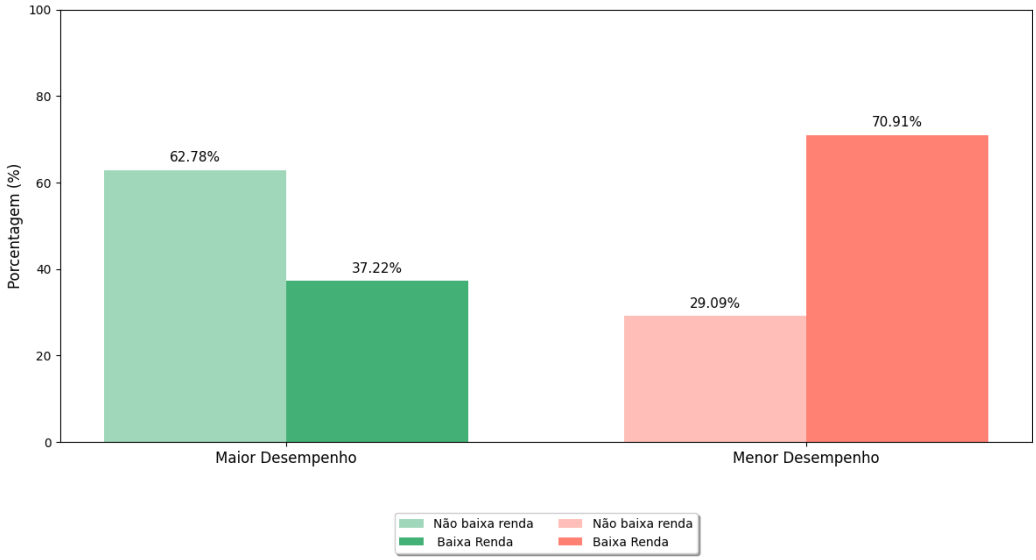
Entretanto, não se pode deixar de lado o grande poder que a disponibilidade de recursos financeiros exerce nesse processo. Não é à toa que os resultados obtidos através da questão 17, relacionada a renda dos participantes, foram tão expressivos no que tange a diferença entre a composição dos grupos de maior e menor desempenho. É digno de nota que a classificação como baixa renda ou não seguiu as definições do próprio exame para cada ano, considerando como baixa renda aqueles com renda familiar bruta mensal per capita de até 1,5 salário mínimo.

Por meio das Figuras 25 e 26, verifica-se quase uma inversão da organização de cada grupo. Enquanto, para os dois anos, os participantes de baixa renda compuseram cerca de 70% do contingente

SciELO Preprints - Este documento é um preprint e sua situação atual está disponível em: https://doi.org/10.1590/SciELOPreprints.18791

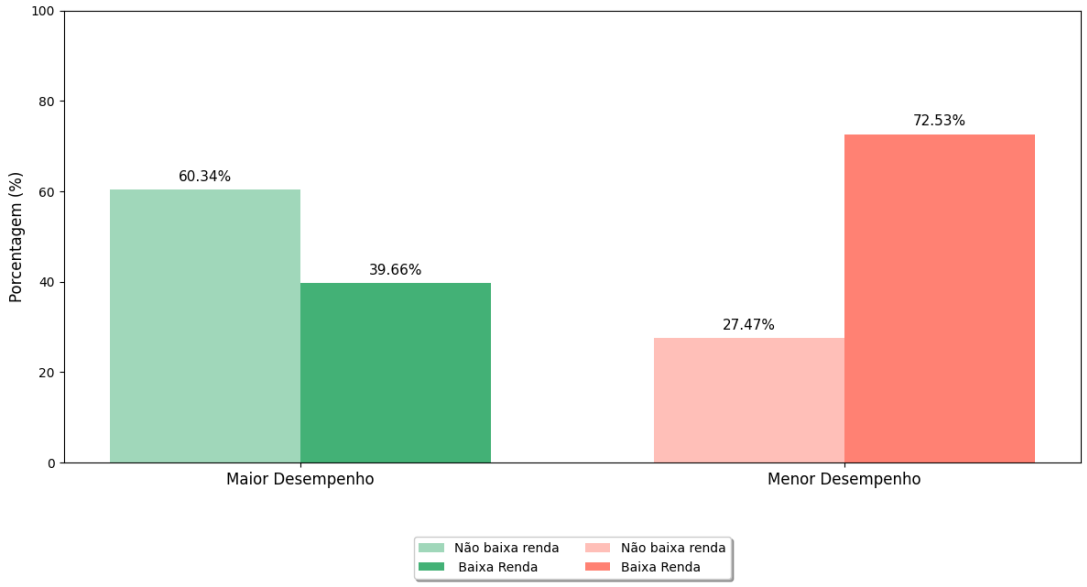
de menor desempenho, cerca de 60% dos participantes não baixa renda se qualificaram no grupo de maior desempenho. Deixando evidente que, mesmo sendo um fator debatido há muito tempo por diversos autores ao longo da literatura, a renda continua sendo um dos principais gargalos para o desenvolvimento dos estudantes, inclusive o coeficiente de Spearman foi maior para renda do que para a escolaridade dos pais, chegando a 0,48 (De Brito et al., 2022; Araújo et al.,2020; Rodrigues et al., 2024).

