

Avaliação Psicométrica do Desempenho em Física dos Alunos de Santarém-PA no ENEM 2019: Uma Abordagem com TCT e TRI⁺*

Ronilson dos Santos Bezerra¹

Glauco Cohen Ferreira Pantoja¹

Universidade Federal do Oeste do Pará
Santarém – PA

Resumo

A Física ainda é vista por muitos alunos como uma disciplina difícil pelos alunos, e os professores enfrentam desafios para aproximá-la da realidade dos estudantes e tornar o ensino mais dinâmico. O ensino da Física no Ensino Médio tem priorizado a resolução de problemas para o ingresso no ensino superior, negligenciando a compreensão dos fenômenos físicos. O Exame Nacional do Ensino Médio, atualmente o segundo maior exame de ingresso no ensino superior, fornece importantes dados sobre o desempenho dos alunos após o Ensino Médio. Neste contexto, uma pesquisa em Santarém com dados do Exame Nacional do Ensino Médio de 2019 utilizou a Teoria Clássica dos Testes e a Teoria de Resposta ao Item para analisar o desempenho dos alunos em Física. Os resultados revelaram que a média de acertos foi de apenas 22,56%, em linha com outros estudos nacionais, que mostram baixos resultados em Física e Ciências da Natureza. Além disso, alunos de menor renda tiveram resultados ainda piores na prova. Alguns itens não conseguiram avaliar adequadamente os alunos, apresentando baixo poder de discriminação.

Palavras-chave: *Ensino de Física; Enem; Avaliação; Psicometria.*

⁺ Psychometric Assessment of the Performance in Physics of Students from Santarém-PA in ENEM 2019: An Approach with TCT and TRI

* Recebido: 7 de janeiro de 2024.
Aceito: 11 de março de 2025.

¹ E-mails: ronilsonsb@gmail.com; glaucopantoja@hotmail.com

Abstract

Physics is perceived as a challenging subject by students, and teachers face challenges to relate it to students' reality and make the teaching more dynamic. Physics education in high school has prioritized problem-solving for college entrance, neglecting the understanding of physical phenomena. The Exame Nacional do Ensino Médio, currently the second-largest college entrance exam, provides essential data on students' performance after high school. In this context, a research study in Santarém with data from the Exame Nacional do Ensino Médio 2019 used Classical Test Theory and Item Response Theory to analyze students' performance in Physics. The results revealed an average score of only 22.56%, consistent with other national studies showing low performance in Physics and Natural Sciences. Moreover, students from lower-income backgrounds had even poorer results on the exam. Some items inadequately evaluated students, showing low discrimination power.

Keywords: Physics Teaching; ENEM; Evaluation; Psychometrics.

I. Introdução

Nos últimos anos, diversas pesquisas vêm sendo realizadas tendo como o foco a aprendizagem dos conceitos de Física pelos alunos do Ensino Médio. Esses estudos têm mostrado grandes desafios nesse processo de ensino-aprendizagem, tanto devido às dificuldades de compreensão por parte dos alunos quanto às dificuldades enfrentadas pelos docentes. Alves e Stachak (2005) afirmam que muitos docentes veem essa disciplina como difícil de ser ensinada, uma vez que enfrentam grandes dificuldades no processo de construção do conhecimento juntamente com seus alunos.

O ensino dessa disciplina ainda é percebido por alguns “como um objeto abstrato, distante da realidade dos alunos, o que gera um desinteresse total pelo trabalho escolar” (Alves; Stachak, 2005, p. 2). Menegotto e Rocha Filho (2008) ratificam essa percepção, afirmando que muitos alunos não compreendem a linguagem utilizada pelos professores de Física. Na concepção desses autores, essa dificuldade pode estar relacionada à “ausência de relações entre os conteúdos estudados e as situações cotidianas do jovem, o que dificulta a construção de significados. Assim, parece que o raciocínio do professor geralmente se encontra fora de sintonia com o do aluno” (Menegotto; Rocha Filho, 2008, p. 308).

Rosa e Rosa (2005) afirmam que a busca por um significado real no ensino da Física no Ensino Médio é uma questão emergente e tem sido objeto de várias pesquisas. No entanto, segundo eles, existe uma controvérsia em torno dos diferentes enfoques dados a essa disciplina nesse nível escolar, pois a Física, conforme apresentada nos livros didáticos e nas salas de aula,

está distante e distorcida de seu propósito real. Para esses autores, ensinar Física vai além da transmissão de conceitos e fenômenos; é proporcionar o aprimoramento pessoal do aluno. Essa visão é corroborada por pesquisas que indicam que o ensino de Física tem se direcionado mais para a preparação dos alunos para provas, visando a resolução de questões em processos seletivos para ingresso no ensino superior (Souza, 2002; Nardi, 1998).

A atual conjuntura do ensino de Física no Brasil reflete não apenas desafios pedagógicos, mas também limitações estruturais e políticas. Segundo Nesi et al. (2021), o cenário educacional é marcado por deficiências na formação de professores, infraestrutura inadequada e currículos desatualizados, aspectos que dificultam o avanço no ensino de Física. Além disso, a Base Nacional Comum Curricular (BNCC), embora represente um direcionamento para modernizar a educação, ainda carece de clareza quanto à sua implementação. Os autores destacam que a formação continuada de educadores é essencial para enfrentar esses desafios, propondo programas como o PARFOR e o MNPEF como alternativas para qualificar profissionais da área. No entanto, a ausência de incentivos financeiros e o tempo adicional exigido para a capacitação desmotivam muitos docentes, evidenciando a necessidade de políticas públicas mais integradas.

Para além disso, Nascimento (2020) destaca as desigualdades na formação e distribuição de professores de Física no Brasil, revelando que apenas 20% dos docentes nas escolas públicas estaduais possuem formação específica na área. O estudo ressalta que as regiões Norte e Centro-Oeste são as mais afetadas, com índices ainda menores de profissionais qualificados, respectivamente 18% e 15%. Além de questões de formação, fatores como gênero e etnia acentuam essas desigualdades, com uma sub-representação de mulheres e indígenas na área. Esses dados reforçam a urgência de políticas públicas direcionadas, capazes de equilibrar a distribuição de profissionais e incentivar a formação de professores em regiões mais carentes. Questionamentos e ponderações, como os mencionados anteriormente, evidenciam a necessidade de pesquisas que busquem compreender a aprendizagem dos alunos, a fim de enfrentar tais desafios. Nesse sentido, é importante destacar o papel significativo do Exame Nacional do Ensino Médio (ENEM) como uma fonte de dados utilizada em pesquisas sobre a aprendizagem dos alunos em diversas áreas. Cumpre destacar também que o ENEM desempenha um papel fundamental, visto que é atualmente a principal forma de ingresso no ensino superior, levando muitos alunos a prestar o exame com o objetivo de conquistar uma vaga universitária. Assim, seus resultados fornecem informações relevantes sobre o desempenho dos alunos no Ensino Médio em diferentes campos do conhecimento. No contexto específico da Física, Sousa Sobrinho e Rodrigues (2019) destacam a importância do ENEM, considerando-o como uma ferramenta orientadora para os professores da disciplina, uma vez que proporciona um “feedback dos conhecimentos adquiridos pelos alunos durante a etapa final da educação básica” (Sousa Sobrinho; Rodrigues, 2019, p. 59).

Barroso, Rubini e Silva (2018) realizaram um estudo com questões de Física de 2009 a 2014 e constataram algumas dificuldades permanentes de compreensão dos conceitos básicos

da disciplina. Segundo os autores, apesar de todo o esforço na área de pesquisas do Ensino de Física desde meados dos anos 1980, houve pouco impacto dos resultados dessas pesquisas no processo de aprendizagem. Já Marcom (2019), apresentou em sua tese uma análise de dados das edições do ENEM de 2009 a 2017, focada nos alunos que se declararam concluintes do Ensino Médio e egressos de escolas públicas no momento da inscrição nos exames. A pesquisa de Marcom (2019) revelou que esses alunos apresentaram índices de acerto abaixo de 50%, e mais de 72% dos candidatos classificaram os itens da prova como difíceis.

Nesse sentido, é possível verificar que há alguns desafios no ensino de Física aos alunos do Ensino Médio, uma vez que a disciplina ainda é vista por muitos estudantes como de difícil aprendizagem (Menegotto; Filho, 2008; Alves; Stachak, 2005; Barroso; Rubini; Silva, 2018). Além disso, em algumas situações, os estudantes não compreendem o conteúdo ensinado, pois este está distanciado da realidade do aluno (Menegotto; Filho, 2008) ou até mesmo distorcido de seu real propósito, concentrando-se mais na preparação dos alunos para os processos seletivos (Rosa; Rosa, 2005).

Esse artigo é produto da dissertação de mestrado do primeiro autor desse trabalho (Bezerra, 2023) e, nesse sentido, busca responder à seguinte questão norteadora: **Qual o desempenho dos alunos do Ensino Médio da cidade de Santarém na prova de Física do Enem 2019?** Para isso, são apresentados os resultados da dissertação citada, que investigou essa questão por meio de análises psicométricas baseadas na Teoria Clássica dos Testes (TCT) e na Teoria de Resposta ao Item (TRI). O estudo teve como objetivo compreender o desempenho dos estudantes locais no exame, evidenciando os desafios enfrentados no ensino de Física e suas implicações para a aprendizagem.

Todavia, antes de realizar uma análise nos dados do ENEM, buscando respostas para esta pergunta, é importante observar o que a literatura nacional já oferece sobre esse tema, uma vez que conforme mencionado acima, o ENEM tem sido fonte de dados para diversas pesquisas sobre esta temática. E neste aspecto, é crescente número de trabalhos na literatura brasileira que se utiliza dos dados do ENEM com o objetivo de diagnosticar problemas na Educação (Bezerra, 2023). Segundo Villar (2021) as pesquisas atuais sobre esse tema, apontam para uma diversidade de abordagens, abrangendo análise e categorização dos itens, exploração dos microdados e questionário socioeconômico, bem como estudos sobre o currículo e a qualidade do ensino básico.

Nesse contexto, buscando compreender a relevância dessas pesquisas, realizamos uma busca no Portal de Periódicos da Capes em outubro de 2022, utilizando os termos “ENEM” e “Exame Nacional do Ensino Médio”, e encontrou 871 trabalhos publicados entre janeiro de 2018 e outubro de 2022. Além disso, foi realizada uma busca semelhante no “Catálogo de Teses e Dissertações” da Capes, onde foram encontradas 6.535 publicações no mesmo período, sendo 3.756 dissertações e 1.969 teses. Refinando a busca para a área da educação, foram encontradas 198 dissertações e 94 teses. Por fim, uma busca no Google Acadêmico com os termos “Teoria

de Resposta do Item” e “Enem” resultou em 730 publicações para o mesmo período (2018 a 2022).

Disso, resultou uma seleção das pesquisas mais importantes para nortear a pesquisa. Elas serão mostradas resumidamente nos parágrafos seguintes, com enfoque nos estudos que utilizaram a Teoria de Resposta ao Item (TRI) e a Teoria Clássica dos Testes (TCT) em pesquisas educacionais. Em seguida, serão apresentados trabalhos que abordam mais diretamente sobre itens de Física no contexto do ENEM.

Silva (2019a) realizou análises psicométricas e pedagógicas da prova da primeira fase da 13^a edição das Olimpíadas Brasileira de Matemática das Escolas Públicas (OBMEP) para alunos do 6º e 7º ano do ensino fundamental. Os resultados obtidos por meio da TCT e TRI revelaram a baixa qualidade da prova para a amostra considerada, sobretudo no que se refere à consistência interna do instrumento. Além disso, a análise pedagógica apontou dificuldades dos alunos na resolução de problemas matemáticos.

A pesquisa de Silva (2019b) analisou os microdados do ENEM 2017, referentes aos alunos de Santarém que responderam a prova de matemática. Os resultados mostraram a dificuldade dos alunos em dominar essa disciplina, com apenas 22,8% de acertos, sendo as escolas privadas com resultados melhores que as públicas. Foi observado que muitos conteúdos exigidos no ENEM fazem parte dos estudos no ensino fundamental, indicando possíveis problemas na transição para o Ensino Médio.

Sousa (2019) avaliou os resultados do ENEM 2017 utilizando a TCT e a TRI, identificando a falta de unidimensionalidade dos itens das provas. Segundo o autor, para alcançar um modelo unidimensional, foram necessárias exclusões de diversos itens em diferentes áreas do conhecimento. O autor sugere a necessidade de mais estudos visando uma melhor compreensão do uso da TRI nas finalidades do ENEM.

