



As “piores” questões de Física do ENEM: Uma análise psicométrica das edições de 2009 a 2019

The “worst” items in Physics of ENEM: A psychometric analysis of the 2009–2019 editions

Patrick Alves Vizzotto*¹

¹Universidade Federal do Sul e Sudeste do Pará, Instituto de Ciências Exatas, 68500-000, Marabá, PA, Brasil.

Recebido em 17 de maio de 2022. Aceito em 18 de junho de 2022.

O desempenho no Exame Nacional do Ensino Médio (ENEM) pode ser usado como critério de seleção para Instituições de Ensino Superior (IES) públicas e para obter financiamento e bolsas em muitas IES privadas. Considerando o potencial de influência do exame no presente e no futuro de uma pessoa no que tange a aspectos pessoais, profissionais e sociais, espera-se que a prova consiga realizar uma seleção justa. Um dos critérios técnicos para esse ideário é a presença de características mínimas de validade e confiabilidade na avaliação. Esse artigo apresenta uma pesquisa que buscou verificar indicadores psicométricos de qualidade do exame. Analisou-se os Microdados do ENEM para as edições de 2009 a 2019, classificando a qualidade dos itens de Física como bons, duvidosos e ruins. Como resultados, observou-se que a maioria dos itens foram considerados não adequados para mensurar a proficiência que se propõe medir na prova de Ciências da Natureza. Espera-se que os resultados apresentados possam contribuir para a reflexão sobre a qualidade de exames em larga escala. Sugere-se que trabalhos com escopo na análise pedagógica dos itens possam ser realizados, a fim de compreender os pontos a se melhorar e os caminhos que devem ser percorridos para promover a qualidade do ENEM.

Palavras-chave: Avaliação, Ensino de Física, Psicomетria, Microdados.

The performance in the National High School Exam (ENEM) can be used as a selection criterion for public Higher Education Institutions (HEIs) and to obtain funding and scholarships in many private HEIs. Considering the potential for influence of the examination on the present and future of a person, regarding personal, professional and social aspects, it is expected that the test can perform a fair selection. One of the criteria for this idea is that the examination has minimal characteristics of validity and reliability. This article presents research that sought to verify psychometric indicators of the quality of the test. The Microdata of the ENEM was analyzed for the editions from 2009 to 2019, classifying the quality of the items as good, dubious and bad. As results, it was observed that most items were considered not adequate to measure the proficiency that is proposed to be measured in the Nature Sciences test. It is expected that the results presented can contribute to the reflection on the quality of large-scale examinations. It is suggested that works with scope in the pedagogical analysis of items considered not suitable can be performed in order to understand the points to be improved and the paths that must be traveled to promote the quality of the examination.

Keywords: Evaluation, Physics Teaching, Psychometrics, Microdata.

1. Introdução

O Exame Nacional do Ensino Médio (ENEM) já foi considerado um dos maiores do mundo [1]. Pode ser classificado como um exame de larga escala, mesmo que a participação na prova seja de natureza voluntária. Criado para ser um instrumento de avaliação do Ensino Médio, com o passar dos anos, teve sua estrutura e finalidade modificadas. Atualmente, a grande adesão de participantes à prova se deve, principalmente, ao fato de que a nota final nessa avaliação pode assegurar o ingresso na maioria das Instituições do Ensino Superior (IES) públicas do país todo, assim como, pode ser requisito para pleitear vagas e subsídios financeiros para custear uma graduação em diferentes instituições privadas também [2].

Na prática, um bom resultado no ENEM pode permitir ao participante uma série de ganhos que conseguem moldar o seu presente e o seu futuro, no que tange a sua formação profissional, a sua qualidade de vida e a sua ascensão social.

Tendo em vista o impacto que o exame pode ter na vida dos participantes, salienta-se a importância da sua qualidade, seja na segurança da sua produção e distribuição, seja na elaboração das questões que compõem a prova. Sobre esse último fator, destaca-se o cuidado que se deve existir para ser garantido que o exame meça aquilo que se propõe. Sobre esse particular, entre outras coisas, defende-se que a prova deve ter certo grau de qualidade psicométrica [3].

A psicomетria é um campo de estudo atribuído à psicologia, que aborda, de modo geral, o processo de elaboração e validação de testes, de modo a estipular e assegurar os parâmetros de validade e confiabilidade de

* Endereço de correspondência: patrickvizzotto@gmail.com

instrumentos usados [3]. A Educação é uma das áreas que fazem uso de conhecimentos da psicometria e os transpõe para esse contexto, aplicando-os no controle de qualidade dos diferentes testes e avaliações presentes no cotidiano escolar, em especial os de larga escala [4].

No cenário internacional, um referencial psicométrico de grande importância para as áreas da educação e da psicologia é o “*Standards for educational and psychological testing*” [5], pois mantém o estado da arte da área sobre os parâmetros psicométricos nos quais os testes psicológicos e educacionais deveriam seguir para que se atestem indícios de qualidade dos instrumentos.

Os exames em larga escala seguem uma extensa metodologia de validação, tendo como base muitos dos critérios psicométricos preconizados pela literatura. Testes como o *Programme for International Student Assessment* (PISA), o Exame Nacional de Desempenho dos Estudantes (ENADE), a Prova Brasil, a *Scholastic Aptitude Test* (nos Estados Unidos), *Gaokao* (na China), entre tantos outros, seguem diferentes metodologias de validação, todas com vistas a certificar que seus instrumentos medem o que se propõem mensurar.