Figura 25 - Desempenho dos estudantes por faixa de renda - ENEM 2022



Fonte: Elaborado pelos autores (2025).

Figura 26 - Desempenho dos estudantes por faixa de renda - ENEM 2023



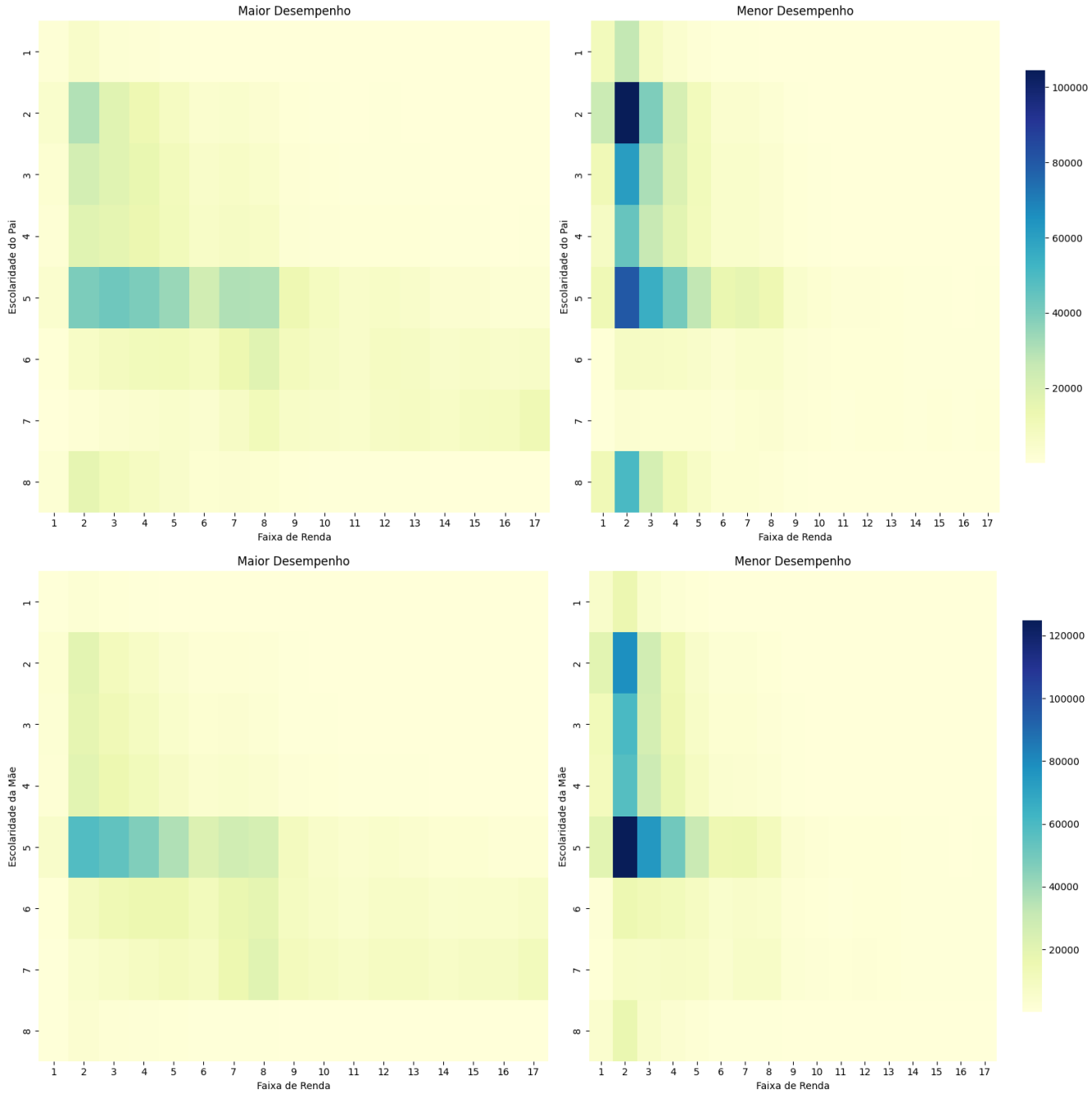
Fonte: Elaborado pelos autores (2025).