Ao analisar os trabalhos diretamente relacionados aos conteúdos de Física exigidos nas provas do ENEM, destaca-se a pesquisa realizada por Villar (2021), que em sua tese, mostrou os resultados de uma análise fatorial realizadas nos itens de Física, contidos nas edições do ENEM no período de 2012 a 2017. O resultado dessa análise, apontou que 83,7% dos itens podem ser considerados difíceis para os alunos do Ensino Médio.

Marcom e Kleinke (2016) focaram seus estudos na análise dos distratores mais escolhidos pelos candidatos nas provas de Física do ENEM de 2009 a 2012. Neste trabalho, os autores constatam que os estudantes utilizavam ideias não científicas ao resolver questões e apresentavam dificuldades em transpor os problemas para outros contextos. Muitos também utilizaram raciocínios intuitivos e formas simbólicas em questões que exigiam abordagem matemática.

Marcom (2019) destaca em sua tese, que a partir da análise do desempenho dos alunos nas edições de 2009 a 2017 do ENEM, é possível verificar que cerca de 70% dos itens de Física são considerados difíceis para os candidatos, com um índice de acerto abaixo de 50%. Essa informação é corroborada por Marcom e Kleinke (2021), que acrescentam a identificação de

dificuldades conceituais e matemáticas através da análise dos erros cometidos. Por sua vez, Rubini (2019) analisou as provas de Física do ENEM entre 2009 e 2017, concluindo que, tanto pela perspectiva da TCT quanto da TRI, as provas não são adequadas para a maioria dos candidatos, evidenciando desafios significativos no ensino de Física.

Por fim, é importante destacar o trabalho de Vizzotto (2022) que realizou uma análise das edições do ENEM de 2009 a 2019, utilizando a TCT e a TRI, e apontou que 79,95% dos itens de Física foram considerados inadequados, devido às distorções nos indicadores psicométricos. O autor enfatiza a necessidade de melhorar a mensuração da proficiência em Ciências da Natureza na prova.

Essas pesquisas apontam para a relevância e a complexidade das questões que envolvem a Física no ENEM, fornecendo bases importantes para discussões sobre melhorias na educação e aprimoramento do exame, sendo evidente a necessidade de estudos mais amplos sobre esta temática. Sendo assim, estabelecemos um recorte para nortear o objetivo geral desse estudo, que ficou definido da seguinte forma: analisar o desempenho dos alunos do município de Santarém-PA² nos itens da disciplina de Física no ENEM de 2019.

Para atingir esse objetivo geral, foram delineados os seguintes objetivos específicos: 1) descrever os dados do desempenho dos candidatos, residentes em Santarém, que responderam os itens de Física do ENEM 2019 e utilizar nesta análise técnicas psicométricas da TCT; 2) estimar os parâmetros de dificuldade, discriminação e acerto ao acaso dos itens citados no objetivo anterior e a habilidade dos participantes utilizando técnicas psicométricas da TRI, e; 3) realizar uma análise descritiva dos itens de Física, do ENEM 2019, mostrando uma das possíveis resoluções para cada item e destacando as alternativas mais escolhidas pelos candidatos de Santarém.

A escolha de Santarém como foco desta pesquisa se justifica pelos desafios educacionais locais e pela carência de estudos específicos sobre o desempenho em Física. Como destacado por Junior e Barroso (2012), compreender avaliações em larga escala, como o ENEM, pode subsidiar políticas públicas e orientar práticas pedagógicas. Estudos anteriores, como o de Silva (2019c), já apontaram o baixo desempenho dos alunos de Santarém em matemática no ENEM 2017, com apenas 22,8% de acertos, refletindo dificuldades que também afetam outras áreas, como Física, e destacando a necessidade de intervenções pedagógicas mais eficazes.

Dessa forma, esta pesquisa busca analisar o desempenho dos alunos da cidade em Física no ENEM 2019, contribuindo para identificar os fatores que dificultam a aprendizagem e propondo reflexões que auxiliem na melhoria do ensino-aprendizagem local. Os resultados

² Fundada em 1661, localizada na região Oeste do Estado do Pará, Santarém é uma das cidades mais antigas da Amazônia. Nesse município encontra-se o Programa de Pós-Graduação em Educação (PPGE) da Universidade Federal do Oeste do Pará, ao qual o autor desse trabalho está vinculado. De acordo com o censo de 2010, Santarém possuía uma população de 294.580 habitantes. Entretanto, conforme as estimativas mais recentes do IBGE, a cidade agora abriga cerca de 308.000 habitantes, o que a enquadra como terceiro maior município do Estado do Pará.

podem subsidiar ações voltadas para formação docente, ajustes curriculares e estratégias pedagógicas que conectem os conteúdos de Física à realidade dos estudantes, promovendo avanços na educação em Santarém e na região amazônica.

II. A avaliação em larga escala e o ENEM

Nesta sessão, apresentamos uma breve revisão teórica sobre avaliação em escala e o Exame Nacional do Ensino Médio. Gonçalves Junior e Barroso (2012) defendem que a compreensão das avaliações em larga escala é fundamental, uma vez que estas podem servir de base para políticas públicas em educação, além de subsidiar professores e influenciar o currículo escolar e a carga horária das disciplinas. Sudbrack e Cocco (2014) também ressaltam essa importância, enfatizando que por meio dela é possível verificar erros e acertos para aprimorar a educação.

O ENEM é atualmente a principal forma de ingresso no ensino superior no Brasil, além de ser considerado a maior avaliação em larga escala do país e o segundo maior processo seletivo do mundo (Marcom, 2019; Brasil, 2022). Criado em 1998 com o objetivo fundamental de “avaliar o desempenho do aluno ao término da escolaridade básica, para aferir o desenvolvimento das competências fundamentais ao exercício pleno da cidadania” (Brasil, 1998, p. 1), o ENEM era composto por 63 questões e uma redação. Até o ano de 2008, a nota dos participantes desse exame era calculada por meio da TCT, que levava em consideração a quantidade de acertos dos itens da prova objetiva e pela nota da redação, ambas variando de 0 a 100. Nesse período, a avaliação do aluno era “baseada unicamente no total de questões respondidas corretamente” (Quaresma; Dias; Barbetta, 2019).

Bezerra (2023) destaca que a partir de 2004 o ENEM deixou de ser apenas uma avaliação de competências aplicada a alunos do Ensino Médio e passou também a ser um exame para permitir o ingresso no Ensino Superior, quando o Programa Universidade para Todos (Prouni) decidiu utilizar os resultados desse exame como critério para concessão de bolsas de integrais e parciais de estudos. Essas mudanças tiveram um impacto positivo no número de inscritos, resultando em um crescimento de mais de 90% nas inscrições em 2005, saltando de 1.552.316 para 3.004.091. De acordo com Brasil (2020b), 67% desses inscritos almejavam uma vaga no ensino superior.

A partir do ano de 2008, o Ministério da Educação (Mec) tornou o ENEM em um processo nacional de seleção para ingresso no ensino superior e certificação do Ensino Médio. Com isso, o número de inscritos subiu para mais de 4 milhões, com 70% desse total de inscritos tendo prestado o exame visando entrar na faculdade ou conseguir pontos para o vestibular, conforme dados divulgados por Brasil (2020b).

No ano seguinte o ENEM teve mudanças no formato, na sua matriz de referência e na forma de cálculo dos scores dos seus participantes. A partir de então, este exame passou a ser constituído por 180 questões objetivas e uma redação. Como parte destas mudanças, a matriz do ENEM passou a ser composta por quatro áreas do conhecimento: a) Linguagens, Códigos e

suas tecnologias; b) Matemática e suas tecnologias; c) Ciências Humanas e suas tecnologias; e d) Ciências da Natureza e suas tecnologias (Brasil, 2019B; Gonçalves Junior; Barroso, 2012). Na concepção de Marcom e Kleinke (2016), estas mudanças foram realizadas com o objetivo de transformar o ENEM num exame de acesso ao ensino superior, nas Universidades Federais, por meio do Sistema de Seleção Unificada (Sisu).

A partir destas mudanças realizadas em 2009 e da criação do Sisu, o ENEM passou a ser o principal processo seletivo para ingresso nas universidades públicas. Com isso, a partir de então o número de participantes do ENEM teve um crescimento exponencial. Sendo assim, embora esse exame não seja aplicado a todos os alunos do nível médio, o resultado do ENEM se tornou bem mais representativo em relação ao total ao aluno de ensino médio no Brasil, devido ao crescimento do número de participantes em todo o país, uma vez que o Exame passou a ser aplicativo em todo o território nacional e, como mencionado acima, passou a ser utilizado para ingressos no ensino superior (Vizzotto, 2022).

De acordo com Gonçalves Junior e Barroso (2012), o ENEM possui alguns vícios de origem, como o fato de não ter sido inicialmente concebido como uma avaliação de sistema, mas sim como uma avaliação individual. Além disso, não é uma avaliação universal nem amostral no que diz respeito aos estudantes do Ensino Médio. No entanto, Bezerra (2023) observa que, devido ao crescimento exponencial no número de participantes, os dados do ENEM agora representam uma parcela significativa da população. Isso corrobora a visão de Gonçalves Junior e Barroso (2012) de que, apesar dessas limitações, o ENEM continua fornecendo indicadores cruciais para avaliar a qualidade da educação básica no Brasil. Tais indicadores podem ser obtidos analisando os resultados fornecidos pelo Inep.

No entanto, apesar da disponibilização dos resultados de todas as edições do ENEM pelo Inep em seu site, muitas escolas enfrentam dificuldades em acessar esses dados devido à sua forma de apresentação. Isso porque, em vez desse resultado fornecer informações detalhadas sobre o desempenho individual de cada aluno, as escolas recebem apenas médias de desempenho por área do conhecimento. Essas médias muitas vezes não refletem de maneira precisa o percentual de alunos que efetivamente dominam os conteúdos, como destacado por Silva (2019c).

A disponibilização dos microdados do Enem pelo Inep representa uma fonte de informações valiosa, mas sua utilidade para a melhoria do ensino ainda é limitada. Isso ocorre porque os dados são fornecidos de maneira agregada, sem detalhamento individualizado, o que dificulta uma análise mais precisa do progresso dos alunos e da identificação de lacunas na aprendizagem. Além disso, o acesso e a interpretação desses microdados exigem infraestrutura adequada e profissionais capacitados, o que pode representar um desafio para muitas escolas. Dessa forma, torna-se fundamental que políticas educacionais promovam a sistematização e a acessibilidade dessas informações, garantindo que os resultados do Enem possam ser utilizados de maneira mais eficaz para subsidiar reflexões sobre o ensino e a aprendizagem, sem, contudo, sobreregar as instituições escolares com essa responsabilidade.

Gonçalves Junior e Barroso (2014) enfatizam a importância da análise dos microdados disponibilizados pelo Inep como uma ferramenta essencial para professores que buscam compreender melhor o desempenho dos alunos no final do Ensino Médio. Neste contexto, nossa pesquisa conduz uma análise psicométrica das questões de Física no ENEM 2019, empregando tanto a TCT quanto a TRI. Essa análise fornece dados valiosos que podem enriquecer o conhecimento de professores e pesquisadores sobre o domínio dos conteúdos de Física pelos alunos, conforme mostrado posteriormente neste artigo.

III. Avaliação psicométrica com uso da TCT e da TRI

Ao longo da história da Psicologia um dos principais desafios tem sido compreender como mensurar fenômenos psicológicos, emocionais e cognitivos (Matsunaga, 2022). Isso inclui não apenas a definição de conceitos complexos, mas também a busca por estratégias adequadas para quantificá-los. Desse modo, a Psicometria, surgida no início do século XX, tornou-se fundamental para o desenvolvimento de instrumentos objetivos que permitissem analisar fenômenos subjetivos, como inteligência, personalidade, criatividade, emoção, percepção e memória (Pasquali, 1997; Pasquali, 2017; Alchieri, 2003).

A psicometria é um campo interdisciplinar da Psicologia que se dedica ao estudo das medidas psicológicas, incluindo a criação e avaliação de testes psicológicos, e estabelece conexões com diversas outras áreas do conhecimento, como a neurociência, sociologia e filosofia (Pasquali, 2017; Matsunaga, 2022). Ela se baseia na teoria da medida em ciências em geral, com ênfase na abordagem quantitativa, fazendo uso de testes, ferramentas essenciais nas ciências psicossociais, para estimar o comportamento psicológico dos indivíduos, com o objetivo de minimizar os erros associados a essa medição (Fernández *et al.*, 1997; Pasquali, 2009). Em sua essência, a Psicometria concentra-se na medição de construtos psicológicos e traços latentes, valendo-se de comportamentos verbais ou motores como representações desses traços (Lord, 1980, apud Pasquali, 1996).

