

doi 10.46943/X.CONEDU.2024.GT07.008

# MULHERES TÊM COMPETÊNCIA PARA AS EXATAS?

Marília Magalhães Maia<sup>1</sup>  
Luiza Helena Félix de Andrade<sup>2</sup>  
Silvio Roberto Fernandes<sup>3</sup>  
Lara Carla Freitas Rodrigues<sup>4</sup>

## RESUMO

A população brasileira é composta majoritariamente pelo público feminino, segundo o Censo 2022, e os percentuais são ainda maiores com relação a parcela que cursa o ensino superior. Entretanto, há uma desproporção quanto a distribuição dessa população com relação aos cursos, havendo uma concentração do público feminino nos cursos na área da saúde e educação, com percentuais superiores 60%, e uma escassez na área das exatas, com o máximo de 30%, segundo o Censo da Educação Superior de 2022. Assim, compreendendo a Matemática como a disciplina de maior relevância para o ingresso e desempenho nos cursos das exatas, este trabalho compara o desempenho dos inscritos no Exame Nacional do Ensino Médio (ENEM), autodeclarados do sexo feminino e masculino, nas provas de Matemática e suas Tecnologias. Através de técnicas de ciência de dados, incluindo Análise de Componente Principal (ACP), aplicadas sobre dados dos candidatos as habilidades avaliadas pelas provas aplicadas no período de 2009 a 2022. A partir desta análise observou-se os maiores percentuais de acerto majoritariamente masculinos. Este resultado é reafirmado ao calcular a média das proficiências de ambas as popula-

1 Mestranda em Engenharia de Teleinformática da Universidade Federal do Ceará - UFC, [mariliamaia@alu.ufc.br](mailto:mariliamaia@alu.ufc.br);

2 Doutora em Engenharia de Teleinformática da Universidade Federal do Ceará - UFC, [luizafelix@ufersa.edu.br](mailto:luizafelix@ufersa.edu.br);

3 Doutor em Sistemas e Computação da Universidade Federal do Rio Grande do Norte - UFRN, [silvio@ufersa.edu.br](mailto:silvio@ufersa.edu.br);

4 Graduada pelo Curso de Ciência e Tecnologia da Universidade Federal Rural do Semi-Árido - UFRSA, [lara.rodrigues@alunos.ufersa.edu.br](mailto:lara.rodrigues@alunos.ufersa.edu.br);

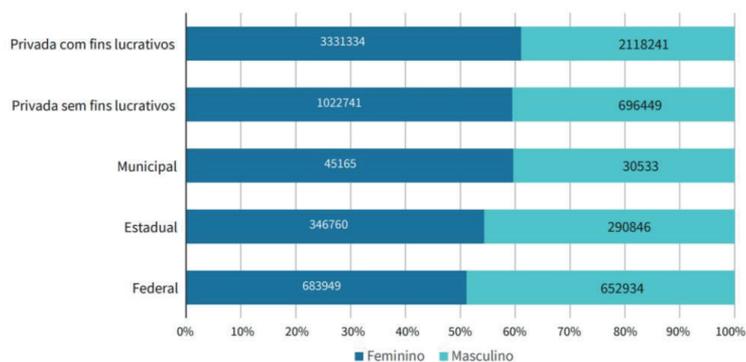
ções, para cada ano, o que demonstrou que apesar das diferenças serem inferiores a 6%, as médias masculinas são maiores do que as femininas. Entretanto, quando se observa o desempenho por habilidade não há predominância de nenhum dos sexos e nem regularidade, além de diferenças percentuais máximas inferiores a 0,3%. Dessa forma, demonstra-se que o desempenho nas provas de matemática não é o único fator determinante para participação feminina nos cursos de exatas.

**Palavras-chave:** Exame Nacional do Ensino Médio, Desempenho, Análise de Componente Principal, Matemática, Público feminino.

## INTRODUÇÃO

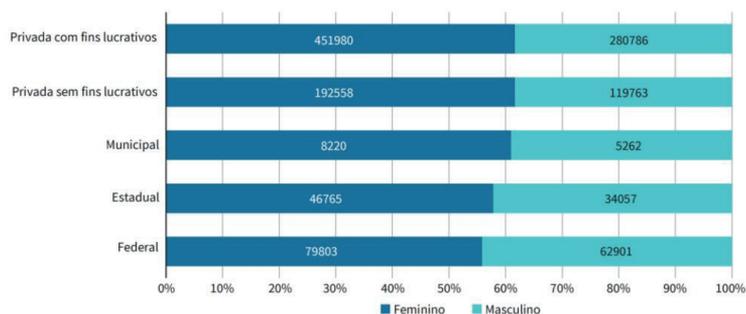
O Censo populacional mais recente realizado, em 2022, informa que o a população brasileira soma mais de 203 milhões de pessoas, dos quais 51,5% são do sexo feminino (Brasil, 2023). Este censo ainda apresenta que entre os jovens entre 15 e 29 anos, o sexo masculino possui maior percentual de participação, sendo está a faixa etária predominante entre os ingressantes nos cursos de graduação, segundo (INEP, 2023). Entretanto, a distribuição por sexo nesta população apresenta uma inversão de predominância, como é observado nos gráficos da Figura 1 e da Figura 2. Encontrados no Censo da Educação Superior de 2022 (INEP, 2023), o gráfico da Figura 1 indica os percentuais, por sexo, do número de alunos ingressantes no em graduações no ano de 2022. Enquanto o gráfico da Figura 2 apresenta o mesmo panorama sobre os concluintes.

**Figura 1** - Número de alunos que ingressaram na graduação em 2022 por sexo, segundo a categoria administrativa.



**Fonte:** Elaborado por Deed/Inep com base em dados do Censo da Educação Superior.

**Figura 2** - Percentual de alunos concluintes de graduação por categoria administrativa, no ano de 2022, segundo o sexo.



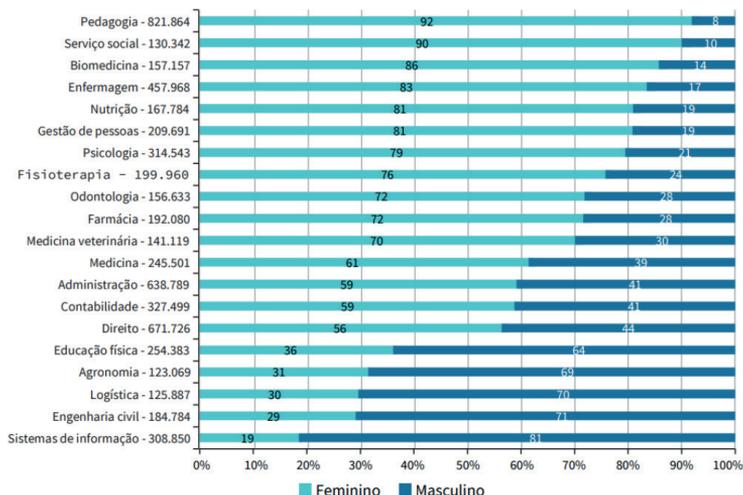
**Fonte:** Elaborado por Deed/Inep com base em dados do Censo da Educação Superior.)