Tendo em vista o impacto que um exame em larga escala como o ENEM pode acarretar ao participante, a preocupação de se ter um teste válido e confiável e as diferentes diretrizes que prescrevem os parâmetros de qualidade dos testes, questiona-se: quais são as características psicométricas dos itens de Física do ENEM?

Para tanto, o objetivo desse artigo é analisar a qualidade psicométrica dos itens de Física do ENEM para os anos de 2009 a 2019. De maneira específica, se buscou: 1) acessar os Microdados do ENEM de 2009 a 2019; 2) selecionar os dados referentes apenas aos itens de Física de cada edição; 3) estimar, via Teoria Clássica de Testes e Teoria de Resposta ao Item, diferentes características psicométricas dos itens de Física; 4) classificar os itens de Física segundo a sua qualidade; e 5) apresentar, para cada edição, o item de Física com a mais baixa qualidade psicométrica aferida.

Para esse estudo considerou-se um item de Física aquela questão da prova de Ciências da Natureza que aborda inteiramente assuntos de Física, ou, ainda, um item de Ciências da Natureza que necessita mobilizar saberes de Física para analisar as alternativas de resposta da questão.

Uma pesquisa com esse escopo mostra-se relevante para a literatura pois auxilia a reflexão sobre a qualidade do ENEM a partir de uma análise de dados reais, de modo a averiguar se os aspectos preconizados para aferir a qualidade do exame são observados no comportamento empírico dos seus itens. Não obstante, tendo em vista o número de participantes de cada edição e o potencial de moldar a vida pessoal e profissional de milhões de cidadãos todos os anos, espera-se que a prova atenda aos requisitos mínimos de validade e fidedignidade para poder se aproximar do ideário de uma avaliação justa,

inclusiva e que cumpra o seu objetivo fim, de maneira efetiva.

A análise de itens do ENEM sob os mais diferentes vieses metodológicos pode ser encontrada na literatura nacional. Destaca-se, no contexto geral, os estudos de Travitzki [1, 6], que abordam o conceito de qualidade do exame e analisam diferentes aspectos das 4 provas de 2009 a 2011; Pontes Junior e colaboradores [7] que investigaram a qualidade psicométrica de alguns itens de Educação Física das edições de 2009 a 2013; Gomes, Golino e Peres [8] que investigaram a fidedignidade dos escores obtidos pelos participantes na edição de 2011; Gonçalves e Almeida que analisaram os indicadores de dificuldade e discriminação das provas de matemática da edição de 2012 [9]; e também Soares, Soares e Santos [2] que analisam as medidas de tendência central dos itens das edições de 2016 a 2018.

No contexto do Ensino de Física pode-se destacar trabalhos como os de Gonçalves Junior e Barroso [10] que analisaram aspectos qualitativos de itens de Física das edições de 2009 a 2011. Entre outras particularidades, os autores notaram que o percentual de acertos nas questões é majoritariamente baixo e os itens que exigem do respondente conhecimentos disciplinares ou raciocínios matemáticos apresentam uma performance considerada de baixo desempenho.

Para Bassalo [11], que analisou aspectos qualitativos da prova de Ciências da Natureza de 2009, quase a totalidade dos itens de física foram construídos com enunciados contendo “pegadinhas”, com informações desnecessárias e, ainda, faltando dados importantes para a resolução do item. Além disso, o autor notou também itens com mais de uma resposta correta.

Já Lopes [12], analisou as questões e os desempenhos dos alunos no ENEM de 2011 e notou que não se observa um alinhamento entre temas comumente abordados no Ensino Médio e os tópicos abordados nos itens da prova. Para o autor, a presença de questões que tecem sobre temas pouco abordados no Ensino Médio pode auxiliar a entender o fato de que a maior parte das questões de física foram consideradas de alta dificuldade para essa edição.

Por sua vez, Marcom e Kleinke [13, 14] analisaram os distratores das edições de 2009 a 2012 [13] e relacionaram o desempenho dos participantes da edição de 2014 em função do sexo [14]. Os autores observaram, neste último estudo, que em 67% dos itens de Física os homens tiveram desempenho superior do que o das mulheres.

Carlos [15] que analisou o desempenho dos estudantes em física das edições de 2009 a 2014, notou que a maior parte dos participantes do ENEM pode ser classificado, para as edições analisadas, nos três níveis de desempenho mais baixos. Para o autor, isso mostra a necessidade de acompanhar o ensino da Física nas escolas e fomentar políticas públicas que atuem pedagogicamente para mitigar os efeitos de tal realidade.

Lemos e Hernandes [16], que realizaram uma análise de conteúdo nas questões de Física dos anos 2011 a 2015, observaram que não existe uma uniformidade no percentual de itens das grandes áreas da Física nas provas analisadas. Não obstante, os autores também notaram a grande quantidade de itens sobre assuntos envolvendo Dinâmica e Eletrodinâmica e questões de análise gráfica.