As contribuições da Psicometria para a Psicologia Educacional se refletem na necessidade de adotar uma abordagem mais centrada na cognição para interpretar os resultados de avaliações, além das expectativas comportamentais, bem como na garantia de que pesquisadores e educadores compreendam os princípios subjacentes à avaliação e à melhoria da qualidade do ensino, especialmente em sistemas que adotam esse paradigma de referência (Braga, 2018). Desse modo, as técnicas psicométricas tornaram-se cruciais para determinar o grau de aptidão dos alunos por meio de instrumentos que são válidos e confiáveis, especialmente em avaliações decisivas (Collares; Grec; Machado, 2012).

Pasquali (2009) destaca a existência de duas vertentes na Psicometria contemporânea: a TCT e a TRI. Ambas são consideradas referenciais que podem ser utilizados tanto na fase de construção e validação quanto na avaliação de instrumentos em Psicologia da Educação, conforme observado por Araújo *et al.* (2020) e podem ser utilizadas na avaliação de construtos cognitivos no contexto educacional, sendo construtos habilidades ou características do sujeito

que não podem ser diretamente medidas. Para medi-los, são necessários instrumentos bem elaborados, visando proporcionar uma medição com o mínimo de erro possível (Araújo *et al.*, 2020).

III.1 A TCT e sua Importância na Avaliação Psicométrica

A TCT representa um marco crucial no desenvolvimento da avaliação psicométrica. Um dos pilares da TCT é sua ênfase no escore total do teste. Assim, a soma das pontuações individuais de cada questão culmina na nota geral do aluno, servindo como um indicador do domínio do conhecimento do avaliado. A TCT também fornece escores brutos ou padronizados, facilitando a análise da confiabilidade e do erro padrão da medida do teste como um todo (Santos *et al.*, 2022).

No entanto, uma das principais restrições da TCT é a sua dependência dos itens específicos da prova, tornando a comparação entre indivíduos que não realizaram a mesma avaliação praticamente impossível. Essa limitação torna-se evidente, por exemplo, quando se busca comparar o desempenho de estudantes em edições diferentes de um exame (Andrade; Tavares; Valle, 2000).

III.2 Entendendo a TRI na Avaliação Psicométrica

A TRI surgiu como uma resposta a algumas limitações encontradas na TCT. Autores como Fernández *et al.* (1990) e Embretson e Reise (2000) apontam que a TRI se propõe a resolver problemas inerentes ao modelo clássico, como a conexão entre a medida e o tipo de teste utilizado, a necessidade de uma amostra representativa na estimação dos parâmetros, e a limitação de usar o escore total como referência para a medida.

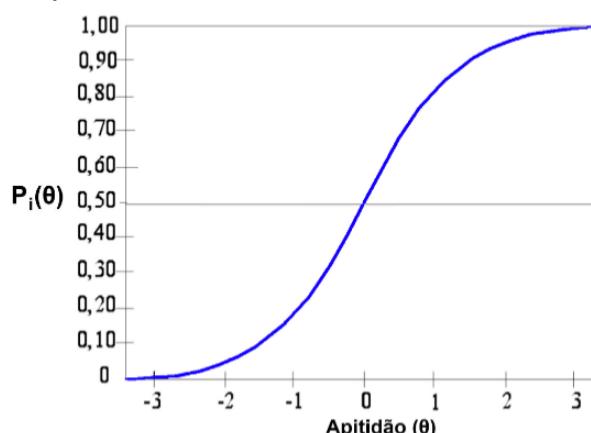
A TRI lida com modelos probabilísticos que estabelecem uma relação entre a probabilidade de acerto em um item específico e o nível de habilidade da pessoa testada (Collares; Grec; Machado, 2012). Essa abordagem da TRI traz várias vantagens em relação à TCT. Sendo que um dos principais benefícios é que os parâmetros dos itens e os níveis do traço latente são independentes, e o modelo é expresso no nível das respostas observadas em vez do escore total do teste. Além disso, a TRI fornece informações valiosas sobre a relação entre as respostas dos indivíduos e seus traços latentes, o que a torna altamente útil na análise de construtos psicológicos (Guewehr, 2007).

Na TRI, a estimativa da habilidade, representada por “theta” (θ), está relacionada à expectativa de que uma pessoa responda perfeitamente aos itens do teste, considerando um ou mais parâmetros. É por esse motivo que alguns consideram que a TRI também pode ser chamada de Teoria do Traço Latente, uma vez que lida com características que não podem ser observadas diretamente, mas que podem ser deduzidas a partir de variáveis secundárias relacionadas a essas características. Ela permite representar a relação entre o grau de aptidão de um sujeito e sua probabilidade de responder corretamente a um item, resultando na curva

característica do item (CCI), conforme pode ser observado no Gráfico 1, que é uma função monotonicamente crescente (Pasquali; Primi, 2003; Araújo *et al.*, 2020).

De acordo com Guewehr (2007), a TRI é uma metodologia avançada e complexa que depende de recursos computacionais especializados. Ela também está intrinsecamente ligada a outras técnicas estatísticas, como análise fatorial, análise de sobrevivência e análise de medidas repetidas, tornando-se uma abordagem versátil na análise de dados (Guewehr, 2007). Além disso, devido à capacidade dessa teoria de lidar com itens individualmente, a TRI se torna muito útil para estudos que exigem comparações entre diferentes grupos de indivíduos ou avaliações distintas (Andrade, Laros; Gouveia, 2010).

Gráfico 1: Curva Característica do Item (CCI).



Fonte: Bezerra (2023), adaptado de Pasquali e Primi (2003).

É importante observar que, apesar das vantagens oferecidas pela TRI em relação à TCT, essas duas abordagens não são mutuamente exclusivas, mas complementares. Ademais, cabe destacar que ao utilizar a TRI é indispensável que se conheçam dois pressupostos fundamentais dessa teoria, conforme mostrados a seguir.

III.2.1 Pressupostos da TRI e seus modelos logísticos

Os pressupostos fundamentais da TRI, conforme destacados por Pasquali e Primi (2003), incluem a unidimensionalidade e a independência local. Embora modelos bidimensionais e multidimensionais da TRI já tenham sido desenvolvidos, a unidimensionalidade continua sendo um requisito essencial quando o constructo mensurado se pretende unidimensional. No contexto do Enem, a modelagem adotada pressupõe a unidimensionalidade das áreas avaliadas, garantindo que a escala de proficiência seja interpretada dentro desse princípio.

Em primeiro lugar, a unidimensionalidade refere-se à suposição de que os itens de um teste estão medindo um único traço latente, representado por θ , que influencia as respostas dos participantes (Fernández *et al.*, 1990). Isso implica que, apesar de muitos comportamentos humanos serem multideterminados, a TRI assume que um traço latente dominante está sendo

medido pelo teste (Sousa, 2019). Embora esta suposição não pode ser diretamente provada e possa ser questionada, é essencial para o funcionamento eficaz dos modelos TRI (Pasquali; Primi, 2003).

Em segundo lugar, a independência local é outra suposição fundamental da TRI. Isso significa que, mantendo constantes as aptidões que afetam o teste, as respostas dos participantes a diferentes itens são estatisticamente independentes (Pasquali; Primi, 2003). Em outras palavras, o desempenho em um item não deve afetar o desempenho em outro, uma vez que cada item é respondido com base no traço latente do indivíduo (Sousa, 2019).

Essas suposições, embora possam parecer simplificadoras, são indispensáveis para o funcionamento adequado dos modelos que utilizam a TRI. Ao aceitar a unidimensionalidade e a independência local, a TRI permite que os itens sejam considerados como unidades independentes que podem ser usadas para medir traços latentes em escalas de habilidade (Quaresma, 2014). Portanto, esses pressupostos são fundamentais para a avaliação psicométrica baseada na TRI e desempenham um papel vital na construção e interpretação de testes psicométricos.

Os modelos logísticos desempenham um papel essencial na TRI, proporcionando uma estrutura matemática sólida para avaliações psicométricas. Esses modelos são fundamentais para a análise de itens de teste, permitindo a estimativa da probabilidade de um indivíduo acertar um item específico com base em seu nível de habilidade latente (θ). Os três modelos logísticos comumente utilizados na TRI são: o modelo logístico de um parâmetro (ML1), o modelo de dois parâmetros (ML2) e o modelo de três parâmetros (ML3).

Modelo Logístico de 1 parâmetro	$Pi(\theta) = \frac{e^{D(\theta-b_i)}}{1 + e^{D(\theta-b_i)}}$
Modelo Logístico de 2 parâmetros	$Pi(\theta) = \frac{e^{Da_i(\theta-b_i)}}{1 + e^{Da_i(\theta-b_i)}}$
Modelo Logístico de 3 parâmetros	$Pi(\theta) = c_i + (1 - c_i) \frac{e^{Da_i(\theta-b_i)}}{1 + e^{Da_i(\theta-b_i)}}$

Fig. 1 – Expressões matemáticas dos três modelos logísticos da TRI.

Fonte: Sousa (2019).

Esses modelos são expressos por meio de equações matemáticas (Fig. 1) que relacionam as variáveis, incluindo a probabilidade de sucesso do participante ($Pi(\theta)$) em relação ao seu nível de habilidade (θ), o índice do item (i), o parâmetro de discriminação (a), a dificuldade do item (b), o índice de acerto ao acaso (c), e outros fatores como a base logarítmica e um fator de escala. A escolha do modelo a ser aplicado depende da complexidade dos itens e dos objetivos da avaliação.

III.3 O ENEM e o uso da TCT e da TRI

No ano de 2009 o ENEM passou por uma significativa mudança em sua metodologia de cálculo das notas, adotando a TRI. Essa mudança foi motivada pela percepção de que o modelo anterior, baseado na TCT, não aproveitava todas as informações disponíveis nas respostas dos candidatos.

Até então, com o uso da TCT, o *score* era obtido pela contagem das respostas corretas, ignorando se o candidato mantinha um padrão coerente de respostas de acordo com sua proficiência. Com a TRI, a coerência das respostas passou a ser levada em consideração. Isso significa que dois alunos que acertem o mesmo número de questões podem obter notas diferentes, dependendo do grau de dificuldade das questões respondidas corretamente (Brasil, 2021). Isso porque, a TRI possibilita dar uma pontuação mais alta para aqueles que acertam tanto os itens mais complexos quanto os mais fáceis da prova, mesmo que o número total de respostas corretas seja igual ao de outro candidato que acertou questões menos complexas (Brasil, 2019c).

Em 2019, por meio de uma nota técnica³, o Inep informou que adota o modelo logístico de 3 parâmetros da TRI e justificou a decisão de usar a TRI no ENEM listando duas finalidades principais: (1) permitir a comparabilidade dos resultados entre os anos e (2) permitir a aplicação do Exame várias vezes ao ano.

IV. Metodologia

Esta pesquisa empregou uma abordagem mista, combinando métodos quantitativos e qualitativos para analisar e comparar o desempenho dos alunos de Santarém que prestaram o ENEM 2019. Utilizando microdados disponibilizados pelo Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira (Inep), conduziu-se uma análise qualitativa que consistiu em uma investigação observatória e interpretativa do desempenho dos alunos na prova de Ciências da Natureza do ENEM 2019, com foco na seção de Física.

Durante a análise quantitativa, foram utilizadas ferramentas estatísticas, e os microdados do ENEM 2019 foram processados com o auxílio do software RStudio. Este facilitou a extração, organização e interpretação dos microdados, incluindo a segmentação dos dados para concentrar-se nos alunos de Santarém.

Os dados utilizados na pesquisa de Bezerra (2023) foram obtidos no site do Inep, que disponibiliza os microdados de avaliações e censos realizados pelo referido instituto. Entende-se por microdados a representação do menor nível de desagregação de dados coletados pelo Inep em suas pesquisas estatísticas. É importante frisar que a pesquisa utilizou os dados do

³ Nota técnica - Assunto: Teoria de Resposta ao Item. Disponível em https://download.inep.gov.br/educacao_basica/enem/nota_tecnica/2011/nota_tecnica_tri_enem_18012012.pdf. Acesso em: 02de Jan 2025

ENEM 2019, pois estes eram os dados da última edição do exame quando a pesquisa foi iniciada.