Ambos os gráficos demonstram a predominância atual do sexo feminino no ensino superior, independente da modalidade administrativa. Estes valores indicam o sexo feminino como aquele que mais buscam a progressão nos estudos, avançando do ensino básico para o ensino superior, a nível de graduação. No entanto, este cenário não foi sempre assim. As transformações sociais ocorridas no pós-guerra, na década de 1960, permitiram que as mulheres passassem a ocupar novos espaços fora do ambiente privado e doméstico, ingressando de maneira mais expressiva no mercado de trabalho e nas universidades. Esse período marcou um avanço significativo no pensamento feminista, que acompanhava as novas configurações sociais e reivindicava igualdade de direitos, especialmente em relação à educação e ao emprego (Gonçalves Carneiro et al., 2019; Pinto et al., 2017).

No Brasil, o hiato educacional de gênero foi superado na década de 1990, com as taxas de indivíduos com curso superior completo entre homens e mulheres mostravam uma distribuição bastante similar, o que reflete um avanço na igualdade de acesso à educação superior. Esse cenário é apresentado em (A. Ricoldi & A. Artes, 2016), que destacaram a reversão do gap de gênero, resultando em uma presença cada vez maior de mulheres no ensino superior. Essa mudança contribuiu para o crescimento da participação feminina em diversas áreas do conhecimento e no mercado de trabalho, consolidando uma maior igualdade de oportunidades educacionais no país.

Entretanto, como também apresentado em (INEP, 2023), a desproporção quanto a distribuição do sexo feminino nos cursos de graduação ainda existe e é elevada. O gráfico apresentado na Figura 3, apresenta os percentuais de participação dos sexos nos 20 maiores cursos de graduação, em número de matrículas em 2022. Verifica-se que o público feminino compoem a maioria em 15 dos 20 maiores cursos de graduação, demonstrando mais uma vez a maior iniciativa por parte do sexo para a obtenção do avanço educacional. Mas observa-se, também, que os cursos os quais a população feminina encontra-se como minoria são cursos da área das exatas, contrapondo a sua participação superlevada nos cursos da área da saúde e educação. Diante disso, questiona-se os motivos que condicionam essa discrepância. Seria a questão de o sexo feminino não possuir bom desempenho nas disciplinas atreladas ao curso e por isso terem menor potencial de competição pelas vagas dos cursos de exatas?

**Figura 3** – 20 maiores cursos em número de matrículas de graduação e os respectivos percentuais de participação por sexo, em 2022



**Fonte:** Elaborado por Deed/Inep com base em dados do Censo da Educação Superior)

A importância de equalizar as oportunidades e direitos entre os sexos é demonstrada pela sua presença nos Objetivos de Desenvolvimento Sustentáveis, pela Organização das Nações Unidas (ONU), ocupando o quinto objetivo, nomeado por “Igualdade de Gênero”. Assim, apesar de a escolaridade não garantir igualdade entre os gêneros, é uma ferramenta de alta relevância e potencial de mudança social (A. Ricoldi & A. Artes, 2016; Melo, 2013). Visando isto, o Programa Futuras Cientistas, lançado pelo Governo brasileiro, teve início das suas atividades em abril de 2024, estimulando alunas do ensino público às áreas das exatas, tanto no ingresso nos cursos de graduação, quando ao desenvolvimento mais completo daquelas que já estão nos cursos (Machado, 2023). A criação deste projeto demonstra a preocupação com a discrepância da participação feminina neste cenário e é uma medida de sanar alguns dos possíveis motivos deste fato.

Nesta etapa é válido ressaltar a diferença entre sexo e gênero. Sua diferenciação se faz necessária para que seja empregado corretamente o sentido do termo, não desconfigurando a discussão abordada. Segundo os conceitos apresentados em (Duarte, 2021; Melo, 2013; Pinto et al., 2017), podem ser descritos os seguintes conceitos. O termo “sexo” pertence ao campo da biologia e se refere às características biológicas, genéticas, anatômicas e fisiológicas que distinguem homens e mulheres. Essas características incluem, por exemplo, os cromossomos (XY para homens e XX para mulheres), os órgãos reprodutivos,

bem como as diferenças hormonais que influenciam o desenvolvimento físico de cada indivíduo. Já o conceito de “gênero” está inserido no contexto social e cultural. Ele envolve o conjunto de papéis, comportamentos, expectativas e responsabilidades que são socialmente atribuídos a homens e mulheres. Esses papéis são aprendidos e incorporados ao longo da vida, através de processos de socialização que variam de acordo com o tempo, o local e a cultura. A construção do gênero ajuda a moldar a percepção que cada pessoa tem de si mesma, influenciando diretamente a compreensão do próprio corpo e a formação da identidade, assim como a maneira de se relacionar com o outro. Portanto, enquanto o sexo é um dado biológico, com classificação binária, masculino e feminino, o gênero é uma construção social que reflete as normas e valores de cada sociedade. Neste trabalho será abordada uma discussão relacionada a comparação de desempenho e habilidade entre os sexos, visto que é esta a informação presente nos formulários de inscrição do ENEM.

Sobre os cursos das exatas, existe uma forte característica em comum, a relação da maioria das disciplinas com a matemática. Essa relação ocorre diante da construção do raciocínio lógico e o fato de que o domínio do conhecimento matemático facilita a compreensão de conteúdos mais complexos em disciplinas como física, química, eletricidade e informática (Crystini & Sehn, 2017). Em (Crystini & Sehn, 2017) destaca-se que muitos alunos que apresentam dificuldades em matemática também encontram problemas em outras disciplinas de exatas, indicando que a matemática funciona como uma base para o desenvolvimento intelectual nessas áreas.

A individualização da Matemática e suas tecnologias dentre as demais disciplinas conhecidas como exatas, no Exame Nacional do Ensino Médio (ENEM), demonstra sua importância. O ENEM foi desenvolvido com o intuito de avaliar o desempenho dos alunos ao fim da escolaridade básica, a nível nacional. Entretanto, passou a ser, também, a principal ferramenta de acesso ao ensino superior, no país, permitindo que as proficiências dos alunos que o prestam sejam utilizadas em diversos processos seletivos de Instituições de Ensino Superior (IES), sendo elas públicas, privadas e internacionais. O exame é composto por 5 provas: Matemática e suas Tecnologias, Linguagens e Códigos, Ciências da Natureza, Ciências Humanas e Redação, totalizando em 180 questões e um texto dissertativo-argumentativo de até 30 linhas, dividindo sua aplicação em dois dias de prova por ano (da Silveira et al., 2015). A prova de Matemática e suas Tecnologias possui 45 questões, com 5 alternativas, cada, sendo cada questão

elaborada de modo que seja capaz de avaliar o conhecimento sobre 1 das 30 habilidades que compõem a Matriz de Referência do ENEM, para a prova em questão. Esta formatação atual passou a ser implementada com a reestruturação realizada em 2009, que teve como um de suas mudanças a aplicação da Teoria de Resposta ao Item (TRI) (Brasil, 2020).