Barroso, Rubini e Silva [17] analisaram as dificuldades na aprendizagem em física a partir da análise das provas do ENEM entre 2009 e 2014. Os autores observaram dificuldades consideradas permanentes no entendimento de conhecimentos básicos de mecânica, terminologia e óptica geométrica, permeadas, muitas das vezes, pela presença de concepções não científicas.

Duarte, Gonçalves Junior e Barroso [18] analisaram pedagogicamente um item do ENEM de 2009 considerado de alto grau de dificuldade. Como resultados, os autores observaram que ele exigia um conhecimento demasiadamente específico de física e apresentava, para obter êxito na sua resolução, uma ênfase na memorização de uma relação matemática que expressava uma grandeza física.

Rodrigues [19] analisou as questões de Física do ENEM de 2017 via taxonomia de Bloom e observou que a maioria dos itens podem ser classificados sob o domínio cognitivo da “análise”, segundo o qual os alunos necessitariam ter condições de classificar o conteúdo estudado, deduzir as equações dos teoremas conhecidos e diferenciar as informações.

Nascimento, Cavalcanti e Ostermann [20] estudaram a relação de itens dos anos de 2009, 2012 e 2015 com aspectos socioeconômicos. Os autores encontraram indícios de que questões de cunho quantitativo estão mais relacionadas a níveis socioeconômicos do que questões com foco qualitativo, privilegiando estudantes com elevado índice de capital.

Silveira, Barbosa e Silva [21] e Silveira, Stilck e Barbosa [22] apontaram problemas nas questões de Física e trouxeram para o debate a importância da qualidade dos itens para a validade do exame, uma vez que é grande a quantidade de questões com falhas, seja no seu enunciado, seja nas alternativas de resposta, ou ainda na necessidade demasiada de uma interdisciplinaridade acrítica.

Por fim, destaca-se também a tese de Rubini [23] que analisou as provas e os desempenhos dos participantes do ENEM dos anos de 2009 a 2017. O autor observou, entre outros aspectos, a presença de concepções não científicas na escolha das alternativas, baixas habilidades de leitura e de interpretação gráficos e tabelas, além do desempenho aquém do esperado em itens que exigiam cálculos matemáticos e resolução de problemas.

Esse autor realizou também uma análise de cunho psicométrico nos itens de Ciências da Natureza até a edição de 2017, representando, para a presente pesquisa, juntamente com o estudo de Travitzki [1], um

referencial teórico de grande importância. Com isso, nesse manuscrito, foi possível avançar no estado da arte, apresentando à literatura da área, para itens de Física, as análises psicométricas também das edições de 2018 e 2019. Ressalta-se que, até a conclusão desse estudo, os Microdados do ENEM para o ano de 2020 e 2021 ainda não estavam disponíveis na página eletrônica do INEP.

Por fim, pode-se notar, ao analisar as produções já existentes, o interesse crescente de pesquisadores com a qualidade do exame. Diferentes frentes de análise podem ser identificadas de modo a trazer ao debate a reflexão sobre diferentes aspectos associados com a qualidade, seja da prova e dos itens, seja do impacto que o exame tem na vida dos participantes e do sistema educacional. Isso sinaliza que o presente artigo pode auxiliar a somar forças nesse debate, analisando aspectos específicos que podem atuar na certificação de qualidade das provas e dos próprios itens.

Sendo assim, na sequência, explana-se sobre os referenciais que deram suporte ao trabalho.

2. Fundamentação Teórica

2.1. A qualidade psicométrica de um teste

O conceito de qualidade pode assumir diferentes significados, a depender do ponto de vista do referencial adotado. Um teste em larga escala, por exemplo, pode ser uma das maneiras de se avaliar a educação, embora o uso de avaliações em larga escala – principalmente as internacionais – no contexto brasileiro, seja objeto de importantes críticas na literatura [6]. Não obstante, foi para esse fim que o ENEM foi criado, há mais de duas décadas.

Na literatura, há autores que corroboram com a importância de testes como esses para tal objetivo [2, 7], em simultâneo, há trabalhos que, por outro lado, abordam um contraponto, mostrando os pontos fracos dessa categoria de avaliação [20–23]. Em ambos os espectros considera-se a realidade e reconhece-se a importância que o ENEM tem para o país, com as suas qualidades e seus pontos a melhorar. A partir dessa postura, justifica-se a preocupação de assegurar que esse tipo de avaliação tenha certa qualidade mínima para fazer cumprir suas metas.

Conforme referido na seção anterior, para atestar a qualidade de uma prova como o ENEM, a literatura da área pode se basear em critérios estabelecidos por diferentes referenciais teóricos e metodológicos. As recomendações referendadas ao redor do mundo recorrem a uma área denominada psicometria [3, 4], esfera que estuda aspectos intrínsecos das pessoas e, para isso, estuda indicadores que representam tais constructos. De maneira prática, entre outras coisas, a psicometria ocupa-se da elaboração e validação de testes para os mais diversos fins. Esses saberes são transpostos para

o contexto educacional e, na literatura, é possível encontrar uma série de produções que visam abordar as normatizações com vistas a assegurar a qualidade de um teste nos mais diferentes setores da educação [5].

Das técnicas mais antigas às mais modernas, um conjunto de análises e procedimentos são realizados para determinar os parâmetros das provas e dos itens. Considera-se que a qualidade dos dados coletados por um questionário, um teste, uma prova, etc., dependem, parcialmente, das suas características psicométricas. Normalmente essas características são aferidas por procedimentos estatísticos sofisticados.