É relevante observar que, o ENEM 2019 incluiu cinco tipos de provas com questões de Ciências da Natureza. Sendo que neste estudo, foi utilizado o Caderno Azul como referência, onde os itens estavam numerados de 91 a 135. Para simplificar a análise e identificar cada item, Bezerra (2023) optou por atribuir a cada item o número correspondente ao seu número no caderno de prova, acrescido da letra “i”. Sendo assim, “i92”, por exemplo, refere-se ao “Item 92 da prova de Ciência da Natureza do Caderno Azul do ENEM 2019”.

Para analisar apenas os itens de Física, foi necessário realizar uma análise dos dados e identificar todos os itens com conteúdos relacionados a esta disciplina. É importante destacar que o ENEM utiliza diferentes provas, embora os itens sejam os mesmos em todas elas, estão organizados de maneira distinta em cada prova. Sendo assim, foi necessário reorganizar todos esses itens em um novo banco de dados para proceder à análise utilizando um mesmo gabarito. Nesta reorganização, colocou-se os 15 itens em colunas (nomeando-as de Q1 a Q15), e as respostas dos alunos foram colocadas nas linhas, abaixo das respectivas colunas, e estas linhas foram chamadas de A1 a A11624. Onde “Q” se refere à questão (item), “A” se refere ao aluno, e o número 11.624 se refere à quantidade de alunos que prestaram a prova do ENEM 2019 na cidade de Santarém-PA. Esses procedimentos foram realizados através de filtros em planilhas eletrônicas e códigos desenvolvidos no software RStudio e foram bem detalhados por Bezerra (2023) em sua dissertação.

IV.1 Software e Pacotes Utilizados

O software RStudio foi escolhido para a análise de dados, pois é um ambiente de programação que permite a utilização da linguagem R, amplamente empregada em estatística e análise de dados. Além disso, o RStudio oferece uma interface intuitiva e recursos de gerenciamento de pacotes que simplificam o processo de análise.

Dois pacotes específicos foram utilizados para a análise dos dados: (1) o pacote “Psych”, que foi empregado para análises utilizando a TCT, uma vez que ele fornece diversas funções para análise usando esta teoria e possui também funções para visualização dos dados; (2) o pacote “Ltm”, utilizado para análises via TRI, permitindo a análise de dados dicotômicos usando modelos de traços latentes.

O uso do software RStudio e dos pacotes específicos para análise psicométrica possibilitou a obtenção de informações importantes, como será mostrado a seguir.

IV.2. Dicotomização dos dados e proporção de erros e acertos

No início da análise, Bezerra (2023) realizou a importação dos dados usando o RStudio e criou dois data frames separados: um para as respostas dos alunos e outro para o gabarito. Em seguida, foi feita a dicotomização dos dados, atribuindo o valor 1 aos itens respondidos corretamente pelos alunos e 0 aos itens respondidos incorretamente por eles.

Após essa dicotomização, a função “descript()” do pacote ltm foi utilizada para avaliar a proporção de erros e acertos em cada item. Essa informação é importante na TCT, pois essa teoria se baseia no escore bruto de cada indivíduo, sendo o resultado obtido pela contagem de questões respondidas corretamente.

IV.3 Correlação Ponto-Bisserial

Por meio da função “descript()”, Bezerra (2023) conduziu uma análise da correlação ponto-bisserial (*rpb*) para cada item. Esta é uma medida crucial, pois o coeficiente de correlação ponto-bisserial possibilita avaliar o poder de discriminação de um item, comparando o escore total dos alunos que realizaram o teste com o dos estudantes que acertaram um item específico.

Essa estatística é amplamente utilizada na Psicometria Clássica, sendo uma derivada da Correlação de Pearson e é essencial para medir a associação entre o acerto de um item e a nota total do teste (Soares, 2018). Espera-se que a correlação seja positiva para a alternativa correta e negativa para as alternativas erradas, comumente chamadas de “distratores.”

O valor da correlação ponto-bisserial varia no intervalo de [-1, 1]. Onde, valores próximos a zero ou negativos indicam que os alunos com bom desempenho geral no teste estão respondendo incorretamente a esse item específico. Isso pode sugerir que o item não distingue efetivamente entre alunos com diferentes níveis de habilidade (Ferreira, 2018). Sousa (2019) toma como critério de discriminabilidade a *rpb* acima de 0,20, portanto a análise dos itens foi feita seguindo esse parâmetro (Bezerra, 2023).

IV.4 Coeficiente Alpha de Cronbach

O coeficiente Alpha de Cronbach é uma medida de confiabilidade do teste e reflete a correlação entre cada item e o teste como um todo, indicando o quanto cada questão contribui para o teste de maneira geral. Conforme Soares, Soares e Santos (2021), a confiabilidade de um teste é interpretada como a capacidade de reproduzir resultados semelhantes quando aplicado por diferentes pesquisadores e em momentos distintos. Isso a torna uma medida crucial na avaliação da qualidade de um teste. Para estimar esse parâmetro, foi usada a função “descript()” do pacote LTM, considerando os 15 itens de Física respondidos pelos alunos de Santarém no ENEM 2019.

Vale ressaltar que, como as questões de Física estão incluídas no caderno de prova de "Ciências da Natureza e suas Tecnologias (CN)", que também abrange itens de Química e Biologia (totalizando 45 questões), esta pesquisa definiu como itens de Física aqueles que faziam parte da prova de CN e tratavam de conteúdos relacionados à Física ou exigiam conhecimentos dessa área para serem resolvidos. Essa identificação foi realizada com base na Matriz de Referência, disponibilizada junto aos Microdados do ENEM 2019. De acordo com Sousa (2019), índices de confiabilidade baseados na fidelidade do teste, como o Alpha de Cronbach, pressupõem que itens consistentes “compartilham uma grande parte da variância, indicando a presença de um único fator sendo medido por esses itens.” Os valores do coeficiente

Alpha de Cronbach podem variar de 0 a 1, sendo que valores mais próximos de 1 indicam maior confiabilidade do teste. A interpretação desses valores é subjetiva, mas é comum considerar que quanto mais próximo de 1, mais confiável é o teste (Freitas; Rodrigues, 2005).

Para Pasquali (2017), coeficientes inferiores a 0,70 nem sequer são considerados aceitáveis. Para ele, baixa consistência como essa sugere que os itens do teste produzem resultados com uma variabilidade significativa para a mesma amostra, o que aumenta o erro associado às medições. Essa interpretação reitera o entendimento de que os 15 itens de Física respondidos pelos alunos demonstram uma baixa precisão para a população examinada.

Embora Pasquali (2017) estabeleça esse limite mínimo de aceitação, a literatura demonstra que, em pesquisas envolvendo dados do ENEM, valores inferiores podem ser encontrados. Soares (2018), por exemplo, identificou um coeficiente Alpha de Cronbach inferior a 0,60 na prova do ENEM de 2009, considerando todos os itens. Além disso, a quantidade de itens afeta diretamente o valor do Alpha (Ferreira, 2018). Ferreira (2018) aponta que testes com poucos itens tendem a gerar um Alpha mais baixo, enquanto testes com mais itens tendem a aumentar o Alpha, embora isso não signifique um aumento na consistência interna. Freitas e Rodrigues (2005) também salientam que esse índice tende a aumentar à medida que mais itens são incluídos, desde que esses itens sejam representativos do mesmo conceito que está sendo medido.

Além disso, segundo Freitas e Rodrigues (2005), outro fator que pode influenciar o Alpha é a amostra, especialmente quando as pessoas que realizaram o teste fazem parte de uma amostra muito homogênea. Por exemplo, se todos os participantes de uma amostra escolherem a mesma resposta para uma questão, a variância será nula e, consequentemente, a confiabilidade calculada será zero (Freitas; Rodrigues, 2005).

IV.5 Análise dos Dados usando os Três Modelos Logísticos da TRI: ML1, ML2 e ML3

Bezerra (2023) conduziu a análise dos dados empregando três modelos logísticos da TRI: ML1, ML2 e ML3. A razão para isso reside na necessidade de compreender as diferenças entre esses modelos na análise dos dados, permitindo, posteriormente, uma comparação dos resultados obtidos em cada um deles.

O modelo ML1, também conhecido como modelo Rasch, parte da premissa de que todos os itens têm igual poder de discriminação, o que nem sempre condiz com a realidade das avaliações. Lord (1952) percebeu essa limitação do ML1 e, consequentemente, aprimorou o modelo, resultando no modelo de dois parâmetros (ML2), que posteriormente foi refinado com a incorporação da função logística (Coimbra, 2018; Birnbaum, 1957; Lord, 1952).

Mesmo ciente das limitações do ML1, foi realizada a análise dos itens nesse modelo para possibilitar a comparação dos resultados com os obtidos nos outros dois modelos logísticos. Em seguida, utilizando o pacote LTM, foi realizada a análise dos dados por meio do modelo logístico ML2, o qual incorpora dois parâmetros: a discriminação e a dificuldade.

O parâmetro de dificuldade, representado por 'b', denota a habilidade mínima que um indivíduo precisa para ter uma alta probabilidade de acertar o item. Nesse sentido, Ferreira (2018) esclarece que quanto maior o valor do parâmetro de dificuldade, mais difícil é o item, requerendo maior habilidade dos indivíduos para respondê-lo corretamente. Por sua vez, o parâmetro de "discriminação", representado por 'a', reflete a capacidade do item de distinguir entre indivíduos com diferentes níveis de proficiência. Esse valor se reflete na curva da Característica do Item (CCI). Curvas mais íngremes indicam maior poder de discriminação, enquanto curvas mais suaves indicam menor poder discriminativo do item (ver Gráfico 1).

Andrade, Tavares e Valle (2000) explicam que, teoricamente, os parâmetros de dificuldade (b) e discriminação (a) podem assumir qualquer valor real entre $-\infty$ e $+\infty$. No entanto, valores negativos no parâmetro de dificuldade são geralmente indesejados. Normalmente, o parâmetro b varia entre -4 e 4 (Ferreira, 2018), mas Tavares (2014) ressalta que esse limite pode ser ajustado pelo pesquisador. No caso do parâmetro de discriminação (a), embora possa variar de $-\infty$ a $+\infty$, espera-se que seja sempre positivo e não muito próximo de zero, indicando um alto poder de discriminação.

Bezerra (2023) também realizou a análise dos dados utilizando o modelo de três parâmetros (ML3). Neste modelo, além dos parâmetros "a" e "b" citados anteriormente, também é possível observar o índice de acerto ao acaso, representado pela letra "c" no modelo matemático. Este parâmetro de acerto ao acaso se refere à probabilidade de um indivíduo acertar um item mesmo que não possua a habilidade suficiente, portanto, sua variação vai de 0 a 1 (Fernández *et al.*, 1990). Por exemplo, um item com um parâmetro de acerto ao acaso de $c=0,10$ indica uma probabilidade de 10% de acerto do item por acaso (Sousa, 2019).

IV.6 Análise Gráfica dos Itens (AGI)

A AGI parte de dois pressupostos fundamentais. Primeiro, que um aluno que responde corretamente a um item sabe mais do que um aluno que escolhe uma alternativa incorreta. Já o segundo pressuposto parte do princípio de que um aluno que acerta mais itens sabe mais do que um aluno que erra mais. Assim, um aluno que responde corretamente a todos os itens possui probabilidade 1 de ter marcado a alternativa correta para cada item, enquanto um aluno que erra todos os itens possui probabilidade 0 de ter escolhido a resposta correta. Portanto, espera-se que a proporção de acertos aumente de 0 para 1 à medida que o escore total aumenta (Rodrigues, 2006; Braga, 2018; Costa, 2015).

Braga (2018) ressalta que a AGI é uma ferramenta valiosa para análises complementares na TCT e TRI em relação à qualidade dos itens. Essa análise é importante visto que, como explicado por Silva (2019c), tanto na TCT quanto na TRI, a expectativa é que os itens de um teste sejam respondidos corretamente por grupos de indivíduos com maior proficiência. Todavia, isso nem sempre ocorre devido à má elaboração dos itens ou à probabilidade de acerto ao acaso.

IV.7 Análise Fatorial

A Análise Fatorial baseia-se na premissa de que várias variáveis observáveis podem ser explicadas por um número menor de variáveis hipotéticas, chamadas de fatores não observáveis. Esses fatores são considerados a causa subjacente das relações entre as variáveis observáveis, e a análise fatorial permite explorar essas relações subjacentes (Rodrigues, 2006).