Como De Farias Souza et al. (2018) bem explica, a falta de recursos financeiros da família provoca um investimento muito inferior ao desejado no quesito da educação dos sucessores, o que colabora para um fenômeno chamado de transmissão intergeracional da pobreza, um efeito cascata onde aqueles que são pobres hoje tem menos acesso à escolarização de qualidade, como consequência tem uma menor proficiência e chances reduzidas de atuar em espaços com remuneração mais alta, o que colabora para a tendência que a renda da família continue sendo baixa. Inclusive, as Figuras 27 e 28 ajudam a verificar esse fenômeno nos anos objeto de estudo desse trabalho, por meio de mapas de calor a partir do cruzamento das informações de escolaridade dos pais e renda da família, e a frequência de candidatos para cada grupo de proficiência.

SciELO Preprints - Este documento é um preprint e sua situação atual está disponível em: https://doi.org/10.1590/SciELOPreprints.18791

O eixo vertical indica a escolaridade dos pais, sendo a faixa 1 o de menor escolaridade e a faixa 7 de maior escolaridade, já a faixa 8 corresponde às respostas "Não sei". O eixo horizontal representa as faixas de renda disponíveis no questionário socioeconômico que seguem uma ordem crescente, sendo a faixa 1 correspondente a "Sem renda" e a faixa 17 a uma renda igual ou superior a vinte salários-mínimos. Além disso, quanto mais escuro o azul, maior quantidade de participantes na região.