A TRI é um modelo estatístico amplamente utilizado para medir traços latentes, como habilidades, atitudes ou aptidões de indivíduos, com base nas respostas fornecidas em testes ou avaliações. Ela avalia a probabilidade de um respondente acertar um item considerando tanto a habilidade do indivíduo quanto as características do item, como dificuldade, discriminação. Foi desenvolvida para superar limitações da Teoria Clássica de Medidas, que assume que todos os itens têm a mesma dificuldade e que as avaliações são igualmente válidas para todos os respondentes. O modelo da TRI aplicado no ENEM é o modelo de três parâmetros, que inclui a probabilidade de acerto casual, além dos parâmetros de dificuldade e discriminação (Fujii, et al., 2021). Sua implementação é de suma importância, pois permite comparar o desempenho de estudantes com diferentes níveis de habilidade de forma mais precisa e adaptar os testes a esses níveis. O uso da TRI também permite a equalização de provas, ou seja, a possibilidade de comparar resultados de diferentes testes ao longo do tempo, garantindo uma medição mais consistente de habilidades (E. A. C. et al., 2009; Klein, 2013). Para cada item (questão) da prova do ENEM, além da habilidade associada, também há 3 parâmetros (NU\_PARAM\_A, NU\_PARAM\_B e NU\_PARAM\_C) que são calculados após a aplicação da prova de cada ano e são usados no cálculo de proficiências dos participantes. NU\_PARAM\_A é definido como o parâmetro de discriminação, o qual diferencia os participantes que dominam dos que não dominam a habilidade avaliada. NU\_PARAM\_B é o parâmetro de dificuldade, o qual está associado à dificuldade do item, sendo que quanto maior seu valor, mais difícil é o item. E NU\_PARAM\_C é o parâmetro de acerto ao acaso, o qual indica a probabilidade de um participante acertar o item não dominando a habilidade exigida.

A inclusão do “acerto ao acaso” na avaliação TRI do ENEM permite que o modelo diferencie entre um acerto devido à habilidade e um acerto por sorte, evitando que indivíduos com baixa proficiência recebam uma avaliação inflada. Assim, no modelo de três parâmetros, a fórmula é ajustada para que o acerto casual não aumente indevidamente a estimativa de proficiência, mantendo assim a precisão da avaliação. Sem o ajuste para o acerto casual, estudantes com baixo

desempenho, que acertam itens ao acaso, poderiam ter suas habilidades superestimadas (Crystini & Sehn, 2017; Fujii et al., 2021; Lopes & Vendramini, 2013).

Visto as principais características e vantagens da metodologia de avaliação do ENEM, baseado no TRI, a questão de pesquisa abordada neste trabalho é utilizar um método computacional que permita comparar o desempenho dos candidatos no ENEM desde 2009 até 2022 e avaliar se o público feminino tem competência para áreas de exatas. Para isso foram aplicadas diversas técnicas de ciência de dados, incluindo ACP (Análise da Componente Principal).

A ACP é uma técnica de processamento de dados não supervisionado, ou seja, não utiliza rótulos de classificação dos dados. Esta técnica é amplamente utilizada em diversos campos de estudo, devido a sua simplicidade e variedade de funções que pode ser aplicado, entre elas a redução da dimensionalidade do conjunto de dados ou a identificação de padrões (De Souza & Poppi, 2012). Ele transforma um conjunto de variáveis correlacionadas em um novo conjunto de variáveis não correlacionadas chamadas “componentes principais”. Essas componentes capturam a maior parte da variação nos dados, permitindo simplificar a análise sem perder informações relevantes. Isso torna o processamento de dados mais eficiente e eficaz, permitindo que padrões importantes emergem enquanto o ruído e a redundância são minimizados (Hongyu et al., 2015).

Diante do exposto, o presente trabalho tem como objetivo avaliar o desempenho de candidatos do sexo feminino e masculino nas provas de Matemática e suas tecnologias, do ENEM, aplicadas entre 2009 e 2022. Com o intuito de averiguar se o desempenho nas provas pode ser um dos motivos para a escassez do sexo feminino nos cursos de graduação de exatas. Como ferramenta de análise, será aplicado a ACP para o reconhecimento de padrões, além da análise estatística sobre as informações sobre as habilidades abordadas nas provas, em conjunto com acertos dos candidatos, e os parâmetros do TRI para cada item, fornecidos nos microdados do ENEM, disponíveis em (Brasil, 2024).

## **METODOLOGIA**

Para alcançar os objetivos deste trabalho, definiu-se os 14 conjuntos de dados disponibilizados pelo INEP, referentes as aplicações das provas do Enem no período de 2009 a 2022. Estes, nomeados de microdados do Enem, são compostos por dois conjuntos de dados, um relacionado aos alunos inscritos no exame, preservando o anonimato, e o outro relacionado as provas, com infor-

mações sobre cada item avaliado. Assim, os dados foram preparados unindo os dois conjuntos, passando pelo processo de refinamento, amostragem e pré-processamento, antes da sua aplicação na análise computacional, para a então obtenção dos resultados que serão discutidos neste trabalho.

No processo de refinamento, foi aplicado um filtro sobre o conjunto de dados referente aos alunos, selecionando apenas os alunos presentes nas 5 provas aplicadas, em cada ano, para garantir a existência de suas respostas as questões e proficiências, a cor da prova de Matemática e suas tecnologias e suas respectivas respostas. Restringimos nossa análise sobre as provas de Matemática e suas tecnologias, tendo em vista que a matemática é a área do conhecimento mais relacionada aos cursos de exatas. Entretanto, mesmo após a filtragem, o volume é muito grande e se fez necessária a amostragem, que foi aplicada com nível de confiança de 99% e erro amostral de 1%, sendo realizada uma seleção aleatória, sobre o conjunto de dados de cada ano.

Após a amostragem, foram vinculadas as informações do conjunto de dados sobre a prova de Matemática e suas Tecnologias (MT) e o conjunto obtido até esta etapa do processo. Do conjunto referente a prova, foram coletados os gabaritos correspondentes a cada cor de prova e a respectiva habilidade avaliada em cada item (questão da prova). A aplicação dos gabaritos sobre as respostas dos alunos foi realizada de forma que fosse armazenada a informação de erro e acerto de cada questão por cada aluno selecionado. Assim, foi formada uma tabela composta por 45 colunas, correspondendo as questões da prova, e cada linha a um dos alunos incluídos na amostragem, de modo atribuiu-se 0 para cada questão errada e 1 para cada questão acertada pelo aluno.

Entretanto, o objetivo deste trabalho está atrelado avaliar a competência do público feminino nas provas MT, levando em consideração, para cada ano, a média de certo por item, o percentual de acerto nas habilidades avaliadas, o percentual de acerto dos itens em cada um dos 3 parâmetros do TRI e a proficiência média calculada no final pelo INEP.