Assim, a fim de avaliar⁴ a unidimensionalidade dos itens de Física, Bezerra (2023) realizou uma análise fatorial confirmatória dos itens, considerando uma variável latente, a saber, o conhecimento de física. Para essa análise, o autor usou o estimador *Weighted Least Squares Means and Variance Adjusted* (WLSMV), por ser adequado para dados dicotômicos. Já a qualidade do ajuste do modelo foi avaliada por meio de índices como o *Tucker Lewis Index* (TLI), o *Comparative Fit Index* (CFI), o *Standardized Root Mean Squared Residual* (SRMR) e o *Root Mean Square Error of Approximation* (RMSEA). Além disso, a confiabilidade do modelo foi avaliada utilizando o coeficiente Alpha de Cronbach (α), Ômega de McDonald (ω) e confiabilidade composta (Li, 2016; Hu; Bentler, 1999).

É importante notar que embora o Alpha de Cronbach seja amplamente utilizado, estudos psicométricos recentes têm recomendado o uso de índices alternativos, uma vez que o Alpha assume que as cargas fatoriais são equivalentes em seu cálculo, o que nem sempre é o caso. Sendo assim, Bezerra (2023) realizou o cálculo do Ômega de McDonald e da confiabilidade composta, considerando que estas são alternativas melhores para tal avaliação (Mcneish, 2018; Flora, 2020).

V. Resultados e discussões

V.1 Análise dos dados utilizando a TCT

A análise empírica dos itens do teste foi executada empregando técnicas estatísticas, utilizando o software RStudio e os pacotes mencionados anteriormente. Bezerra (2023) avaliou diversos parâmetros que os itens do teste devem cumprir para terem adequação, no sentido de medir com precisão o que se propõem a medir (Rodrigues, 2016; Pasquali, 2017).

A proporção de erros e acertos nos 15 itens de Física do Enem 2019, conforme apresentado na Tabela 1, e observou-se que a proporção de respostas certas foi baixa (Bezerra, 2023). Surpreendentemente, quatro desses itens tiveram uma taxa de erro superior a 80%. Apenas dois itens, Q3 e Q9, obtiveram taxa de acerto superior a 30%, correspondendo, respectivamente, a 3.549 e 3.553 alunos do total de 11.624 participantes da amostra.

Constatou-se que a média de acertos nos itens de Física, entre os alunos de Santarém, foi de 22,56% (Bezerra, 2023). Embora este número seja baixo, é importante ressaltar que pesquisas anteriores que analisaram os itens do ENEM também apontaram para uma baixa

⁴Todas essas análises foram realizadas no ambiente de programação R, utilizando o pacote "lavaan" para a análise fatorial confirmatória de itens e o pacote "semTools" para avaliar a confiabilidade da variável latente do modelo.

porcentagem de acertos nas questões de Física. Por exemplo, Rubini (2019) calculou a média de acertos nas questões de Física nas edições do ENEM de 2009 a 2017, e a média encontrada para o período foi de 27,5%. Para o ano de 2013, a média foi de 21,3%, um valor menor do que o encontrado no ENEM de 2019 (Rubini, 2019). Embora se trate de grupos distintos, pois Rubini (2019) trabalha com a amostra nacional e nós trabalhamos com o município de Santarém (Bezerra, 2023), há consistência no baixo valor de média de acertos nos dois contextos.

Tabela 1 – Proporção de erros e acertos dos alunos de Santarém que prestaram o ENEM 2019 nos 15 itens com conteúdos relacionados à Física.

<i>c</i>	<i>PERCENTUAL DE ERRO</i>	<i>PERCENTUAL DE ACERTO</i>
<i>Q1</i>	84,97%	15,03%
<i>Q2</i>	77,15%	22,85%
<i>Q3</i>	69,48%	30,52%
<i>Q4</i>	76,70%	23,30%
<i>Q5</i>	86,32%	13,68%
<i>Q6</i>	74,54%	25,46%
<i>Q7</i>	85,68%	14,32%
<i>Q8</i>	77,80%	22,20%
<i>Q9</i>	69,47%	30,53%
<i>Q10</i>	77,60%	22,40%
<i>Q11</i>	71,54%	28,46%
<i>Q12</i>	76,55%	23,45%
<i>Q13</i>	79,00%	21,00%
<i>Q14</i>	82,86%	17,14%
<i>Q15</i>	71,96%	28,04%
<i>Média</i>	77,44	22,56

Fonte: Bezerra (2023)

A análise do coeficiente ponto-bisserial referente aos 15 itens de Física mostrou que os valores obtidos estão consideravelmente mais próximos de zero do que de um. Especificamente, dos 15 itens, 11 deles apresentaram coeficientes menores do que 0,30, um valor considerado o mínimo em muitas pesquisas. Todavia, mesmo assim, dois itens (Q5 e Q7) obtiveram coeficientes inferiores a esse limite.

É fundamental destacar que, de acordo com a literatura, quando esse coeficiente é negativo ou muito próximo de zero para um item específico, é possível interpretar que esse item não está cumprindo seu papel de distinguir entre os participantes. Não por acaso, esse coeficiente também é denominado “poder discriminativo do item” (PRIMI, 2012). Portanto, para uma prova do ENEM, era esperado encontrar índices maiores, sugerindo uma correlação mais significativa entre os itens.

Na continuidade da análise, o exame do coeficiente Alpha de Cronbach dos 15 itens de Física resultou em 0,1921. Freitas e Rodrigues (2005) sugerem que valores entre 0,75 e 0,60

são considerados moderados para esse parâmetro, enquanto os abaixo de 0,30 indicam confiabilidade “muito baixa”. Dessa forma, torna-se evidente que o coeficiente Alpha de Cronbach encontrado para o conjunto de itens está significativamente aquém do esperado, indicando uma baixa consistência no conjunto de itens analisado.

Portanto, após a análise inicial do Alpha de Cronbach, entendemos que o baixo coeficiente encontrado nesta pesquisa pode ser atribuído à quantidade reduzida de itens, uma vez que apenas 15 itens de Física foram utilizados, ou talvez à grande proporção de alunos que selecionaram a mesma alternativa para determinadas questões (Bezerra, 2023). Contudo, conduzimos testes adicionais com um maior número de itens, e embora tenha observado uma melhoria nesse coeficiente, ele ainda permaneceu abaixo das expectativas (Bezerra, 2023).

Portanto, é mais provável que a quantidade de alunos que escolheram as mesmas alternativas, juntamente com o número total de itens, tenha influenciado a redução do Alpha. Isso pode ser observado uma vez que dos 15 itens analisados, 13 deles apresentaram uma taxa de erro superior a 70%, e desses, quatro itens tiveram uma taxa de erro superior a 82%. Isso sugere que a maioria dos alunos optou por respostas incorretas e, sendo assim, muitas questões receberam uma resposta esmagadoramente unânime. Isso pode indicar uma amostra de alunos com níveis de habilidade semelhantes, ou a falta delas, na área do conhecimento em questão.

V.2 Análise Exploratória dos Itens por meio da TRI

Nesta etapa, conduzimos uma análise detalhada dos dados utilizando os modelos ML1, ML2 e ML3 da TRI (Bezerra, 2023). Inicialmente, foi realizada uma análise com o modelo logístico ML1, permitindo determinar a dificuldade de cada item na Prova de Física do ENEM 2019. Em seguida, com o modelo ML2, foi possível verificar tanto o índice de dificuldade quanto a capacidade de discriminação de cada item. Posteriormente, ao empregar o modelo ML3 para analisar os dados novamente, foi possível encontrar o índice de acerto ao acaso, a capacidade de discriminação e a dificuldade de cada item (ver Tabela 2).

Ao interpretar os resultados com base nos modelos logísticos ML1, ML2 e ML3 da TRI, mostrados na Tabela 2, torna-se evidente que, considerando a unidimensionalidade dos dados, esses itens demonstram uma capacidade limitada de discriminação. Na referida tabela, também fica evidente que os dois itens (Q5 e Q7) que apresentaram baixa consistência, conforme o coeficiente ponto-bisserial mostrado anteriormente, também apresentaram baixos índices de dificuldade e discriminação nas análises realizadas com os modelos logísticos de 2 e 3 parâmetros (ML2 e ML3). Além disso, é possível notar que a maioria dos itens não apresentou uma mudança significativa em seus índices de dificuldade e discriminação entre as análises conduzidas com o modelo ML2 e ML3. Isso sugere que, sob a perspectiva da TRI, esses itens da Prova de Física do ENEM 2019 não conseguiram avaliar adequadamente o conhecimento dos alunos, independentemente do modelo logístico adotado para calcular os escores.

Tabela 2 – Índices encontrados para os 15 itens de Física, do ENEM 2019, após análise usando os modelos ML1, ML2 e ML3 da TRI.

Item	TCT ponto- bisserial	TRI					
		ML1	ML2		ML3		
		Dificuldade	Dificuldade	Discriminação	Acerto ao acaso	Dificuldade	Discriminação
Q1	0,2777	1,923	3,322	0,552	0,1	2,546	1,429
Q2	0,3965	1,366	1,208	1,323	0,049	1,287	1,626
Q3	0,4146	0,931	0,92	1,109	0,168	1,169	2,665
Q4	0,2807	1,338	5,747	0,209	0,001	5,532	0,218
Q5	0,1743	2,041	-28,059	-0,066	0,051	-24,065	-0,096
Q6	0,3037	1,209	4,172	0,262	0,207	3,071	1,037
Q7	0,1345	1,984	-5,266	-0,348	0,058	-3,387	-0,747
Q8	0,2841	1,406	3,68	0,35	0,174	2,654	1,314
Q9	0,2848	0,931	10,047	0,082	0,006	9,121	0,093
Q10	0,3208	1,394	2,343	0,566	0,12	2,223	1,089
Q11	0,2858	1,041	7,734	0,12	0,044	7,141	0,153
Q12	0,2452	1,329	30,474	0,039	0,116	23,703	0,079
Q13	0,2882	1,484	5,712	0,235	0	5,205	0,258
Q14	0,2556	1,755	5,312	0,302	0,012	5,015	0,336
Q15	0,2844	1,064	7,026	0,135	0,268	2,67	2,43

Fonte: Bezerra (2023).

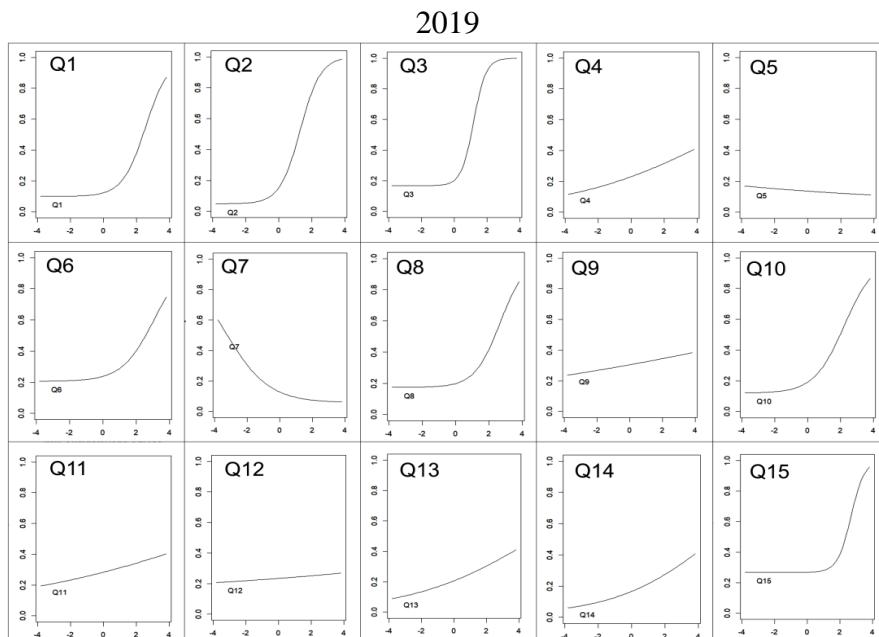
Vale ressaltar que em várias pesquisas que utilizam a TRI, são realizadas análises estatísticas comparativas para determinar qual dos modelos logísticos da TRI é mais adequado para os dados. Nesse sentido, usamos os três modelos, pois contempla o que está na literatura e, ao mesmo tempo, satisfaz o critério do Inep, que originalmente utiliza o modelo de três parâmetros (Bezerra, 2023). Dessa forma, demos ênfase ao ML3, pois é o modelo logístico usado pelo INEP e isso proveria melhor entendimento aos critérios adotados pelo instituto.