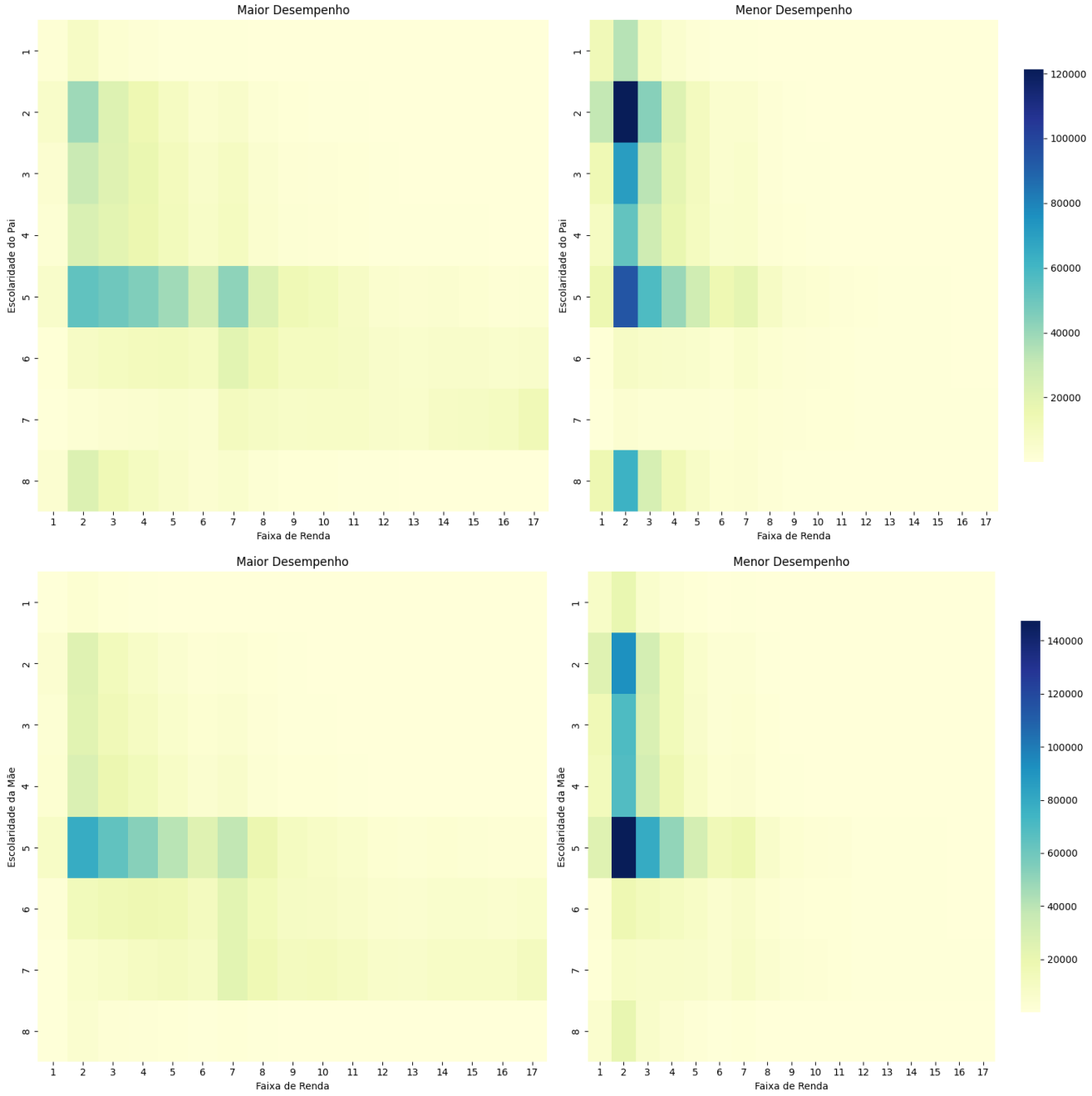
Figura 27 - Relação entre a escolaridade dos pais, renda e proficiência para o ENEM 2022



Fonte: Elaborado pelos autores (2025).

SciELO Preprints - Este documento é um preprint e sua situação atual está disponível em: https://doi.org/10.1590/SciELOPreprints.18791

Figura 28 - Relação entre a escolaridade dos pais, renda e proficiência para o ENEM 2023



Fonte: Elaborado pelos autores (2025).

É notório que, no grupo de maior desempenho, em ambos os anos, existiu uma ocorrência maior de estudantes com pais de maior escolaridade, representados pelo eixo vertical, e uma variação maior na renda familiar, no eixo horizontal, sendo possível ter participantes até a faixa 17, a mais alta renda familiar. Enquanto isso, no grupo de menor desempenho, a faixa de renda fica muito mais localizada à esquerda do gráfico, ou seja, há uma maior incidência de pessoas com menos recursos financeiros e uma maior concentração de participantes com pais de escolaridade mais baixa, quando comparados ao grupo de maior desempenho. Também é possível observar que, nos dois casos, há uma recorrência muito maior na faixa 8, "Não sei", para os pais do que para as mães, o que pode representar famílias sustentadas pelas mães ou candidatos sem registro de pai.

SciELO Preprints - Este documento é um preprint e sua situação atual está disponível em: https://doi.org/10.1590/SciELOPreprints.18791

CONSIDERAÇÕES FINAIS

A pesquisa conduzida por esse trabalho evidenciou, por meio da aplicação da decomposição PARAFAC e das análises correlacionais, que a renda familiar e a escolaridade dos pais foram os principais fatores socioeconômicos associados à proficiência dos estudantes que realizaram o ENEM nos anos de 2022 e 2023. Os resultados dialogam com estudos clássicos e recentes, demonstrando que o contexto social continua a exercer grande influência no desempenho escolar dos estudantes.

Isso não significa dizer, porém, que as disparidades educacionais são causadas exclusivamente pela ausência de recursos ou pela baixa escolaridade dos pais, mas evidencia que apenas oferecer o acesso físico à escola não é suficiente para a garantia de um ensino e aprendizagem de qualidade para todos. Sendo assim, é de grande relevância o contínuo desenvolvimento e aplicação de políticas públicas que busquem mitigar o efeito que as disparidades socioeconômicas causam nos indivíduos no âmbito escolar.