Para avaliar o desempenho dos candidatos em relação as habilidades, utilizamos o método ACP, o qual necessitou de um pré-processamento dos dados. É importante destacar que cada questão é relacionada apenas a apenas uma habilidade, mas uma data habilidade pode ser avaliada em mais de uma questão da prova. O novo conjunto de dados passou a ser formado por 30 colunas, correspondendo as habilidades da matriz curricular, e cada linha a um aluno. Para cada aluno aplica-se uma média sobre os acertos e erros das questões relacionadas

a uma mesma habilidade. Assim, por exemplo, para uma habilidade que possui, em um determinado ano, 4 questões relacionadas, caso o aluno tenha acertado 4 das 4, então será preenchido pelo valor 1 na coluna que corresponder àquela habilidade. Caso ele tenha acertado apenas 3 das 4, será preenchido por 0,75 e assim por diante, até o caso do erro das 4 questões, impondo o preenchimento pelo valor 0. A Figura 4 apresenta um exemplo desse conjunto de dados final.

**Figura 4** – Exemplo dos conjuntos de dados implementados no ACP.

HAB1	HAB2	HAB3	HAB4	HAB5	HAB6	HAB7	HAB8	HAB9	HAB10	...	HAB21	HAB22	HAB23	HAB24	HAB25	HAB26	HAB27	HAB28	HAB29	HAB30
0.0	0.0	0.0	1.0	0.0	0.000000	0.0	0.0	0.0	0.0	...	0.0	1.0	1.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
0.5	0.5	0.5	1.0	0.0	0.000000	0.5	0.0	0.0	0.0	...	0.0	0.0	1.0	0.0	0.0	0.0	1.0	0.0	0.0	1.0
0.5	0.5	0.0	0.0	0.0	0.000000	0.0	0.0	0.0	1.0	...	1.0	0.0	0.0	0.0	0.5	0.5	0.5	0.0	0.0	0.0
0.5	1.0	0.0	0.0	0.0	0.000000	0.0	0.0	0.0	0.0	...	0.0	0.0	1.0	0.0	0.0	0.5	0.0	0.0	0.0	1.0
0.5	0.5	0.5	0.0	1.0	0.666667	1.0	1.0	0.5	1.0	...	1.0	1.0	1.0	1.0	0.0	0.0	0.5	0.0	0.0	0.0
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
0.5	0.0	0.5	0.0	0.0	0.666667	0.0	0.0	0.5	0.0	...	0.0	0.0	0.0	1.0	0.5	0.5	0.0	0.0	0.0	0.0
0.0	0.5	0.5	1.0	0.0	0.000000	0.0	1.0	0.5	0.0	...	0.0	0.0	0.0	0.5	0.5	0.5	1.0	0.5	0.0	0.0
0.0	0.5	0.5	0.0	0.0	0.333333	0.5	0.0	0.0	0.0	...	0.0	0.0	0.0	0.0	0.5	0.0	0.5	0.5	0.0	1.0
0.0	0.5	0.5	0.0	0.0	0.333333	0.0	1.0	0.5	0.0	...	0.0	0.0	0.0	0.5	0.0	0.5	0.0	0.0	0.0	0.0
0.5	0.0	0.0	1.0	0.0	0.666667	0.5	0.0	0.0	0.0	...	0.0	0.0	0.0	0.5	0.0	0.5	0.5	0.0	0.0	0.0

Em seguida, o algoritmo ACP foi usado nos 14 conjuntos de dados individualmente por meio de uma implementação na linguagem de programação Python. O algoritmo ACP recebe como parâmetro a quantidade de componentes principais para o qual o conjunto de variáveis da entrada deve ser reduzida e retorna a variância de cada uma delas, de modo que o somatório delas indicam o percentual de relevância para a representação inicial. Assim, como estamos interessados em mensurar a importância de cada habilidade, utilizamos como parâmetro para o ACP a mesma quantidade de habilidades (30). Após o processamento, as habilidades são colocadas em ordem decrescente em termos de sua relevância retornada pelo algoritmo. A partir disso, foram observadas as ordenações na composição da primeira componente principal de cada ano, sendo coletadas as habilidades que se apresentavam mais frequentemente entre as 10 primeiras, em todos os anos.

Por fim, dividindo os candidatos em dois grupos, por sexo, e o percentual de acertos da habilidade dos dois grupos. Além disso, para cada grupo, foi realizada uma média das proficiências na prova de Matemática e suas tecnologias, e, a partir dos parâmetros do TRI, encontramos as questões com maior grau de dificuldades, com maior discriminação e com maior probabilidade de acerto ao acaso, por ano, e os respectivos percentuais de acerto destas questões em cada grupo. Para isso, foi necessário relacionar o código da prova do candidato para

encontrar a posição relativa as questões com valores máximos dos parâmetros do TRI, e só então verificar se houve acerto deste item.

Todos os processos anteriores foram realizados, separadamente, para cada ano analisado. Os resultados encontrados serão apresentados na próxima seção e servirão de base para as discussões seguintes..

## RESULTADOS E DISCUSSÃO

O Enem é a maior avaliação educacional do país e possui um grande volume de inscritos. Diante disso, a aplicação do processo de filtragem sobre os microdados referentes aos alunos pode ser vista na Tabela 1, juntamente com a apresentação da porcentagem de alunos do sexo feminino e masculino, em cada ano analisado.

**Tabela 1** – Tabela com indicação das quantidades de alunos inscritos, quantidade absoluta e percentual de inscritos do sexo feminino e quantidade de alunos selecionados pelo filtro.

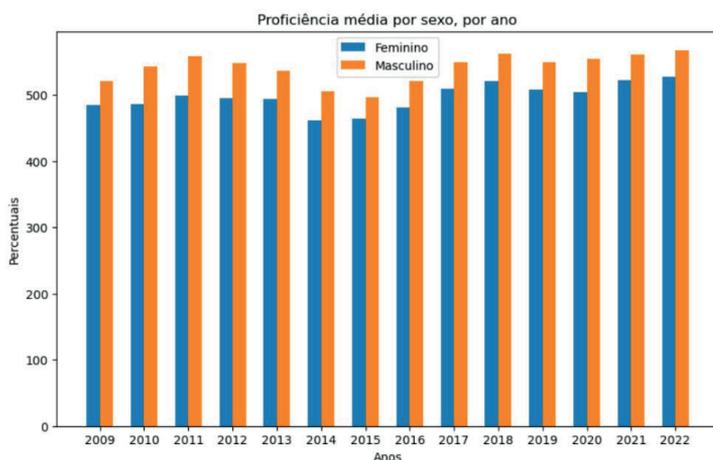
Ano	Inscritos	Feminino	Percentual Feminino	Selecionados
2009	4148720	2482273	59,83%	2420738
2010	4611614	2749651	59,62%	3234222
2011	5366931	3187590	59,39%	3853313
2012	5791065	3416410	58,99%	4079465
2013	7173563	4185354	58,34%	5006981
2014	8722248	5069514	58,12%	5945926
2015	7746427	4460448	57,58%	5598037
2016	8627179	4982539	57,75%	5663011
2017	6731278	3946714	58,63%	4423761
2018	5513733	3257698	59,08%	3891167
2019	5095171	3031760	59,50%	3699333
2020	5783109	3468805	59,98%	2499159
2021	3389832	2090526	61,67%	2136539
2022	3476105	2120519	61,00%	2311963

A Tabela 1 demonstra que o público feminino compõe a maioria dos inscritos no exame, todos os anos, com participação mínima de 57%. Este fato surpreende pelo fato de que na educação básica o sexo feminino não é maioria, atingindo o marco de 50,9% apenas sobre os alunos do ensino médio, segundo o Censo Escolar de 2023. O que demonstra que apesar de haver, praticamente,

uma equivalência na população escolar entre os sexos, o público feminino apresenta maior presença na prestação do exame, o qual é conhecido como “o principal mecanismo de acesso ao ensino superior”.