Além da análise usando os três modelos da TRI, foi analisado o gráfico da Curva Característica do Item (CCI) para as questões, conforme pode ser visualizado no Gráfico 2. Nesse tipo de gráfico, o eixo horizontal (eixo X) representa a habilidade medida pelo item, enquanto o eixo vertical (eixo Y) representa a probabilidade de resposta correta a esse item. E, conforme pode ser observado no Gráfico 2, vários itens não conseguiram cumprir adequadamente sua função em um exame como o ENEM. Isso ocorre porque, neste modelo, não é esperada uma inclinação decrescente nas curvas (Andrade; Tavares; Valle, 2000). Assim, uma inclinação decrescente indicaria que, conforme a habilidade dos alunos aumenta, a probabilidade de resposta correta diminui, o que, obviamente, não é o desejado. Precisamente

isso é observado nos itens Q5 e Q7. Curiosamente, esses são os mesmos itens que apresentaram discriminação negativa, conforme pode ser visualizado na Tabela 2.

No Gráfico 2, é possível observar que os três itens de Física com as maiores dificuldades, ou seja, Q12, Q9 e Q11, apresentam uma inclinação crescente. No entanto, essa inclinação não segue o padrão esperado; em vez disso, vê-se uma linha reta, o que evidencia a baixa capacidade de discriminação desses itens. Isso significa que tanto os alunos com maior habilidade quanto os com menor habilidade tinham probabilidades muito semelhantes de responder corretamente a esses itens, conforme explicado por Andrade, Tavares e Valle (2000).

Gráfico 2 – Curva característica do item referente à cada item da Prova de Física do ENEM.



Fonte: Bezerra (2023).

Diante dos resultados apresentados, especialmente as baixas taxas de discriminação e dificuldade, decidimos realizar novas análises, desta vez excluindo os itens Q5 e Q7 (Bezerra, 2023). O objetivo era avaliar se a consistência interna do teste melhoraria com a exclusão desses dois itens considerados problemáticos. Sendo assim, foi calculado novamente o coeficiente Alpha de Cronbach, que desta vez apresentou o valor de 0,2401 para os 13 itens. Isso representou uma melhoria em relação ao resultado anterior; no entanto, esse coeficiente ainda pode ser considerado muito baixo (Freitas; Rodrigues, 2005).

Após a exclusão desses itens, também foi realizada uma análise usando o ML3 da TRI e, embora tenha havido alguma melhora nos resultados, em relação aos anteriores, quando as questões Q5 e Q7 estavam presentes, esse incremento não foi tão significativo quanto o esperado, considerando a remoção dos dois itens com dificuldades negativas. Além do mais, os itens Q9 e Q12 continuaram apresentando altos índices de dificuldade e índices de discriminação abaixo de zero, juntamente com outros cinco itens que também mantiveram

índices de discriminação negativos. Assim, a exclusão desses dois itens não foi suficiente para melhorar substancialmente os resultados da análise conduzida com a TRI. Diante disso, surgiu a necessidade de novas análises para responder algumas perguntas. Entre elas, destacam-se duas: a) Os dados referentes às 15 questões de Física do ENEM 2019 não podem ser analisados usando modelos unidimensionais? b) Esses dados têm realmente uma consistência interna?

Para responder a essas perguntas, decidimos tomar as seguintes medidas: 1) realizar uma análise fatorial dos dados relativos às 15 questões de Física do ENEM 2019 e 2) ampliar a base de dados para analisar não apenas os 15 itens de Física, mas também os outros itens da prova de Ciências da Natureza, totalizando 45 questões (Bezerra, 2023). Desses, 15 eram os itens de Física. O objetivo dessa análise era determinar se os baixos índices encontrados estavam relacionados à separação dos itens de Física dos demais de Ciências da Natureza.

É importante enfatizar que os índices negativos de discriminação já foram apontados na literatura para questões de Física, assim como em outras questões relacionadas à prova de Ciências da Natureza. Rubini (2019), por exemplo, ao analisar a prova de Ciências da Natureza do ENEM de 2017, identificou que três itens de Física apresentaram valores de discriminação negativos ou próximos de zero. Segundo ele, a análise das curvas características dos itens revelou que, em média, as questões de Ciências da Natureza são consideradas difíceis quando analisadas pela TRI, corroborando os resultados com baixas taxas de acertos. Além disso, algumas questões com índices de discriminação negativos indicam que elas não se encaixam no modelo ML3 da TRI (Rubini, 2019). O autor chama a atenção para esse fato, uma vez que esses índices negativos indicam que os alunos com escores mais altos têm menor probabilidade de acertar as questões do que os alunos com escores mais baixos (Rubini, 2019), o que é contrário ao esperado em um teste que utiliza a TRI e, até mesmo, a TCT.

V.3 Análise dos 45 itens de Ciências da Natureza e suas Tecnologias do ENEM 2019 utilizando a TCT e a TRI

Conforme explicado anteriormente, os resultados obtidos usando a TCT e a TRI na análise dos 15 itens de Física do ENEM 2019 não foram satisfatórios. Na análise dos itens, foi encontrada uma alta proporção de erros, conforme mencionado anteriormente. Além disso, ao examinar a correlação ponto-bisserial, foram encontrados valores consideravelmente abaixo do esperado. Devido a isso, surgiu a necessidade de análises estatísticas adicionais a fim de compreender a causa dos coeficientes tão baixos, culminando na realização de uma análise fatorial, cujo resultado será apresentado posteriormente.

Por outro lado, a análise do coeficiente Alpha de Cronbach, como mencionado anteriormente, também apresentou índices baixos. Esses achados suscitaram uma análise mais aprofundada para determinar se a quantidade de itens estava impactando negativamente o Alpha, ou se os itens de Física do ENEM 2019 realmente apresentavam baixa consistência interna. Para responder a este questionamento, decidimos ampliar a base de dados, incluindo os

45 itens da prova de Ciências da Natureza e Suas Tecnologias, abrangendo não apenas as questões de Física, mas também de Química e Biologia (Bezerra, 2023).

É importante destacar que essa expansão na análise das questões de Ciências da Natureza visava principalmente verificar se o aumento no número de itens afetaria significativamente os índices sob a TCT e a TRI. O objetivo era determinar se a quantidade de itens ou o isolamento dos 15 itens de Física em um grupo maior de 45 questões poderia ser a causa dos índices insatisfatórios.

Cabe frisar que a prova de Ciências da Natureza aplicada no ENEM 2019 era composta por 45 itens que abrangiam tópicos de Física, Química e Biologia, como indicado no item 4.3 do Edital do ENEM 2019 (Brasil, 2019a).

Inicialmente, constatou-se ser notável que a proporção de erros é substancialmente alta, em linha com as observações feitas ao analisar exclusivamente as questões de Física (Bezerra, 2023). Em média, a proporção de erros na prova de Ciências da Natureza atinge 73,35%, enquanto a proporção de acertos é de 26,65%. Comparativamente, a prova de Física registra uma proporção de acertos de 22,56% e uma de erros de 77,44%.

Inicialmente, foi analisado o coeficiente Alpha de Cronbach para os 15 itens de Física, resultando em um valor de 0,1921, o que indica baixa consistência interna. Para fins comparativos, o coeficiente foi calculado também para os 45 itens da prova de Ciências da Natureza, obtendo-se um valor de 0,5556. Essa diferença pode estar associada ao maior número de itens, fator que, segundo Ferreira (2018) e Freitas (2005), tende a aumentar o Alpha, embora não necessariamente signifique maior consistência interna. No entanto, como o objetivo deste estudo é a análise aprofundada do desempenho dos alunos especificamente em Física, a discussão segue focada na interpretação dos índices psicométricos dos itens dessa disciplina, buscando compreender os desafios enfrentados pelos estudantes nessa área.

É importante observar que Vizzotto (2022) realizou uma análise do coeficiente Alpha de Cronbach para as provas de Ciências da Natureza realizadas no período de 2009 a 2019. Para essa análise, o autor utilizou os dados de participantes de todas as edições do Enem nesse período, aplicando filtros específicos para delimitar uma amostra nacional. Após a aplicação desses filtros, a amostra utilizada em sua pesquisa consistiu em 6.981.815 participantes para as onze edições do ENEM (de 2009 a 2019). A média do coeficiente Alpha de Cronbach para esse período foi de 0,667. No entanto, nos anos de 2013 e 2018, esse coeficiente caiu abaixo de 0,60. Portanto, esses resultados indicam que a prova de Ciências da Natureza nesses dois anos apresentou baixa consistência interna (Freitas; Rodrigues, 2005).

Vale destacar que o valor do coeficiente Alpha de Cronbach encontrado por Vizzotto (2022) para o ano de 2019 difere daquele obtido em nossa pesquisa devido à utilização de amostras diferentes (Bezerra, 2023). Haja vista que, enquanto Vizzotto baseou sua análise em uma amostra nacional, consideramos apenas os dados dos alunos de Santarém (Bezerra, 2023). Contudo, a pesquisa de Vizzotto (2022) demonstra que, em média, a prova de Ciências da Natureza não apresenta alta consistência interna, o que está de acordo com o que obtivemos.

Ao analisar a correlação ponto-bisserial dos 45 itens da prova de Ciências da Natureza, observamos que esse coeficiente continua em níveis baixos, em linha com os resultados da análise dos itens de Física isoladamente (Bezerra, 2023). Observou-se que dos 45 itens, apenas 11 apresentaram coeficientes superiores a 0,30, um valor frequentemente utilizado em trabalhos acadêmicos sobre TCT e TRI, enquanto 19 questões alcançaram coeficientes iguais ou superiores a 0,20, considerados o mínimo aceitável (Bezerra, 2023).

Em resumo, os resultados revelam uma consistência interna inadequada nos itens da prova de Ciências da Natureza, em especial nos itens de Física, tanto na análise utilizando a TCT quanto a TRI. Essa falta de consistência pode comprometer a avaliação das habilidades dos alunos e, por conseguinte, a validade da prova do ENEM 2019 nesse aspecto. Os achados também corroboram as conclusões de Vizzotto (2022) em relação à consistência interna geral das provas de Ciências da Natureza ao longo de vários anos do ENEM.

A análise dos resultados sob a perspectiva da TRI revela que, apesar do aumento na quantidade de itens durante a pesquisa, não houve uma melhora substancial nos índices de dificuldade e discriminação em comparação com os resultados da análise anterior que se concentrou apenas nas 15 questões de Física. Em vez disso, alguns itens continuaram exibindo índices de discriminação negativos, indicando que a probabilidade de um aluno acertar esses itens é a mesma, independentemente de suas habilidades. Isso também se aplica aos índices de dificuldade.

Para concluir a análise da prova de Ciências da Natureza, realizamos uma comparação entre as médias das notas dos alunos de Santarém que concluíram o ensino médio em escolas públicas ou privadas (Bezerra, 2023). No entanto, constatou-se que no momento da inscrição para o exame, o preenchimento deste campo não era obrigatório, resultando na falta de informações completas. Sendo assim, dentre os inscritos, apenas 186 alunos declararam ser provenientes de escolas particulares, 2153 declararam ter frequentado escolas públicas, enquanto mais de 9 mil não informaram o tipo de escola. Devido a essa falta de dados, tornou-se inviável realizar uma comparação significativa considerando todas as escolas.

Dado o resultado insatisfatório da comparação entre escolas públicas e privadas, o autor decidiu explorar um aspecto mencionado na tese de Nascimento (2019): a condição socioeconômica. Sendo assim, foram analisados os dados referentes ao “Questionário Socioeconômico” do ENEM 2019, especificamente a pergunta 006: “*Qual é a renda mensal de sua família? (Some a sua renda com a dos seus familiares)*”. Observou-se que para responder a esta pergunta, os alunos podiam escolher entre diferentes faixas salariais, desde “A” (nenhuma renda) até “Q” (R\$ 19.960,00).