No Brasil, alguns programas já desempenham um papel relevante no combate à desigualdade, mas seus impactos podem ser potencializados quando aliados a evidências que orientem intervenções mais específicas na área educacional. Dessa forma, ao utilizar uma abordagem estatística robusta aplicada em um banco de dados nacional e recente, a presente pesquisa se mostra relevante por fornecer evidências empíricas essenciais para fundamentar tanto a elaboração de novas políticas públicas, como no contribuir no aperfeiçoamento das ações que já existentes.

De fato, compreender como fatores extrínsecos à escola podem afetar o desempenho dos estudantes é fundamental para a elaboração de ações efetivas e inclusivas, pois saber onde reside o problema possibilita a implementação mais direcionada das políticas públicas aos estudantes em situação de vulnerabilidade, o que se torna um passo fundamental para impulsionar o rompimento da transmissão intergeracional da pobreza e, como consequência, passos para uma sociedade mais justa vão sendo dados através de um dos caminhos mais consistentes que temos de transformação, a educação.

Sugestões para a continuidade desse trabalho incluem a aplicação do método PARAFAC2 para análise dos dados sem a necessidade de se realizar uma amostragem no banco de dados. Além disso, a análise pode ser realizada por regiões do país, para se verificar quais padrões se repetem e diferem para cada região. Outro trabalho que pode ser derivado é a utilização deste método para análise comparativa de grupos antes e depois de serem assistidos por alguma política pública, para avaliação do seu impacto social.

REFERÊNCIAS

ALENCAR, Nataniele dos Santos; COSTA, Edward Martins; KHAN, Ahmad Saeed; MARIANO, Francisca Zilania; ARAÚJO, Jair Andrade de. Efeito da educação dos pais sobre o rendimento escolar dos filhos via dependência administrativa da escola. *Economia Aplicada*, São Paulo, v. 27, n. 1, p. 67–90, 2023. https://doi.org/10.11606/1980-5330/ea184088

ARAÚJO, J. M.; MARTIN, D. G.; FERREIRA, M. A. M.; FARIA, E. D. Fatores determinantes do desempenho educacional no Sudeste brasileiro. *Revista Eletrônica Gestão & Sociedade*, Belo Horizonte, v. 14, n. 38, p. 3507–3525, 2020. Disponível em: https://www.researchgate.net/publication/340738558. Acesso em: 20 ago. 2025.

BARROS, Ricardo Paes de; MENDONÇA, Rosane Silva Pinto de; SANTOS, Daniel Domingues dos; QUINTAES, Giovani. Determinantes do desempenho educacional no Brasil. *Pesquisa e Planejamento Econômico*, Rio de Janeiro, v. 31, n. 1, p. 1–42, 2001. Disponível em: http://repositorio.ipea.gov.br/handle/11058/5065. Acesso em: 12 mar. 2025.

BASSETTO, C. F. Aplicação do teste qui-quadrado sobre a associação entre proficiência em matemática e fatores socioeconômicos: uma abordagem com dados do SARESP. *Proceeding Series of the Brazilian*

SciELO Preprints - Este documento é um preprint e sua situação atual está disponível em: https://doi.org/10.1590/SciELOPreprints.18791

Society of Computational and Applied Mathematics, v. 8, n. 1, 2021. https://doi.org/10.5540/03.2021.008.01.0372

BOURDIEU, Pierre. *Cultural reproduction and social reproduction*. In: KARABEL, Jerome; HALSEY, A. H. (org.). Power and ideology in education. New York: Oxford University Press, 1977.

BRASIL. *Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira*. Microdados do Enem 2022. Brasília: INEP, 2023. Disponível em: https://www.gov.br/inep/pt-br/acesso-a-informacao/dados-abertos/microdados/enem. Acesso em: 3 jul. 2024.

BRASIL. *Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira*. Microdados do Enem 2023. Brasília: INEP, 2024. Disponível em: https://www.gov.br/inep/pt-br/acesso-a-informacao/dados-abertos/microdados/enem. Acesso em: 3 jul. 2024.

BRASIL. Ministério da Educação. *Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira*. Resultados 2023 [Apresentação]. Brasília: INEP, 2024. Disponível em: https://dowload.inep.gov.br/ideb/apresentacao_ideb_2023.pdf. Acesso em: 18 jul. 2025.