2.2. Validade e confiabilidade

Em suma, a literatura sempre aborda dois conceitos fundamentais no que tange à avaliação de uma avaliação: a validade e a confiabilidade. Por validade, entende-se a característica que atesta se o instrumento em questão mensura aquilo que se destina medir. Já a confiabilidade busca verificar a hipótese de que, se o mesmo instrumento fosse aplicado a uma mesma pessoa em intervalos de tempo diferentes, esta apresentaria desempenhos estatisticamente semelhantes. Ou seja, validade refere-se mais a atestar a coerência entre a medida, o seu objetivo a medir e a confiabilidade, relaciona-se mais com a reprodutibilidade do instrumento [3, 4].

Considera-se que, em pesquisas com seres humanos, o critério preconizado pela confiabilidade pode se tornar de difícil aferição, visto que uma pessoa pode ser influenciada por diferentes fatores no intervalo entre um teste e outro, o que inviabilizaria a premissa. Em função disso, diferentes procedimentos estatísticos buscam suprir essa dificuldade e fornecer indicadores que representem aspectos da confiabilidade do instrumento. Um dos testes mais utilizados é o coeficiente Alfa de Cronbach¹ [3]. Em geral, a confiabilidade analisa o teste por inteiro, enquanto há outros indicadores psicométricos que analisam item a item.

Já para aferir a validade de um instrumento ou de um conjunto de itens, busca-se uma avaliação a partir de especialistas da área e do público-alvo para o qual o teste se destina [3]. Há uma série de procedimentos para a aferição da validade, sendo o índice de validade de conteúdo [24] e o coeficiente de validade de conteúdo [25] alguns deles. Basicamente, quando se cria um novo instrumento, os especialistas e o público-alvo avaliam a qualidade de cada item em aspectos como a adequação semântica, o entendimento, a interpretação, as alternativas de respostas, entre outros. Os melhores itens são classificados e os piores, eliminados. Os que ficam participam de etapas posteriores, que podem ser testes piloto ou outros procedimentos estatísticos.

¹ Embora já existam outros testes com maior acurácia para esta verificação, como o ômega de McDonald, que não considera que cada item tenha a mesma importância para o instrumento como um todo.

2.3. O padrão empírico das questões

Diferentes métodos para averiguar o comportamento dos itens podem ser empregados para a avaliação de testes. Existem duas categorias comumente utilizadas para analisar a qualidade de testes e itens: a Teoria Clássica de Testes (TCT) e a Teoria de Resposta ao Item (TRI).

A TCT busca mensurar o desempenho de um participante a partir do total de acertos em um teste, ou seja, considera-se o escore final para determinar as características de qualidade. Para conferir qualidade ao teste, é necessário considerar o instrumento na totalidade. Tais procedimentos seguem pressupostos como a imposição de que itens iguais sejam respondidos por todos os participantes, em iguais circunstâncias e com uma quantidade mínima de respondentes [2].

No entanto, apenas com o escore bruto não é possível realizar uma comparação adequada entre indivíduos que obtiveram bons escores em uma prova fácil e pessoas com baixos desempenhos em uma prova difícil, por exemplo. Em função de limitações como essa, mostra-se necessária a existência de indicadores que afirmem a dificuldade dos testes e dos itens, bem como a capacidade do item discernir aqueles participantes que possuem daqueles que não possuem o conhecimento averiguado no teste, indicador chamado discriminação.

Já a TRI, no contexto educacional, busca averiguar a proficiência da pessoa para determinado conhecimento e não apenas um escore, como o total de respostas corretas [2]. De modo geral, é medido um traço latente (habilidade), aspecto intrínseco que se manifesta por meio da resposta de alguns itens e representam indicadores desse traço latente. Ou seja, a TRI se fundamenta na hipótese de a pessoa acertar ou errar uma questão segundo o seu conhecimento sobre o assunto [1].

A TRI não substitui a TCT, elas fazem parte de procedimentos complementares, com seus potenciais e suas limitações. De modo geral, indicadores iguais, aferidos independentemente pelas duas teorias, tendem a apresentar um comportamento semelhante, como os índices de dificuldade e discriminação do item, por exemplo [2]. A superioridade da TRI, dentre outros aspectos, centra-se na possibilidade de comparar proficiências de diferentes pessoas e grupos através de diferentes instrumentos [1].

O ENEM, até o ano de 2008, recorria à TCT. A partir de 2009, o exame passou por uma reformulação e a TRI veio em substituição como método de validação dos itens e estimativa do desempenho dos participantes. O modelo usado no ENEM é o de 3 parâmetros, no qual é descrita a probabilidade de um respondente acertar uma questão dependendo da sua proficiência. Esse modelo, como o nome diz, tem como condição conhecer 3 parâmetros sobre o item: o parâmetro “a” corresponde à discriminação do item; o “b” é a sua dificuldade; e o “c” corresponde a chance de uma pessoa com baixo desempenho responder à questão de maneira

correta. Para cada item espera-se que o parâmetro “a” seja alto. Já para o parâmetro “b” deseja-se que a distribuição seja equilibrada, não tendo itens fáceis nem difíceis como maioria; em simultâneo, deseja-se que o parâmetro “c” seja o menor possível (<20% para um item com 5 alternativas de resposta) [1].