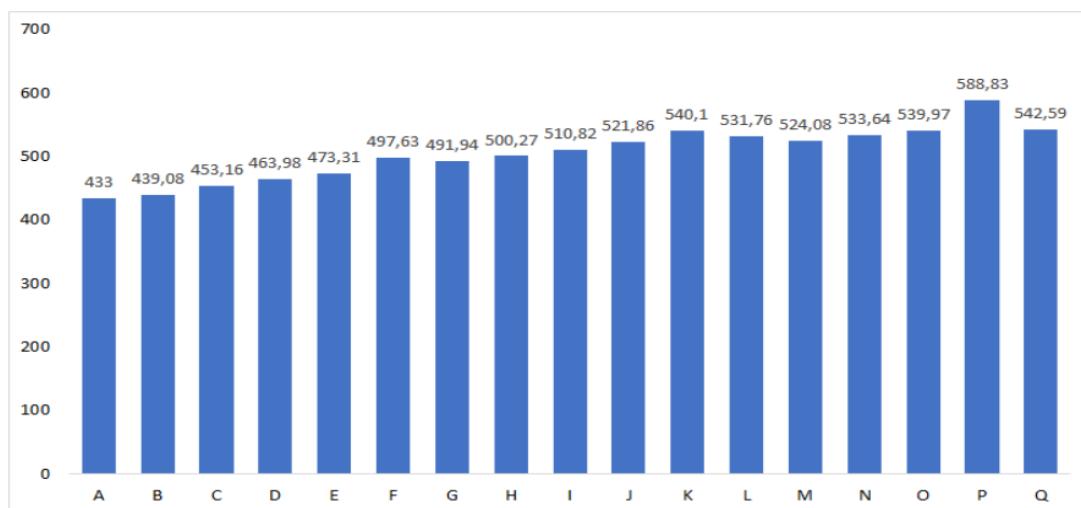
Após a execução da análise das respostas dos alunos à referida pergunta, observou-se que a maioria dos alunos se enquadrava nas faixas B e C, com renda familiar variando de R\$ 998,00 a R\$ 1.996,00. Ao comparar as extremidades do gráfico, ficou evidente que a média dos alunos que declararam não ter renda foi de 433, enquanto a média daqueles com renda superior

a R\$ 19.000,00 atingiu 542. Embora a diferença possa parecer pequena, fica claro no Gráfico 3 que as notas em Ciências da Natureza aumentam proporcionalmente à faixa salarial.

É importante frisar que a relação entre o desempenho dos estudantes no ENEM e sua condição socioeconômica é um aspecto que merece atenção. Kleinke (2017) realizou uma pesquisa abrangente sobre este tema, analisando todos os itens de Física respondidos pelos alunos nas edições do ENEM de 2011, 2012 e 2014. Seu estudo revelou uma correlação entre o desempenho dos estudantes em Física e seu status socioeconômico. As médias dos alunos pertencentes ao topo da classe alta eram até o dobro da média dos alunos na classe baixa.

Kleinke (2017) também observou diferenças nos coeficientes Alpha de Cronbach quando considerou a classe social dos participantes. Esses resultados sugerem que o construto das avaliações no ENEM reflete uma formação educacional que favorece a classe alta, enquanto as classes baixa e média enfrentam desafios na compreensão e resolução das provas. Tais resultados destacam a necessidade urgente de reformulações no ensino e avaliação em Física. Nesse sentido, Kleinke (2017) destaca a importância de professores, pesquisadores e elaboradores considerarem essas informações para melhorar o ensino e a avaliação da formação em Física dos estudantes.

Gráfico 3 – Nota dos alunos de Santarém, na Ciências da Natureza do ENEM 2023, por faixa salarial.



Fonte: Bezerra (2023)

Diante dos resultados obtidos nesta pesquisa, alinhados com as conclusões de Kleinke (2017), é razoável supor que a média da prova de Ciências da Natureza para os alunos de Santarém e o coeficiente Alpha de Cronbach possam ter sido influenciados pela condição socioeconômica dos estudantes. No entanto, para validar essas suposições, pesquisas adicionais são necessárias, abordando a relação entre o desempenho no ENEM e vários fatores socioeconômicos e educacionais. Kleinke (2017) destaca a importância de uma análise aprofundada dessas informações para melhorar o ensino e a avaliação da Física, reconhecendo

a necessidade de equidade educacional e a superação das disparidades socioeconômicas que afetam o desempenho dos alunos.

V.4 Resultados da análise factorial dos 15 itens de Física do ENEM 2019

Os resultados da análise factorial mostraram que o modelo unidimensional não foi rejeitado, uma vez que os seguintes valores foram obtidos: χ^2 [df] = 141.577[90], CFI = 0.965, TLI = 0.960, SRMR = 0.20 e RMSEA = 0.007[IC90%, 0.005, 0.009]. Sendo assim, os valores encontrados estão acima do que convencia a literatura, onde rejeita-se um modelo quando TLI ou CFI foi inferior à 0.90 ou o RMSEA for superior à 0.08 ou ainda se o SRMR for maior que 0.06" (Bezerra, 2023, p. 97). Sendo assim, o modelo não foi rejeitado.

Ao observar as cargas fatoriais dos itens, foi possível identificar duas cargas fatoriais moderadas (Q1 e Q10) e duas fortes (Q2 e Q3). Por sua vez, as cargas dos itens 5 e 12 não foram estatisticamente significativas ($p > 0,05$) e os demais itens apresentaram cargas fatoriais baixas ou desprezíveis. Considerando os valores absolutos das cargas fatoriais, as estatísticas descritivas foram: média = 0,20 (DP = 0,18), mínimo = 0,02 e máximo = 0,62.

Por fim, concluímos que os resultados da Análise Fatorial demonstraram que "o modelo de uma dimensão não foi rejeitado, contudo, não é possível afirmar que a dimensão identificada é robusta, na medida em que os itens apresentaram cargas fatoriais baixas e o fator não exibiu confiabilidade aceitável" (Bezerra, 2023, p. 98).

VI. Considerações finais

Essa pesquisa realizou uma análise sobre o desempenho dos alunos de Santarém na prova de Ciências da Natureza e suas Tecnologias do ENEM 2019, com foco específico nos itens relacionados à disciplina de Física. A análise quantitativa e qualitativa dos itens de Física apontou para um baixo índice de acerto, uma preocupação que se estende nacionalmente, corroborando dados apresentados por Rubini (2019), onde a média de acertos dos itens de Física de 2009 a 2017 foi de apenas 27.5%.

Ressaltamos que o baixo desempenho em Física, e em Ciências da Natureza como um todo é preocupante, mas não é exclusividade dos alunos de Santarém, uma vez que outras pesquisas como as de Marcom e Kleinke (2016), Marcom (2019) e Vizzotto (2022), indicam baixos índices de acerto nas questões relacionadas à Ciência da Natureza e suas Tecnologias em âmbito nacional.

A expansão da análise para todos os itens da prova de Ciências da Natureza do ENEM 2019, empregando técnicas da TCT e TRI, permitiu tirar conclusões importantes. A ideia era verificar se ao aumentar a quantidade de itens, a consistência da prova melhoraria e isso poderia indicar que os baixos índices encontrados na TRI poderiam ter sido causados devido à baixa quantidade de itens (quando foram analisados somente os 15 itens de Física) ou devido à análise isolada destes itens. Contudo, a pesquisa com os 45 itens até apresentou melhorias no

coeficiente Alpha de Cronbach, porém os demais índices permaneceram baixos, sugerindo limitações no poder discriminativo e na consistência interna da prova. Por sua vez, a análise fatorial exploratória realizada, buscou identificar a dimensionalidade dos itens de Física e confirmou unidimensionalidade do modelo, porém, com confiabilidade questionável devido a cargas fatoriais baixas.

Bezerra (2023) destaca a necessidade de uma análise socioeconômica a partir dos dados dos inscritos no ENEM, uma vez que a pesquisa indicou uma correlação entre o desempenho dos alunos em Ciências da Natureza e suas Tecnologias e a condição financeira. Essa descoberta alinha-se com pesquisas anteriores de Kleinke (2017) e Marcom (2019), que sugerem uma influência significativa das condições socioeconômicas nas notas dos alunos, especialmente nos itens de Física.

Em síntese, os resultados da pesquisa dessa pesquisa mostram o baixo desempenho dos alunos de Santarém ao responderam os itens de Física do ENEM 2019 e sugerindo investigações futuras para dirimir algumas questões que foram surgindo durante as análises dos dados, especialmente da relação entre o desempenho do aluno e a sua condição socioeconômica.

Referências bibliográficas

- ALCHIERI, J. C. **Avaliação psicológica**: conceito, métodos e instrumentos. [S.l.]: Casa do Psicólogo, 2003.
- ALVES, V. C.; STACHAK, M. **A importância de aulas experimentais no processo ensino-aprendizagem em física**: eletricidade. In: SIMPÓSIO NACIONAL DE ENSINO DE FÍSICA, XVI, 2005, Rio de Janeiro. p. 1-4.
- ANDRADE, D. F.; TAVARES, H. R.; VALLE, R. da C. **Teoria da Resposta ao Item**: conceitos e aplicações. [S.l.: s.n.], 2000.
- ANDRADE, J. M. de; LAROS, J. A.; GOUVEIA, V. V. O uso da teoria de resposta ao item em avaliações educacionais: diretrizes para pesquisadores. **Avaliação Psicológica**, v. 9, p. 421-435, dez. 2010. ISSN 1677-0471.
- ARAUJO, A. L. S. O. *et al.* **Teoria de Resposta ao Item**. SBC, 2020. Disponível em: https://metodologia.ceie-br.org/wp-content/uploads/2018/10/cap2_8.pdf. Acesso em: 30 dez. 2023.
- BARROSO, M. F.; RUBINI, G.; SILVA, T. da. Dificuldades na aprendizagem de física sob a ótica dos resultados do ENEM. **Revista Brasileira de Ensino de Física**, (SciELO), v. 40, n. 4, jun. 2018.

BEZERRA, R. dos S. **O desempenho em física dos alunos de Santarém (PA) que prestaram o ENEM 2019:** análise psicométrica utilizando a teoria clássica dos testes e a teoria de resposta ao item. 2023. 158 f. Dissertação (Mestrado em Educação) – Universidade Federal do Oeste do Pará, Santarém.

BIRNBAUM, A. Efficient design and use of tests of mental ability for various decision-making problems. School of Aviation Medicine, USAF., 1957.

BRAGA, M. B. P. **Escala de Proficiências em Concepções Térmicas:** Diagnóstico Psicométrico de Estudantes em Portugal e Brasil. 2018. Tese (Doutorado) - Universidade de Coimbra, Portugal.

BRASIL. Ministério da Educação, Instituto Nacional de Estudos e Pesquisas Educacionais.

INEP. Documento Básico ENEM. [S.l.: s.n.], 1998.

BRASIL. Edital n. 14, de 21 de março de 2019. Diário Oficial da União, Brasília, DF, 2019. ISSN 1677-7069.

BRASIL. Ministério da Educação, Instituto Nacional de Estudos e Pesquisas Educacionais.

INEP. 2019. Acesso em: 17 ago. 2020.

BRASIL. Ministério da Educação, Instituto Nacional de Estudos e Pesquisas Educacionais.

INEP. Método de correção das provas reconhece o conhecimento e a coerência do participante. 2019. Disponível em: <https://www.gov.br/inep/pt-br/assuntos/noticias/enem/metodo-de-correcao-das-provas-reconhece-o-conhecimento-e-a-coerencia-do-participante>. Acesso em: 19 jun. 2020.

BRASIL. Ministério da Educação, Instituto Nacional de Estudos e Pesquisas Educacionais.

INEP. Exame Nacional de Desempenho dos Estudantes (Enade), 2020. Disponível em: <https://www.gov.br/inep/pt-br/areas-de-atuacao/avaliacao-e-exames-educacionais/enade>. Acesso em: 01 ago. 2020.

BRASIL. Ministério da Educação, Instituto Nacional de Estudos e Pesquisas Educacionais.

INEP. Histórico. 2020. Acesso em: 14 mar. 2022.

BRASIL. Ministério da Educação, Instituto Nacional de Estudos e Pesquisas Educacionais.

INEP. Saeb. 2020. Acesso em: 07 ago. 2020.

BRASIL. Ministério da Educação, Instituto Nacional de Estudos e Pesquisas Educacionais. MEC lança novo Saeb com avaliação para todos os anos a partir do 2º do ensino fundamental. 2020. Acesso em: 03 ago. 2020.

BRASIL. Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira. Exame Nacional do Ensino Médio – ENEM: Procedimentos de análise [recurso eletrônico]. Brasília, DF: Inep, 2021.

BRASIL. A segunda maior prova de acesso ao ensino superior do mundo - MEC. 2022. Acesso em: 05 fev. 2022.

BRITTON, E.D; SCHNEIDER, S. A. Large-scale assessments in science education. [S.l.: s.n.], 2007. p. 1007-1040.

COHEN, A. Organizational commitment and turnover: A meta-analysis. **Academy of Management Journal**, v. 36, n. 5, p. 1140–1157, 1993.

COIMBRA, P. S. B. **Um olhar reflexivo sobre a prática pedagógica de docentes da matemática na educação básica de Santarém**. 2018. Dissertação (Mestrado) - Universidade Federal do Oeste do Pará.