BRO, Rasmus. PARAFAC: tutorial and applications. *Chemometrics and Intelligent Laboratory Systems*, v. 38, n. 2, p. 149–171, 1997. Disponível em: https://www.researchgate.net/profile/Rasmus-Bro/publication/223735298_PARAFAC_tutorial_and_applications_Chemom_Intell_Lab_Syst/links/5dae9414a6fdccc99d92a30b/PARAFAC-tutorial-and-applications-Chemom-Intell-Lab-Syst.pdf. Acesso em: 10 fev. 2025.

BRO, Rasmus; KIERS, Henk A. L. A new efficient method for determining the number of components in PARAFAC models. *Journal of Chemometrics*, v. 17, n. 5, p. 274–286, 2003. Disponível em: https://three-mode.leidenuniv.nl/pdf/b/brokiers2003joch.pdf. Acesso em: 05 jan 2025.

CAPRARA, B. M. Condição de classe e desempenho educacional no Brasil. *Educação & Realidade*, v. 45, n. 4, p. e93008, 2020. http://dx.doi.org/10.1590/2175-623693008

COLEMAN, James S. et al. Equality of educational opportunity. Washington: US Government Printing Office, 1966.

DE BRITO, J. J. R. T. et al. A relação entre nível socioeconômico e proficiência em matemática de estudantes pernambucanos do 9º ano através da mineração de dados educacionais. Amazônia: Revista de Educação em Ciências e Matemáticas, v. 18, n. 41, p. 112–126, 2022. Disponível em: https://periodicos.ufpa.br/index.php/revistaamazonia/article/view/13547/9558#. Acesso em 15 maio 2025.

DE FARIAS SOUZA, W. P. S.; DE OLIVEIRA, V. R.; ANNEGUES, A. C. Background familiar e desempenho escolar: uma abordagem não paramétrica. *Pesquisa e Planejamento Econômico*, v. 48, n. 2, 2018. Disponível em: https://ppe.ipea.gov.br/index.php/ppe/article/view/1673/1270. Acesso em 20 jan 2025.

DIAMBO, F. P. T.; BRANCO, M. L. Relação família-escola: percepções de professores e pais/encarregados de educação numa escola pública em Angola. *Revista Iberoamericana de Educación*, v. 85, n. 2, p. 53–71, 2021. https://doi.org/10.35362/rie8523718

FEIJÓ, J. R.; FRANÇA, J. M. S.; PINHO, V. R. Desempenho dos estudantes ao final do ensino médio: mensurando a influência direta e indireta da educação dos pais. *Revista Brasileira de Economia*, v. 76, n. 1, p. 30–56, 2022. https://doi.org/10.5935/0034-7140.20220003

SciELO Preprints - Este documento é um preprint e sua situação atual está disponível em: https://doi.org/10.1590/SciELOPreprints.18791

GONÇALVES, F. O.; FRANÇA, M. T. A. Transmissão intergeracional de desigualdade e qualidade educacional: avaliando o sistema educacional brasileiro a partir do SAEB 2003. *Ensaio: Avaliação e Políticas Públicas em Educação*, v. 16, n. 61, p. 639–662, out. 2008. https://doi.org/10.1590/S0104-40362008000400009

HIPPOLYTO, Luzia de Queiroz; ANDRIOLA, Wagner Bandeira; NUNES, A. O. Uso da análise multivariada (PCA) e da álgebra multilinear (PARAFAC) para averiguar a associação entre o clima escolar e o desempenho acadêmico de alunos. *Educação & Linguagem*, v. 10, n. 1, p. 1–22, 2023. Disponível em: https://www.fvj.br/revista/wp-content/uploads/2023/07/1_REdLi.2023.1.pdf. Acesso em 05 de março de 2025.

KOLDA, Tamara G.; BADER, Brett W. Tensor decompositions and applications. *SIAM Review*, v. 51, n. 3, p. 455–500, 2009. Disponível em: https://www.kolda.net/publication/TensorReview.pdf. Acesso em 10 nov de 2024.

KOSSAIFI, Jean et al. TensorLy: tensor learning in Python. *Journal of Machine Learning Research*, v. 20, n. 18, p. 1–6, 2019. Disponível em: https://jmlr.org/papers/volume20/18-277/18-277.pdf. Acesso em 22 de maio de 2025.