Por fim, considera-se a TCT e a TRI como ferramentas que podem auxiliar na resposta à pergunta de pesquisa. Assim, a seção seguinte irá expor quais testes serão empregados para aferir a qualidade dos itens de Física, bem como quais critérios devem ser usados para interpretar os resultados.

3. Metodologia

3.1. Caracterização da pesquisa

Essa é uma pesquisa de abordagem quantitativa, de natureza básica, com objetivos de pesquisa descritiva e exploratória, que investiga bancos de dados e os analisa através de técnicas de estatística [26, 27].

3.2. Os Microdados do ENEM

Os Microdados consistem em arquivos que contêm informações sobre as mais diversas avaliações ou pesquisas. Podem conter informações sobre as questões, os desempenhos, dados de caracterização social dos participantes, as alternativas assinaladas, etc. Qualquer cidadão pode ter acesso aos Microdados por meio do site do INEP (<https://www.gov.br/inep/pt-br/area-de-informacao/dados-abertos/microdados/enem>). Os arquivos são compilados de modo a serem acessados através de determinados softwares estatísticos. No contexto nacional, o INEP é o maior compilador de Microdados voltados para pesquisas na educação, gerando informações sobre o ENEM, a Prova Brasil, o censo da Educação Superior, o Censo Escolar, o PISA (brasileiro), entre outros [1].

3.3. Recorte temporal

Os dados analisados foram obtidos a partir dos Microdados do ENEM do ano de 2009 até 2019. O recorte temporal escolhido para a pesquisa se justifica porque, a partir de 2009, o ENEM passou por uma reformulação de grande monta, cujas principais mudanças foram: a inserção de uma nova matriz de referência; a mudança na estrutura da prova; o acréscimo da TRI como metodologia de validação e geração da nota final dos candidatos; entre outros, como: o aumento do uso do exame para o ingresso no Ensino Superior público, financiamento estudantil e bolsas de estudo para instituições privadas, etc.

3.4. Indicadores de qualidade via TCT e TRI

A fim de estabelecer critérios para aferir a qualidade das provas e dos itens, os seguintes indicadores foram

utilizados. Essa seção apresentará um breve conceito sobre o teste e a diretriz sobre como ele deve ser interpretado.

Índice de Dificuldade: uma das maneiras de verificar a dificuldade de um item, via TCT, é observar a média de acertos dessa questão. Itens com baixa média de acertos são considerados difíceis, enquanto itens com alta taxa de acertos, são considerados fáceis. De acordo com Vilarinho [24], questões com dificuldade igual ou superior a 0,7 podem ser consideradas fáceis; entre 0,7 e 0,3 são de dificuldade média; e abaixo de 0,3 devem ser consideradas difíceis. Já quando este indicador é observado a partir da TRI, por meio do parâmetro “b”, há autores que defendem que os valores adequados para dificuldade do item devem estar entre -3 e $+3$ [2] ou -4 e $+4$ [1].

Índice de Discriminação: Afere o quanto um item consegue diferenciar pessoas com diferentes níveis de proficiência. Em uma avaliação sobre um determinado constructo, indivíduos com habilidade devem alcançar pontuações diferentes daquelas pessoas sem habilidade. Esse índice será analisado nesta pesquisa por meio do parâmetro “a” da TRI. Para interpretá-lo, será usada a seguinte classificação: discriminação boa se o parâmetro “a” for maior ou igual a 0,5; discriminação duvidosa se for entre 0,2 e 0,5; e discriminação ruim se for abaixo de 0,2 [1]. Índices de discriminação negativos poderiam sinalizar que a probabilidade de responder corretamente uma questão diminui com o aumento da proficiência [7]. Baixo valor de discriminação significa que, tanto pessoas com baixa habilidade quanto os com alta habilidade, tem probabilidades semelhantes de acertar o item. Quanto maior o valor de discriminação da questão, maior será a contribuição dela para a medida da habilidade (também chamada “proficiência”).

Coefficiente Alfa de Cronbach: é uma medida de confiabilidade do teste. Relaciona a correlação entre cada item e a prova toda, ou seja, o quanto cada questão contribui para a prova, de modo geral. A sua premissa considera que os itens de uma prova são formas paralelas de se mensurar o conhecimento que se deseja medir. Para interpretá-lo, o teste gera um valor entre 0 e 1, sendo os valores mais próximos de 1, indicativos de maior consistência interna. Não há consenso sobre um valor mínimo para atribuir uma consistência satisfatória. Comumente, usa-se valores a partir de 0,5, 0,6, 0,7. Para essa pesquisa usar-se-á o valor de 0,6, conforme indicado por estudos da área [3, 4].

Correlação item total: é uma medida do item. Indica o quanto ele se correlaciona com o escore total do instrumento integral. Itens com alto nível desse indicador sugerem que contribuem fortemente para a consistência interna do instrumento. Por outro lado, itens com baixo valor, provavelmente estão contribuindo para uma menor de consistência interna do instrumento. Em uma validação de questionário, geralmente sugere-se que itens com

baixo valor de correlação item total sejam eliminados do instrumento, pois, sem eles, a confiabilidade geral tende a melhorar [1]. A correlação item total foi calculada incluindo todos os 45 itens da prova de Ciências da Natureza.