Sendo o desempenho no exame medido pelas proficiências calculadas através da TRI, analisa-se a média das proficiências na prova de Matemática e suas tecnologias, ou seja, a nota final divulgada pelo INEP que será usada nos processos seletivos das IES. Na Figura 8 são apresentadas as médias de proficiência masculinas e femininas ao longo dos anos, considerando todos os participantes inscritos que foram selecionados pela filtragem, sem amostragem.

**Figura 8** – Valores médios das proficiências na prova de Matemática e suas tecnologias de cada ano, por sexo.



A Figura 8 demonstra que as proficiências médias do público masculino são maiores do que as do público feminino em todos os anos. Entretanto, apesar de ser visível a diferença entre as médias, a Tabela 3, apresenta os valores das médias e as diferenças percentuais entre masculino e feminino.

**Tabela 3** – Proficiências médias por sexo, por ano.

Ano	2009	2010	2011	2012	2013	2014	2015	2016	2017	2018	2019	2020	2021	2022
<b>Média Feminina</b>	484,359	486,495	499,668	495,793	493,833	461,931	464,621	481,205	509,838	521,053	507,751	504,652	521,669	526,898
<b>Média Masculina</b>	520,566	542,557	558,735	548,3	537,125	505,528	496,857	521,333	549,122	562,444	549	554,671	561,466	567,819
<b>Percentual da diferença</b>	3,6207	5,6062	5,9067	5,2507	4,3292	4,3597	3,2236	4,0128	3,9284	4,1391	4,1249	5,0019	3,9797	4,0921

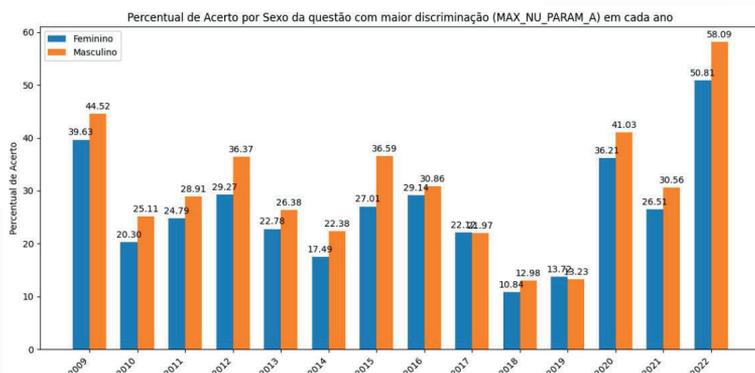
Os valores expostos na Tabela 3 demonstram que as diferenças são inferiores a 6%, ou seja, 60 pontos, sendo a diferença média próxima a 4,5%. Ainda assim, realizamos um teste t para amostras independentes, um teste estatístico que compara a proficiência dos dois grupos. Neste teste, a hipótese nula ( $H_0$ ) supõe que não há diferença entre as médias entre os dois grupos. E a hipótese alternativa ( $H_1$ ) supõe que há uma diferença entre as médias (diferença  $\neq 0$ ). Realizamos o teste por ano (comparações independentes) e o teste global (análise combinada) e em todos eles *valor-p* foi zero, o que significa rejeição de  $H_0$ , e como os valores de *t-estatístico* variaram entre -21 e -31, indicam que a média de um grupo está distante da média do outro, em múltiplos desvios-padrão.

Contudo, para uma análise mais profunda, avaliamos o desempenho em relação as habilidades as quais os itens estão relacionados e em relação às questões com valores máximos em cada parâmetros do TRI em cada ano. Para tanto, foi necessário realizar uma amostragem sobre o conjunto de alunos selecionados, com nível de confiança de 99% e erro amostral de 1% resulta no mesmo valor de 16592 como o valor adequado de representação, os quais foram usados para os demais resultados apresentados a seguir.

Desta forma, primeiro analisamos o desempenho dos públicos masculino e feminino considerando as questões com valores máximos em cada parâmetros do TRI em cada ano, como pode ser visto a seguir.

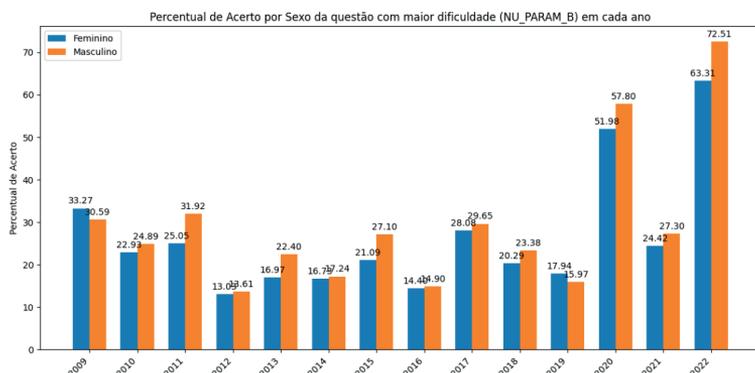
A Figura 9 apresenta o percentual de acerto da questão com maior valor de discriminação de cada ano, ou seja, o parâmetro NU\_PARAM\_A do TRI, de forma comparativa entre os participantes masculino e feminino. Percebe-se que em todos os anos, o masculino teve maior percentual, com diferença máxima de 7,28 pontos percentuais em 2022 e mínima de 0,15 em 2017.

**Figura 9** – Percentual de acertos masculino e feminino da questão com maior discriminação em cada ano



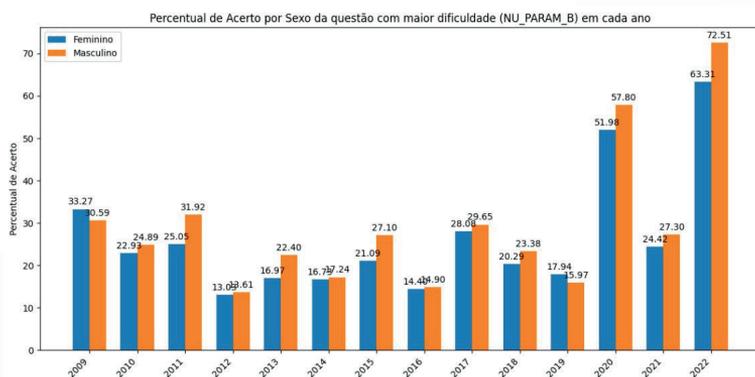
A Figura 10 apresenta a comparação entre os dois públicos em relação a questão com maior dificuldade em cada ano e comportamento é muito semelhante a anterior. Percebemos que em todos os anos o público masculino apresenta médias superiores, com maior diferença em 2022, com 9,2 pontos percentuais.