COLLARES, C. F.; GREC, W. L. P.; MACHADO, J. L. M. Psicometria na garantia de qualidade da educação médica: conceitos e aplicações. **Sci Health**, v. 3, n. 1, p. 33-49, 2012.

COSTA, R. Q. G. da. **Análise da prova da primeira fase da OBMEP como subsídio para orientar a prática docente**. 2015. Dissertação (Mestrado) - Universidade de Brasília, 2015.

CRONBACH, L. J. Coefficient alpha and the internal structure of tests. **Psychometrika**, Springer, v. 16, n. 3, p. 297-334, 1951.

EMBRETSON, S. E; REISE, S. P. **Item response theory**. [S.l.]: Psychology Press, 2000.

FERNÁNDEZ, J. M. *et al.* **Teoría de respuesta a los ítems**: un nuevo enfoque en la evaluación psicológica y educativa. [S.l.]: Pirámide, 1990.

FERNÁNDEZ, J. M. *et al.* **Introducción a la teoría de respuesta a los ítems.** [S.l.]: Pirámide, 1997.

FERREIRA, E. A. **Teoria de Resposta ao Item TRI – análise de algumas questões do ENEM:** habilidades 24 a 30. 2018. Dissertação (Mestrado) – Universidade Federal de Grande Dourados.

FLORA, D. B. Your coefficient alpha is probably wrong, but which coefficient omega is right? A tutorial on using R to obtain better reliability estimates. **Advances in Methods and Practices in Psychological Science**, Sage Publications Sage CA, Los Angeles, v. 3, n. 4, p. 484-501, 2020.

FREITAS, A. L. P.; RODRIGUES, S. G. **A avaliação da confiabilidade de questionários: uma análise utilizando o coeficiente alfa de Cronbach.”** [S.l.]: Psychology Press, 2005.

FRIAS, J. L. D. Uma ferramenta para a obtenção e análise de dados do ENEM. 2015. Dissertação (Mestrado) - PUC-Rio.

GOMES, C. M. A.; GOLINO, H. F.; PERES, A. J. de S. Fidedignidade dos escores do exame nacional do ensino médio (ENEM). **Psico**, EDIPUCRS, v. 51, n. 2, p. e31145, fev. 2020. Disponível em: <https://revistaseletronicas.pucrs.br/revistapsico/article/view/31145>.

GONÇALVES JUNIOR, W. P; BARROSO, M. F. As questões de física e o desempenho dos estudantes no ENEM. **Revista Brasileira de Ensino de Física**, v. 36, p. 1-16, 2014.

GONÇALVES JUNIOR, W. P.; BARROSO, M. F. **Enem:** Os itens e o desempenho dos estudantes em 2009. 2012. 220 f. Dissertação (Mestrado do Programa de Ensino de Física – Instituto de Física, UFRJ.

GUEWEHR, K. Teoria da resposta ao item na avaliação de qualidade de vida de idosos. 2007. Dissertação (Mestrado) – UFRGS.

HU, L.; BENTLER, P. M. Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. **Structural Equation Modeling: A Multidisciplinary Journal**, v. 6, n. 1, p. 1-55, 1999.

IBGE. INSTITUTO BRASILEIRO DE GEOGRAFIA E ESTATÍSTICA - População/Santarém. 2022. Disponível em: <https://cidades.ibge.gov.br/brasil/pa/santarem/panorama>. Acesso em: 18 jan. 2022.

KLEINKE, M. U. Influência do status socioeconômico no desempenho dos estudantes nos itens de física do Enem 2012. **Revista Brasileira de Ensino de Física**, v. 39, n. 2, 2017.

LI, C.-H. **Confirmatory factor analysis with ordinal data**: Comparing robust maximum likelihood and diagonally weighted least squares. [S.l.: s.n.], 2016.

LORD, F. **A theory of test scores**. [S.l.: s.n.], 1952.

LORD, F. M. **Applications of Item Response Theory to Practical Testing Problems**. [S.l.]: Routledge, 1980.

MARCOM, G. S. **O ENEM, Indicadores Formativos e Ensino de Física**. 2019. Tese (Doutorado) - Unicamp, SP.

MARCOM, G. S.; KLEINKE, M. U. Análises dos distratores das questões de física em exames de larga escala. **Caderno Brasileiro de Ensino de Física**, Universidade Federal de Santa Catarina, v. 33, n. 1, p. 72-91, 2016.

MARCOM, G. S.; KLEINKE, M. U. Indicadores formativos para o ensino de física através do exame nacional do ensino médio (ENEM). **Caderno Brasileiro de Ensino de Física**, Universidade Federal de Santa Catarina, v. 38, n. 3, p. 1388-1419, dez. 2021. Disponível em: <https://doi.org/10.5007/2175-7941.2021.e77046>. Acesso em: 15 mai. 2022.

MATSUNAGA, L. H. **O que é psicometria?** 2022. Disponível em: <https://ibpad.com.br/comunicacao/o-que-e-psicometria>. Acesso em: 15 mai. 2022.

MCDONALD, R. P. **Test theory**: A unified treatment. [S.l.]: Psychology Press, 2013.

MCNEISH, D. Thanks coefficient alpha, we'll take it from here. **Psychological methods**, American Psychological Association, v. 23, n. 3, p. 412, 2018.

MENEGOTTO, J. C.; ROCHA FILHO, J. B. da. Atitudes de estudantes do ensino médio em relação à disciplina de física. **Revista Electrónica de Enseñanza de las Ciencias**, v. 7, n. 2, p. 298-312, 2008.

MOREIRA, C. R. B. Psicometria: limitações e avanços. In: **Discursos Interdisciplinares por uma Educação Transformadora**. FAMEN, 2020. p. 123-134. Disponível em: <https://doi.org/10.36470/famen.2020.l4c11>. Acesso em: 10 jan. 2021.

MUNIZ, F. **Teoria Clásica de Los Tests**. [S.l.]: Pirámide, 1994.

NARDI, R. **A pesquisa em ensino de física**. [S.l.]: Escrituras, 1998.

NASCIMENTO, M. M. O professor de Física na escola pública estadual brasileira: desigualdades reveladas pelo Censo escolar de 2018. **Revista Brasileira de Ensino de Física**, v. 42, p. e20200187, 2020.

NASCIMENTO, M. M. O acesso ao ensino superior público brasileiro: um estudo quantitativo a partir dos microdados do Exame Nacional do Ensino Médio. 2019. Tese (Doutorado) UFRGS.

NESI, E. R. *et al.* Perspectivas e desafios atuais no ensino de física. **Brazilian Journal of Development**, v. 7, n. 2, p. 17285-17298, 2021.

PASQUALI, L. **Teoria e métodos de medida em ciências do comportamento**. [S.l.: s.n.], 1996.

PASQUALI, L. **Psicometria**: teoria e aplicações: a teoria clássica dos testes psicológicos. [S.l.]: Ed. da UnB, 1997.

PASQUALI, L. Psychometrics. **Revista da Escola de Enfermagem**, USP, v. 43, n. 1, p. 992-999, 2009.

PASQUALI, L. **Psicometria**: teoria dos testes na psicologia e na educação. [S.l.]: Editora Vozes Limitada, 2017.

PASQUALI, L.; PRIMI, R. Fundamentos da teoria da resposta ao item: Tri. Avaliação Psicológica. **Interamerican Journal of Psychological Assessment**, v. 2, n. 2, p. 99-110, 2003.

PRIMI, R. Psicometria: fundamentos matemáticos da teoria clássica dos testes. **Avaliação Psicológica**, Instituto Brasileiro de Avaliação Psicológica, v. 11, n. 2, p. 297-307, 2012.

QEDU. Pará: Aprendizado | QEdu. 2022. Disponível em: <https://qedu.org.br/uf/15-para/> aprendizado. Acesso em: 01 out. 2022.

QUARESMA, E. de S. A. **Modelagem para construção de escalas avaliativas e classificatórias em exames seletivos utilizando teoria da resposta ao item uni e multidimensional.** 2014. Tese (Doutorado) - Universidade de São Paulo.

QUARESMA, E. de S. A.; DIAS, C. T. dos S.; BARBETTA, P. A. **Escalas avaliativas e classificatórias em exames seletivos:** uma abordagem utilizando a teoria da resposta ao item. 2019.

R CORE TEAM. R (4.2). [Computer software]. R Foundation for Statistical Computing. [S.1.], 2022. Disponível em: <https://cran.r-project.org/bin/windows/base/>. Acesso em: 01 jan. 2022.

RODRIGUES, M. M. Proposta de análise de itens das provas do sob a perspectiva pedagógica e a psicométrica. **Estudos em Avaliação Educacional**, v. 17, n. 34, p. 43-78, 2006.

RODRIGUES, R. **Aplicação da teoria da resposta ao item na avaliação das habilidades matemáticas em alunos do ensino médio de uma escola de Belo Horizonte.** 2016. Dissertação (Mestrado) - Pontifícia Universidade Católica de Minas Gerais.

ROSA, C. W.; ROSA, Á. B. Ensino de física: objetivos e imposições no ensino médio. **Revista Electrónica de Enseñanza de las Ciencias**, v. 4, n. 1, 2005.

RUBINI, G. M. **O que o Enem revela sobre a Aprendizagem em Física na Educação Básica.** 2019. Tese (Doutorado) - Universidade Federal do Rio de Janeiro, Rio de Janeiro.

SALSA, I. da S.; RAMOS, I. C. de O.; SANTOS, R. B. **Qual o recado dos erros em questões de estatística descritiva no Enem 2016 para a educação básica?** In: CONEDU, VI, 2019, Fortaleza.

SANTARÉM. História de Santarém | Portal de Turismo de Santarém. 2022. Disponível em: <https://turismo.santarem.pa.gov.br/historia-de-santarem>. Acesso em: 25 de jan. 2022.

SANTOS, C. L. *et al.* O impacto da pandemia na aprendizagem da matemática nas turmas de 9º ano de 2021 da rede municipal de Canindé. **Revista Missionária**, v. 24, n. 1, p. 21-33, 2022.

SILVA, A. C. da. **Analise Psicométrica e Pedagógica da prova da primeira fase da Obmep respondida por um grupo de alunos do 6º e 7º ano do ensino fundamental.** 2019. Monografia (TCC) - Universidade Federal do Oeste do Pará, Santarém, PA.

SILVA, W. L. G. **Avaliação em larga escala como política do estado**: um estudo comparativo entre a Teoria Clássica dos Testes e a Teoria da Resposta ao Item na Olimpíada Brasileira de Matemática das Escolas Públicas (OBMEP). 2019. Dissertação (Mestrado) - Universidade Federal do Oeste do Pará, Santarém, PA.

SILVA, W. N. da. Análises e Reflexões sobre os objetos do conhecimento de matemática aplicados no Enem de 2017 no município de Santarém-Pa. 2019. Dissertação (Mestrado) - Universidade Federal do Oeste do Pará, Santarém, PA.

SOARES, D. J. M. **Teoria clássica dos testes e teoria de resposta ao item aplicadas em uma avaliação de Matemática básica**. 2018. Dissertação (Mestrado) - Universidade Federal de Viçosa.

SOARES, D. J. M. ; SOARES, T. E. A.; DOS SANTOS, W. Análise da qualidade psicométrica da prova de matemática do Exame Nacional do Ensino Médio brasileiro de 2018. **Actualidades Investigativas en Educación**, v. 21, n. 1, p. 86-115, 2021.

SOUSA, L. A. de. **Análise comparativa do exame nacional do ensino médio (ENEM) via teoria clássica dos testes e teoria de resposta ao item**. 2019. Tese (Doutorado) - Universidade Federal de Viçosa.

SOUSA SOBRINHO, A. de; RODRIGUES, A. P. A física no exame nacional do ensino médio (ENEM). **Acta Tecnológica**, v. 14, n. 2, p. 57-73, 2019.

SOUZA, T. C. F. **Avaliação do ensino de física**: um compromisso com a aprendizagem. [S.l.: s.n.], 2002.

SUDBRACK, E. M.; COCO, E. M. Avaliação em larga escala no Brasil: potencial indutor de qualidade? **Roteiro**, Editora Unoesc, v. 39, n. 2, p. 347-370, 2014.

VILLAR, R. P. Os itens de física do ENEM classificados por análise fatorial exploratória. 2021. Tese (Doutorado) - Unicamp.

VIZZOTTO, P. A. As “piores” questões de física do ENEM: Uma análise psicométrica das edições de 2009 a 2019. **Revista Brasileira de Ensino de Física**, v. 44, 2022.



Direito autoral e licença de uso: Este artigo está licenciado sob uma [Licença Creative Commons](#)