Coefficiente de correlação bisserial: verifica se aquelas pessoas que obtiveram um desempenho satisfatório no teste, tenderam a assinalar as alternativas corretas dos itens. O cálculo desse indicador acontece através de uma análise de correlação entre o escore das pessoas e a taxa de escolha de cada alternativa. Espera-se que o valor de correlação seja positivo para a alternativa correta e negativo para todas as outras alternativas erradas (também chamadas “distratores”). Esse coeficiente tem potencial de identificar itens com problemas em sua formulação ou com erros no gabarito, pois se a correlação for negativa para a alternativa correta, pode sugerir que o item teve mais acertos vindos de pessoas com baixo desempenho. Ao mesmo tempo, se algum distrator tiver correlação positiva, sinaliza que, por alguma razão, tal alternativa está atraindo mais o grupo de pessoas com bom desempenho. Isso pode configurar um problema de discriminação do item, sendo sugerida a sua remoção ou adequação. Para interpretar esse indicador, nessa pesquisa, será considerado um item bom se a alternativa correta apresentar uma correlação bisserial igual ou acima de 0,3; duvidoso se o valor for entre 0,15 e 0,30; e ruim se o índice for abaixo de 0,15. Também será classificado como duvidoso o item que apresentar valores positivos para qualquer distrator e/ou se a alternativa correta demonstrar o coeficiente com valor negativo [1].

Acerto ao acaso: aponta a probabilidade de uma pessoa não proficiente no teste acertar o item por acaso. Se esse valor for maior que zero, significa que, para alguns respondentes, houve “chute”. Ou seja, o parâmetro sinaliza que houve respostas corretas de indivíduos que não teriam habilidade de conhecer a resposta correta [2, 6]. Para o caso do ENEM, a literatura costuma determinar como valores aceitáveis aqueles abaixo de 20% (0,2), pois é uma prova em que cada item possui 5 alternativas de resposta [26].

Ajuste do modelo: é um indicador que mostra se o modelo de 3 parâmetros da TRI (empregado no ENEM) se ajusta aos dados analisados, fornecendo as informações de discriminação, dificuldade e potencial de acerto casual de maneira adequada. De modo geral, diz-se que um ajuste não satisfatório não consegue garantir que os parâmetros obtidos sejam invariantes. Ou seja, um modelo com um ajuste inadequado não fornecerá essas informações com precisão, prejudicando a estimativa dos parâmetros, bem como, da proficiência analisada. O teste de ajuste do modelo gerará um p-valor para cada item que deve ser estatisticamente significativo para sinalizar um ajuste adequado do modelo, ou seja, o p-valor deve ser menor que 0,05. Itens com p-valor

abaixo de 0,05 são considerados bons; já quando o p-valor estiver no intervalo entre 0,05 e 0,10, o item será definido como duvidoso; por fim, um p-valor acima de 0,10 representa um item com ajuste considerado não adequado [1].

Curva Característica do Item (CCI): Os parâmetros analisados na TRI podem ser observados graficamente por meio da CCI. Ela é uma curva em forma de “S”, em que a escala do eixo horizontal corresponde a habilidade que o item mede e o eixo vertical representa a probabilidade de os respondentes sem a habilidade assinalarem a resposta correta ao acaso. Um item com discriminação alta apresentará uma curva tendendo ao formato de “S”. Discriminação baixa, a curva tenderá a um formato menos acentuado. Quanto mais à esquerda do gráfico, menor é a dificuldade do item. Por outro lado, quanto mais à direita, maior é a sua dificuldade. O parâmetro “c”, por fim, pode ser observado analisando em que ponto do eixo vertical a curva e o eixo se cruzam.

Curva de Informação do Item (CII): sabe-se que um item não discrimina igualmente para toda a faixa de habilidade medida. Assim, a CII observa a precisão que o item tem para os diferentes níveis de habilidade. Ou seja, possibilita analisar o quanto uma questão possui de informação psicométrica para a aferição da habilidade. É uma curva que auxilia a verificação do quanto um item contribui para mensurar a habilidade que o teste se propõe a medir, nas diferentes faixas dessa habilidade. A CII está diretamente relacionada com a CCI, sendo que, quanto maior for a inclinação da CCI, maior será a CII [25].

3.5. Classificação dos itens

Os itens considerados pertencentes a Física foram aqueles que abordaram inteiramente assuntos de Física, ou, ainda, aqueles itens de Ciências da Natureza que necessitavam mobilizar saberes de Física para analisar as alternativas de resposta da questão. De acordo com essa regra, cada item de Ciência da Natureza foi avaliado e classificado de acordo.

De modo a analisar o comportamento empírico das questões de Física do ENEM e classificá-las dentro de um padrão de qualidade, usou-se alguns dos indicadores anteriormente mencionados. A classificação aplicada nesta pesquisa seguirá a mesma estrutura proposta por Travitzki [1], que pode ser observada na Tabela 1.

Após a análise de cada questão, reunir-se-á os itens em uma classificação global, baseada na proposta de Travitzki [1], que os rotulará em: **Item duvidoso** (indicador global): quando o item for considerado duvidoso em pelo menos 3 dos 4 indicadores, ou ruim em pelo menos 1 indicador. **Item ruim** (indicador global): quando for ruim em pelo menos 2 dos 4 indicadores. Se os itens não se encaixarem nesses parâmetros, então significa que foram considerados bons.