**Figura 10** - Percentual de acertos masculino e feminino da questão com maior dificuldade em cada ano



A Figura 11 apresenta as médias de acerto em relação a questão de maior probabilidade de acerto ao acaso, em cada ano para os dois públicos. O comportamento é semelhante aos dois gráficos anteriores, ou seja, o público masculino com superioridade em todos os anos. No entanto, este parâmetro revela um fator punitivo no cálculo da proficiência, ou seja, neste caso, quanto menor melhor. Isso revela que o público feminino apresenta mais coerência nas respostas, de forma geral. Aqui, a maior diferença foi em 2010 com 14,82 percentuais.

**Figura 11** - Percentual de acertos masculino e feminino da questão com maior dificuldade de acerto ao acaso em cada ano



A segunda análise foi relacionada ao desempenho dos dois grupos em relação às habilidades. Para tanto, os 14 conjuntos de dados (relativo aos anos de 2009 a 2022), foram utilizados, separadamente, no algoritmo ACP com o intuito de realizar um reconhecimento de padrões, identificando quais das variáveis que compõem o conjunto são as mais importantes. Assim, ao final desse processo, obtêm-se uma lista das habilidades ordenadas pela relevância delas na avaliação da prova de Matemática e suas tecnologias, de cada ano. Para exemplificar, a Figura 5 apresenta o gráfico resultante referente aos dados de 2019. Os demais gráficos estão disponíveis em uma pasta de acesso livre através do link<sup>5</sup>.

**Figura 1** – Gráfico com a ordem de importância na composição da primeira Componente Principal, sobre os dados de 2019.



Observa-se que ocorre uma redução acentuada no potencial de interrelação entre as habilidades. Considerando que, ao variar a composição dos dados implementados, visto a aleatoriedade no processo de amostragem, ocorre uma pequena variação na ordem das habilidades mais importantes, avaliar as 10 primeiras, dentre as 30 habilidades apresentada, permite que seja considerada esta variação. Assim, a Tabela 2 apresenta a frequência das habilidades no grupo das 10 principais, de cada ano.

Ao analisar a Tabela 2, constata-se que a habilidade 1 é aquela que se apresenta, em todos os anos analisados, entre as 10 principais habilidades da prova. Então, pode-se indicá-la como aquela que tem maior potencial de influência sobre as demais. Ou seja, considerando a matemática como uma ciência de conhecimentos acumulativos (Fujii et al., 2021) e que a habilidade 1 indica a

<sup>5</sup> <https://drive.google.com/drive/folders/12cpYaENLNvXAKZdff5eOBhnmGemheBUK?usp=sharing>

capacidade de reconhecimento dos números nas suas variadas representações e contextos, tê-la bem desenvolvida interfere positivamente para o desenvolvimento das demais.

**Tabela 2** – Tabela com a frequência das habilidades entre as 10 primeiras habilidades que compõem as primeiras

Frequência	Habilidade	Matriz de referência do ENEM
14	1	Reconhecer, no contexto social, diferentes significados e representações dos números e operações - naturais, inteiros, racionais ou reais.
9	15	Identificar a relação de dependência entre grandezas.
8	7	Identificar características de figuras planas ou espaciais.
7	10	Identificar relações entre grandezas e unidades de medida.
7	17	Analisar informações envolvendo a variação de grandezas como recurso para a construção de argumentação.
7	20	Interpretar gráfico cartesiano que represente relações entre grandezas.
6	3	Resolver situação-problema envolvendo conhecimentos numéricos.
6	12	Resolver situação-problema que envolva medidas de grandezas.
6	13	Avaliar o resultado de uma medição na construção de um argumento consistente.
6	16	Resolver situação-problema envolvendo a variação de grandezas, direta ou inversamente proporcionais.
6	18	Avaliar propostas de intervenção na realidade envolvendo variação de grandezas.
5	2	Identificar padrões numéricos ou princípios de contagem.
5	5	Avaliar propostas de intervenção na realidade utilizando conhecimentos numéricos.
5	11	Utilizar a noção de escalas na leitura de representação de situação do cotidiano.
5	21	Resolver situação-problema cuja modelagem envolva conhecimentos algébricos.
4	6	Interpretar a localização e a movimentação de pessoas/objetos no espaço tridimensional e sua representação no espaço bidimensional.
4	9	Utilizar conhecimentos geométricos de espaço e forma na seleção de argumentos propostos como solução de problemas do cotidiano.
4	14	Avaliar proposta de intervenção na realidade utilizando conhecimentos geométricos relacionados a grandezas e medidas.
4	22	Utilizar conhecimentos algébricos/geométricos como recurso para a construção de argumentação.
4	23	Avaliar propostas de intervenção na realidade utilizando conhecimentos algébricos.

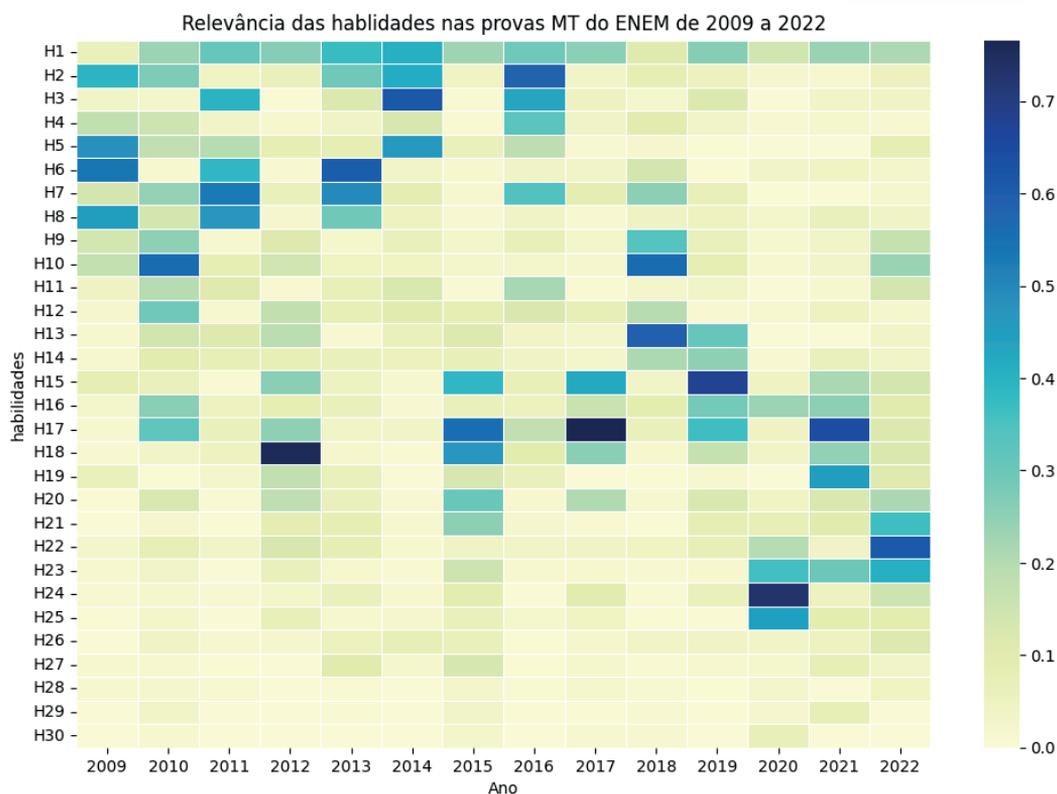
Frequência	Habilidade	Matriz de referência do ENEM
3	4	Avaliar a razoabilidade de um resultado numérico na construção de argumentos sobre afirmações quantitativas.
3	8	Resolver situação-problema que envolva conhecimentos geométricos de espaço e forma.
3	19	Identificar representações algébricas que expressem a relação entre grandezas.
3	24	Utilizar informações expressas em gráficos ou tabelas para fazer inferências
2	25	Resolver problema com dados apresentados em tabelas ou gráficos.
2	27	Calcular medidas de tendência central ou de dispersão de um conjunto de dados expressos em uma tabela de frequências de dados agrupados (não em classes) ou em gráficos
1	26	Analisar informações expressas em gráficos ou tabelas como recurso para a construção de argumentos.
1	30	Avaliar propostas de intervenção na realidade utilizando conhecimentos de estatística e probabilidade.
0	28	Resolver situação-problema que envolva conhecimentos de estatística e probabilidade
0	29	Utilizar conhecimentos de estatística e probabilidade como recurso para a construção de argumentação.