KOSSAIFI, Jean. Selecting the number of components in PARAFAC models. TensorLy Documentation, 2021. Disponível em: https://tensorly.org/viz/stable/auto_examples/plot_selecting_aminoacids_components.html. Acesso em: 7 jun. 2024.

KOSSAIFI, Jean. TLViz — visualising and analysing component models. 2024. Disponível em: https://tensorly.org/viz/. Acesso em: 24 jan. 2025.

OECD. PISA 2022 results: factsheets – Brazil. Paris: OECD Publishing, 2023. Disponível em: https://www.oecd.org/pisa/publications/pisa-2022-results.htm. Acesso em: 1 jul. 2025.

PORTELA, Romualdo. Da universalização do ensino fundamental ao desafio da qualidade: uma análise histórica. Educação & Sociedade, Campinas, v. 28, n. 100, p. 661–690, 2007. https://doi.org/10.1590/S0101-73302007000300003

RODRIGUES, L. C. F. et al. Desempenho educacional e desigualdade social: um estudo de sua relação através da análise de componentes principais (ACP). In: CONEDU, 10., 2024, Campina Grande. Anais do X CONEDU. Campina Grande: Realize Editora, 2024. Disponível em: https://editorarealize.com.br/artigo/visualizar/112751. Acesso em: 5 maio 2025.

SENKEVICS, Adriano Souza; CARVALHAES, Flavio; RIBEIRO, Carlos Antonio Costa. Mérito ou berço? Origem social e desempenho no acesso ao ensino superior. Cadernos de Pesquisa, v. 52, p. e09528, 2022. https://doi.org/10.1590/198053149528

SILVA, T. E. V. da et al. Ferramentas de processamento e análise de informação aplicadas ao ensino de engenharia. In: XXXI Simpósio Brasileiro de Telecomunicações - SBrT2013. Fortaleza: Sociedade Brasileira de Telecomunicação, 2013. Disponível em: https://biblioteca.sbrt.org.br/articles/1430. Acesso em 4 jun 2025.

SILVA, T. E. V. da. Educometrics: from theory to application. 2017. Tese (Doutorado em Engenharia). Universidade Federal do Ceará, Fortaleza, 2017. Disponível em: http://repositorio.ufc.br/handle/riufc/28972. Acesso em: 10 jun. 2024.

SciELO Preprints - Este documento é um preprint e sua situação atual está disponível em: https://doi.org/10.1590/SciELOPreprints.18791

YOKOTA, T. Very basics of tensors with graphical notations: unfolding, calculations, and decompositions. arXiv preprint, arXiv:2411.16094, 2024. Disponível em: https://arxiv.org/abs/2411.16094. Acesso em: 10 jan. 2025.

ZAR, Jerrold H. *Biostatistical analysis*. 5. ed. Upper Saddle River: Pearson Prentice Hall, 2010.

Submetido: XX/XX/20XX
Preprint: XX/XX/20XX
Aprovado: XX/XX/20XX
Editor de seção:

DECLARAÇÃO DE FINANCIAMENTO

O presente trabalho foi realizado com apoio do Programa de Iniciação Científica Institucional da UFERSA (PICI/UFERSA), EDITAL 21/2024.

DECLARAÇÃO SOBRE DISPONIBILIDADE DE DADOS

Os dados utilizados neste estudo são públicos e estão disponíveis no portal do Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira (INEP). Os códigos desenvolvidos para o tratamento e análise dos dados não estão disponíveis publicamente, mas podem ser fornecidos pelos autores mediante solicitação.

DECLARAÇÃO DE USO DE INTELIGÊNCIA ARTIFICIAL

Durante a preparação deste trabalho, os autores utilizaram a ferramenta Napkin AI, na etapa metodológica para a elaboração da Figura 5. Além disso, os autores utilizaram a ferramenta ChatGPT, versão GPT-5.2, para revisão gramatical e ortográfica. Todo conteúdo produzido com o auxílio dessas ferramentas foi revisado criticamente pelos autores, que assumem total responsabilidade pelo conteúdo, pela precisão das informações e pela versão final do manuscrito.

CONTRIBUIÇÃO DE AUTORIA

Autor 1 – Curadoria de dados, análise formal, Metodologia, Redação.
Autor 2 – Conceitualização, Obtenção de financiamento, Supervisão, Validação, Escrita: revisão e edição.
Autor 2 – Supervisão, Validação, Escrita: revisão e edição.