Tabela 1: Critérios para classificação dos itens.

| Classificação do item | TCT | | TRI | |
|-----------------------|------------------------------|-------------------------------|--------------------------------|------------------------|
| | Correlação item total (CORR) | Coefficiente bisserial (BISS) | Parâmetro de discriminação (A) | Ajuste do modelo (FIT) |
| Bom | $CORR \geq 0,30$ | $BISS \geq 0,30$ | $a \geq 0,5$ | $FIT < 0,05$ |
| Duvidoso | $0,15 < CORR < 0,30$ | $0,15 < BISS < 0,30$ | $0,2 < a < 0,5$ | $0,05 < FIT < 0,10$ |
| Ruim | $CORR \leq 0,15$ | $BISS \leq 0,15$ | $a \leq 0,2$ | $FIT \geq 0,10$ |

Fonte: Travitzki [1].

Tabela 2: Número de inscritos e amostra selecionada para cada edição.

| Ano | Inscritos | Amostra | Ano | Inscritos | Amostra | Ano | Inscritos | Amostra |
|-------------|-----------|---------|-------------|-----------|-----------|-------------|-----------|---------|
| 2019 | 5.095.270 | 512.707 | 2015 | 7.746.427 | 336.720 | 2011 | 5.380.856 | 817.224 |
| 2018 | 5.513.747 | 543.571 | 2014 | 8.722.248 | 1.030.138 | 2010 | 4.626.094 | 727.174 |
| 2017 | 6.731.341 | 305.701 | 2013 | 7.173.563 | 1.011.596 | 2009 | 4.148.720 | 571.247 |
| 2016 | 8.627.367 | 349.597 | 2012 | 5.791.065 | 776.140 | | | |

Fonte: Microdados do INEP.

Adicionalmente, para algumas discussões desta pesquisa, classificar-se-á os itens como adequados e não adequados. Os itens adequados são aqueles considerados como bons e os não adequados, por sua vez, são todos aqueles sendo classificados como duvidosos ou como ruins.

3.6. Softwares e filtragem dos dados

Para acessar os Microdados do ENEM e calcular os indicadores psicométricos necessários utilizou-se o software R (estatística) [28], com os pacotes “*mirt*” [29] para os cálculos da TRI e o “*psych*” [30] para as análises via TCT. Tabelas e gráficos foram elaborados através do Microsoft Excel.

Ao somar o número de inscritos no ENEM nas 11 edições analisadas chega-se ao número de 69.556.698 pessoas. Esse quantitativo, demasiadamente grande, mostra a necessidade de realizar-se um recorte amostral nos dados. De modo a definir critérios para esse recorte, foram aplicados os seguintes filtros, a fim de analisar somente os dados daqueles que: 1) eram participantes da primeira aplicação; 2) responderam todos os itens; 3) estudaram em escola regular; 4) concluíram ou estavam concluindo o Ensino Médio no ano em que fizeram a prova; 5) estiveram presentes em todas as provas da edição; 6) receberam o caderno azul.

Assim, conforme pode ser analisado na Tabela 2, o recorte acarretou uma redução substancial do número de dados analisados, fixando a amostra a um total de 6.981.815 participantes.

Os parâmetros da TRI foram estimados para cada amostra (de cada ano analisado) individualmente. Considerando que o INEP não disponibiliza, para cada edição, os parâmetros de cada item, realizou-se uma análise comparativa com outro resultado da literatura, a fim de verificar a coerência dos dados aqui gerados.

Para isso, destacou-se o já mencionado trabalho de Rubini [23], que estimou os parâmetros da TRI do

ENEM das edições de 2009 a 2017, utilizando a população inteira de participantes de cada edição, ou seja, respondentes de todas as cores das provas.

Já que no presente trabalho optou-se por uma análise amostral, realizar uma comparação (por meio de uma análise paramétrica de correlação) entre os parâmetros dessas duas análises independentes mostrou-se uma metodologia adequada para verificar a confiabilidade dos parâmetros calculados para esta pesquisa.

Dessa maneira, observou-se que os parâmetros dos itens da TRI (dos anos de 2009 a 2017) não são idênticos, mas quase todos são razoavelmente equivalentes, com pequenas diferenças que podem ser supostas como sendo causadas pela diferença na amostra utilizada (correlação de Pearson para o parâmetro a: 0,994, sig. <0,000; para o parâmetro b: 0,998, sig. <0,000; para o parâmetro c: 0,990, sig. <0,000). Assim, esse resultado ajudou a atestar a coerência dos parâmetros gerados para o presente trabalho.

Destaca-se também que os itens de Ciências da Natureza foram classificados em itens de Física, Química e Biologia de acordo com a rotulação realizada por Rubini [23].

Ainda, destaca-se que outros indicadores poderiam ter sido empregados na análise da qualidade dos itens, sendo as opções aqui usadas fruto de um recorte. Não foram identificadas, nos gabaritos disponibilizados nos Microdados, questões de Física que tenham sido anuladas.

A seguir, a seção dos resultados apresentará o relatório das características psicométricas dos itens de Física.