Assim como podem ser indicadas as habilidades que se apresentam no final da Tabela 2 como sendo aquelas que menos influenciam as demais, o que faz sentido por serem habilidades mais complexas e específicas.

É importante observar que as habilidades mais relevantes, de acordo com o ACP, variam em anos diferentes. E que apesar da habilidade mais relevante muitas vezes ter uma diferença relativamente alta em relação às demais, como no ano de 2019 (Figura 5), ainda assim é importante observar as demais habilidades, pois a proficiência deve ser construída como um conjunto de habilidades. Desta forma, a Figura 6, apresenta um mapa de calor em relação a relevância de cada habilidade, calculadas a partir do ACP, ao longo dos 14 anos, de forma independente ano a ano, ou seja, são calculadas as relações entre as habilidades apenas do mesmo ano. Esta relevância diz respeito a interpelação e covariância entre as variáveis usadas no cálculo, ou seja, as 30 habilidades. Neste mapa, quanto mais escuro representa maior relevância e quanto mais claro menor. Como as habilidades são associadas a conhecimentos cumulativos, dependentes de habilidades anteriores, observa-se que as últimas habilidades (mais próximas da H30) são as de menor valor, mas não de forma linear. Por exemplo, as habilidades com

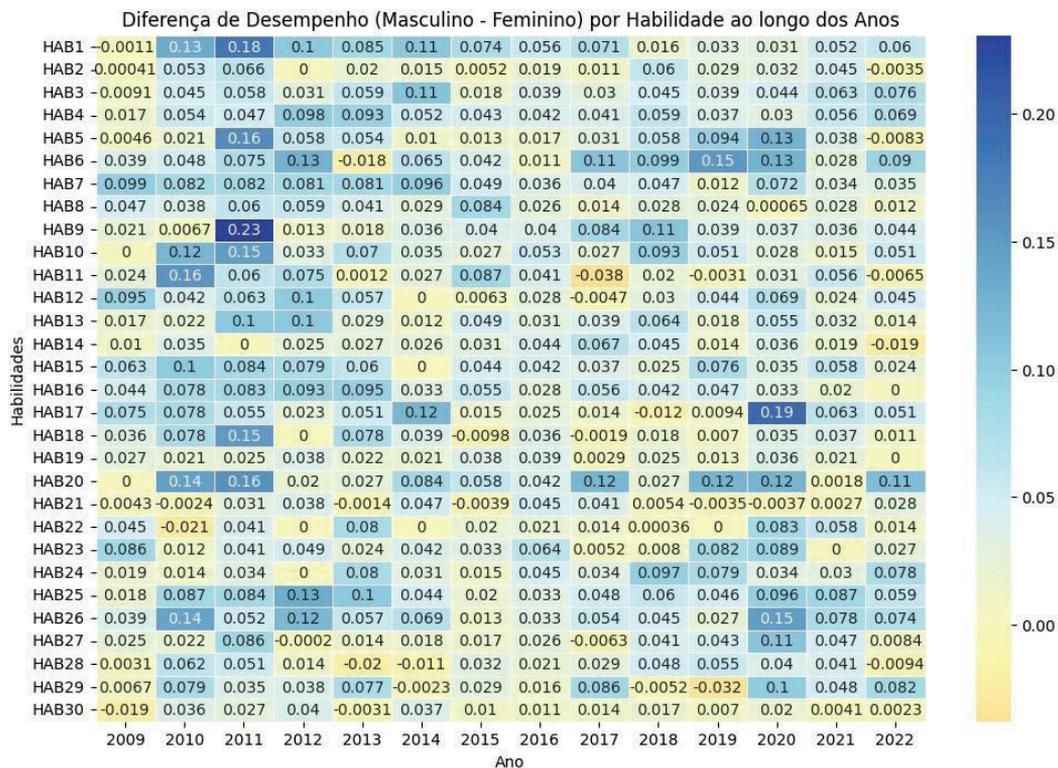
maiores valores gerais são H18, H17 e H24, respectivamente nos anos de 2012, 2017 e 2020, mas com valores relativamente baixos em outros anos. A habilidade H1, muito frequente entre as 10 principais, como observado anteriormente na Tabela 2, ao longo dos anos está quase sempre em um espectro entre azul e verde, mostrando que ela é relevante, mas nunca a de mais alta intensidade.

**Figura 6** – Mapa de calor do grau de relevância das habilidades (H1 a H30) em cada ano, calculadas pelo ACP



Ainda, no intuito de apresentar a comparação de acertos entre os participantes do sexo masculino e feminino, a Figura 7 apresenta outro mapa de calor, que expressa a diferença entre as médias de acertos dos participantes do sexo masculino e do sexo feminino para cada habilidade, por ano.

**Figura 7** – Diferença entre as médias de acertos dos estudantes do sexo masculino e feminino, por habilidade, por ano.



Os valores do gráfico da Figura 7 correspondem a diferença das médias de acerto do público masculino e o feminino por habilidade. Vale lembrar que uma mesma habilidade pode se apresentar em mais de um item da prova, isso quer dizer, que algum candidato poderia ter 0,5 acerto para uma habilidade que aparece em dois itens e o ele acertou apenas um deles. Assim, no gráfico, vemos a média masculina subtraída da feminina em cada habilidade de cada ano. Isto quer dizer que os valores negativos, que se apresentam nos tons mais claros (próximos do laranja), indicam habilidades que o sexo feminino apresentou maior média de acerto do que o masculino no ano indicado. Assim como observa-se que os valores em tons mais escuros (em azul), correspondem as médias maiores do público masculino. E os valores zero (em amarelo) indicam média iguais ou habilidades que não foram avaliadas no ano indicado. Portanto, no cenário de todos os anos apresentado nesta figura, não há uma predominância de nenhum dos sexos para nenhuma das habilidades, mesmo que as diferenças sejam mais acentuadas para o público masculino como nas habilidades HAB1 e HAB9 em 2011 e HAB17 em 2020, estas diferenças são muito pequenas, chegando no

máximo a 0,23. Desta forma, entendemos que participantes femininos, mesmo com menor proficiência, tem tanta competência em MT quando participantes masculinos.