DECLARAÇÃO DE CONFLITO DE INTERESSE

Os autores declaram que não há conflito de interesse com o presente artigo.

SciELO Preprints - Este documento é um preprint e sua situação atual está disponível em: https://doi.org/10.1590/SciELOPreprints.18791

APÊNDICE A - DESCRIÇÃO DAS QUESTÕES SOCIOECONÔMICAS ESTUDADAS

Questão	Descrição
Q00	Faixa etária
Q01	Sexo
Q02	Estado civil
Q03	Cor/raça
Q04	Nacionalidade
Q05	Situação de conclusão do Ensino Médio
Q06	Ano de Conclusão do Ensino Médio
Q07	Tipo de escola do Ensino Médio
Q08	Tipo de instituição que concluiu ou concluirá o Ensino Médio
Q09	Dependência administrativa (Escola)
Q10	Localização (Escola)
Q11	Situação de funcionamento (Escola)
Q12	Escolaridade do pai
Q13	Escolaridade da mãe
Q14	Ocupação do pai
Q15	Ocupação da mãe
Q16	Quantas pessoas moram na residência
Q17	Renda mensal da família
Q18	Em sua residência trabalha empregado(a) doméstico(a)?
Q19	Na sua residência tem banheiro?
Q20	Na sua residência tem quartos para dormir?
Q21	Na sua residência tem carro?
Q22	Na sua residência tem motocicleta?
Q23	Na sua residência tem geladeira?
Q24	Na sua residência tem freezer (independente ou segunda porta da geladeira)?
Q25	Na sua residência tem máquina de lavar roupa?
Q26	Na sua residência tem máquina de secar roupa (independente ou em conjunto com a máquina de lavar roupa)?
Q27	Na sua residência tem forno micro-ondas?
Q28	Na sua residência tem máquina de lavar louça?
Q29	Na sua residência tem aspirador de pó?
Q30	Na sua residência tem televisão em cores?
Q31	Na sua residência tem aparelho de DVD?
Q32	Na sua residência tem TV por assinatura?
Q33	Na sua residência tem telefone celular?
Q34	Na sua residência tem telefone fixo?
Q35	Na sua residência tem computador?
Q36	Na sua residência tem acesso à Internet?

Fonte: Elaborado pelos autores com base no Dicionário de Microdados do ENEM (INEP, 2025).

Este preprint foi submetido sob as seguintes condições:

- Os autores declaram que os necessários Termos de Consentimento Livre e Esclarecido de participantes ou pacientes na pesquisa foram obtidos e estão descritos no manuscrito, quando aplicável.
- Os autores declaram que a elaboração do manuscrito seguiu as normas éticas de comunicação científica.
- Os autores declaram que estão cientes que são os únicos responsáveis pelo conteúdo do preprint e que o depósito no SciELO Preprints não significa nenhum compromisso de parte do SciELO, exceto sua preservação e disseminação.
- Os autores declaram que os dados, aplicativos e outros conteúdos subjacentes ao manuscrito estão referenciados.
- O manuscrito depositado está no formato PDF.
- Os autores declaram que a pesquisa que deu origem ao manuscrito seguiu as boas práticas éticas e que as necessárias aprovações de comitês de ética de pesquisa, quando aplicável, estão descritas no manuscrito.
- Os autores declaram que uma vez que um manuscrito é postado no servidor SciELO Preprints, o mesmo só poderá ser retirado mediante pedido à Secretaria Editorial do SciELO Preprints, que afixará um aviso de retratação no seu lugar.
- Os autores concordam que o manuscrito aprovado será disponibilizado sob licença Creative Commons CC-BY.
- O autor submissor declara que as contribuições de todos os autores e declaração de conflito de interesses estão incluídas de maneira explícita e em seções específicas do manuscrito.
- Os autores declaram que o manuscrito não foi depositado e/ou disponibilizado previamente em outro servidor de preprints ou publicado em um periódico.
- Caso o manuscrito esteja em processo de avaliação ou sendo preparado para publicação mas ainda não publicado por um periódico, os autores declaram que receberam autorização do periódico para realizar este depósito.
- O autor submissor declara que todos os autores do manuscrito concordam com a submissão ao SciELO Preprints.