## CONSIDERAÇÕES FINAIS

Motivado pela escassez de alunas do sexo feminino nos cursos de graduação de exatas, este trabalho buscou responder o questionamento se participantes deste sexo teriam menos habilidades ou competência em matemática, e este seriam um dos motivos desse fato. Diante disso, foi realizada a análise de desempenho dos participantes nas provas de Matemática e suas tecnologias, do ENEM, aplicadas entre 2009 e 2022, por sexo, aplicando diversas técnicas de ciência de dados, incluindo o ACP como método de reconhecimento de padrões e a análise estatística das proficiências dos candidatos e candidatas nas provas.

Obteve-se que o sexo feminino possui menores percentuais de acertos nas principais habilidades de cada ano, do que o sexo masculino, mas não em todas as habilidades. Dessa forma, considerando a capacidade de medição de competência, pelo modelo da TRI implementado no ENEM, foi verificado que o percentual de acerto dos participantes masculinos é maior na questão de maior dificuldade, maior discriminação e maior probabilidade de acerto ao acaso, em cada ano. Em todos os casos, as diferenças são pequenas, e desta forma, impactando diretamente sobre as proficiências, provocam pequenas diferenças entre as proficiências dos dois grupos. Entretanto, essa pequena diferença não é reproduzida na participação desses grupos nos cursos de exatas, sendo encontrados valores muito mais elevados. Dessa forma, é possível supor que, em vista da desproporção dos percentuais, o desempenho nas provas de matemática não é o único e nem o fator mais importante sobre o a taxa de participação nos cursos de exatas pelo sexo feminino.

Como trabalhos futuros indica-se a busca por outras justificativas para a desproporção do sexo feminino nas exatas, a partir de informações socioeconômicas e investigação da influência de fatores culturais. Ainda, apresenta-se como possível trabalho futuro o estudo semelhante aplicado sobre a análise dos motivos do desbalanceamento do sexo masculino em cursos da área da saúde e educação.

## REFERÊNCIAS

RICOLD, A.; ARTES, A.. Mulheres no ensino superior brasileiro: espaço garantido e novos desafios. **GEA - Grupo Estratégico de Análise da Educação Superior no Brasil**, n. 6, 2016.

BRASIL. Censo Brasileiro de 2022. **IBGE - Instituto Brasileiro De Geografia E Estatística**, 2023. Disponível em: <https://censo2022.ibge.gov.br/panorama/>

FUJII, T. R. C.; DE SOUZA, A. D. P.; FÜRKOTTER, M.; BORGATTO, A. F.; CÚRI, M.. Estudo sobre Construção de Escalas com Base na Teoria da Resposta ao Item: Avaliação de Proficiência em Conteúdos Matemáticos Básicos. **Bolema - Mathematics Education Bulletin**, V. 35, n. 71, P. 1876–1898, 2021. Disponível em: <https://doi.org/10.1590/1980-4415V35N71A29>

CRYSTINI, L.; SEHN, E.. A influência da matemática no ensino das ciências exatas. **RECIT - Revista Eletrônica Ciência Inovação e Tecnologia**. V. 8, n. 16, 2017. Disponível em: <https://periodicos.utfpr.edu.br/recit>

SILVEIRA, F. L. et al. Exame nacional do ensino Médio (ENEM): Uma análise crítica. **Revista Brasileira de Ensino de Física**. V. 37, n.1, P. 1101, 2015. Disponível em: <https://doi.org/10.1590/S1806-11173710001>

SOUZA, A. M.; POPPI, R. J.. Experimento didático de quimiometria para análise exploratória de óleos vegetais comestíveis por espectroscopia no infravermelho médio e análise de componentes principais: um tutorial, parte i. **Quim. Nova**, V. 35, n. 1, P. 223 – 229, 2012. Disponível em: <http://www.soft82.com/get/download/>

DUARTE, V.. Gênero / Gender. Conceitos-chave em Sociologia da Infância. Perspetivas Globais. **UMinho Editora**. P. 259–267, 2021. Disponível em: <https://doi.org/10.21814/uminho.ed.36.32>

ARAUJO, E. A. C. et al. Teoria da Resposta ao Item Teoria da Resposta ao Item. **Revista Escola de Enfermagem USP**. V. 4, P. 1000-1008, 2009. Disponível em: [www.ee.usp.br/reeusp/](http://www.ee.usp.br/reeusp/)

LOPES, F. L.; VENDRAMINI, C. M. M.. Igualização de provas acadêmicas via Teoria de Resposta ao Item. **Psico – USF**. V. 18, n. 1, P. 141-150, 2013. Disponível em: <https://www.scielo.br/j/psuf/a/Yjh4yd5BP53spJk5H8jhMDN/?lang=pt&format=pdf>

MACHADO, G. et al. Futuras Cientistas. **Câmara Brasileira do Livro**. 2023. Disponível em: <https://drive.google.com/file/d/1zFEIXp2TX-cx3QnDC5gXth6i8XgP6aHb/view>

CARNEIRO, S. G. et al. Mulheres nas ciências de exatas, engenharia e computação: uma revisão integrativa. **FINOM – Humanidades e tecnologias**. V. 20, 2020. Disponível em: [https://core.ac.uk/outputs/287360751/?utm\\_source=pdf&utm\\_medium=banner&utm\\_campaign=pdf-decoration-v1](https://core.ac.uk/outputs/287360751/?utm_source=pdf&utm_medium=banner&utm_campaign=pdf-decoration-v1)

HONGYU, K.; LÚCIA, V.; SANDANIELO, M.; JUNIOR, G J. O.. Análise de Componentes Principais: resumo teórico, aplicação e interpretação. **Engineering and Science**. V. 1, n. 5, 2016. Disponível me: <https://doi.org/10.18607/ES20165053>

INEP. Resumo técnico do censo da educação superior 2022. **Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira**. 2023. Disponível em: <https://www.gov.br/inep/pt-br/areas-de-atuacao/pesquisas-estatisticas-e-indicadores/censo-da-educacao-superior/resultados>

KLEIN, R.. Alguns aspectos da Teoria de Resposta ao Item relativos à estimação das proficiências. **Ensaio: avaliações e políticas públicas na Educação**. V. 21, n. 78, P. 35-56, 2013.

MELO, M. C. M. A.. Gênero e universidade: a presença da mulher aluna nos Cursos do Centro de Ciências Exatas e Tecnologia da Universidade Federal do Maranhão. **Universidade Federal do Maranhão**, 2013.

PINTO, E. J. S.; CARVALHO, M. E. P.; RABAY, G.. As relações de gênero nas escolhas de cursos superiores. **Revista Tempos e Espaços Em Educação**, V. 10, n. 22, P. 47-58, 2017. Disponível em: <https://doi.org/10.20952/revtee.v10i22.6173>.