4. Resultados

Os resultados obtidos serão particionados em três seções: a primeira, abordando algumas características gerais das 11 provas; a segunda, apontando o resultado da qualidade dos itens para cada prova, incluindo em que classificação cada item foi alocado; e, por fim, na terceira seção, o item com o conjunto de qualidades psicométricas

Tabela 3: Consistência interna e dificuldade das provas.

| Ano | Alfa de Cronbach | Dificuldade da prova de CN | Ano | Alfa de Cronbach | Dificuldade da prova de CN | Ano | Alfa de Cronbach | Dificuldade da prova de CN |
|------|------------------|----------------------------|------|------------------|----------------------------|------|------------------|----------------------------|
| 2019 | 0,601 | 0,28 | 2015 | 0,658 | 0,27 | 2011 | 0,746 | 0,33 |
| 2018 | 0,580 | 0,25 | 2014 | 0,640 | 0,28 | 2010 | 0,752 | 0,33 |
| 2017 | 0,625 | 0,27 | 2013 | 0,589 | 0,26 | 2009 | 0,720 | 0,35 |
| 2016 | 0,664 | 0,26 | 2012 | 0,764 | 0,30 | | | |

Fonte: Autor.

Tabela 4: Classificação de cada item analisado.

| Ano | Total de itens de Física | Quantidade de itens Bons | Quantidade de itens Duvidosos | Quantidade de itens Ruins | Itens Bons | Itens Duvidosos | Itens Ruins |
|------|--------------------------|--------------------------|-------------------------------|---------------------------|----------------------------|--------------------------------|----------------------------------------------------|
| 2019 | 16 | 3 (18,75%) | 3 (18,75%) | 10 (62,5%) | 94, 98, 100 | 92, 109, 119 | 102, 106, 111, 117, 121, 126, 131, 132, 134, 135 |
| 2018 | 15 | 1 (6,67%) | 4 (26,67%) | 10 (66,66%) | 97 | 92, 108, 112, 115 | 103, 104, 120, 122, 123, 125, 128, 129, 131, 134 |
| 2017 | 16 | 4 (25%) | 5 (31,25%) | 7 (43,75%) | 101, 105, 107, 115 | 99, 120, 128, 129, 133 | 93, 103, 108, 110, 112, 127, 131 |
| 2016 | 15 | 2 (13,33%) | 2 (13,33%) | 11 (73,33%) | 57, 74 | 55, 88 | 47, 49, 54, 59, 63, 64, 66, 77, 82, 84, 86 |
| 2015 | 15 | 5 (33,33%) | 7 (46,67%) | 3 (20%) | 63, 68, 82, 85, 86 | 49, 50, 57, 65, 75, 79, 88 | 53, 64, 70, |
| 2014 | 15 | 2 (13,33%) | 8 (53,33%) | 5 (33,33%) | 55, 76 | 50, 57, 66, 68, 72, 84, 87, 90 | 46, 62, 64, 67, 82 |
| 2013 | 16 | 0 (0%) | 3 (18,75%) | 13 (81,25%) | – | 48, 52, 57 | 46, 61, 65, 66, 72, 75, 76, 79, 82, 83, 85, 87, 89 |
| 2012 | 16 | 7 (43,75%) | 4 (25%) | 5 (31,25%) | 47, 50, 54, 61, 64, 71, 72 | 77, 78, 83, 88 | 55, 60, 67, 73, 74 |
| 2011 | 15 | 2 (13,33%) | 7 (46,67%) | 6 (40%) | 66, 86 | 54, 56, 60, 63, 70, 77, 84 | 46, 67, 73, 74, 78, 80 |
| 2010 | 16 | 6 (37,50%) | 7 (43,75%) | 3 (18,75%) | 52, 56, 58, 59, 63, 89 | 47, 48, 54, 68, 78, 81, 84 | 50, 67, 70 |
| 2009 | 16 | 4 (25%) | 5 (31,25%) | 7 (43,75%) | 5, 14, 19, 20 | 24, 30, 31, 32, 37 | 17, 18, 27, 35, 38, 39, 45 |

Fonte: Autor.

mais inadequado de cada edição será apresentado com maior detalhamento.

4.1. Confiabilidade e dificuldade das provas

A Tabela 3 apresenta as informações de consistência interna e dificuldade das 11 provas de Ciências da Natureza analisadas.

Pode-se perceber que, das 11 provas, 2 delas (2013 e 2018) apresentaram um coeficiente de confiabilidade abaixo do valor desejado ($<0,60$). No que tange a dificuldade da prova, notou-se que todas as edições analisadas se encaixam como provas difíceis, dado que os índices de acertos se concentram entre 0,25 e 0,45 [1]. A prova mais difícil de todas foi a de 2018 e a menos difícil foi a de 2009. É possível notar que a prova mais difícil foi também uma das que tiveram menor coeficiente de confiabilidade.

4.2. Qualidade dos itens

Após a análise de cada um dos indicadores pode-se chegar na classificação de itens. Conforme referido na

seção de metodologia, eles foram organizados em 3 categorias: itens bons; itens duvidosos; e itens ruins, conforme os critérios adotados.

Ao todo, as 11 edições analisadas contaram com 171 itens de Física. Desse total, apenas 36 (21,05%) puderam ser considerados itens bons, enquanto 55 (32,16%) foram classificados como duvidosos e outros 80 (46,78%), rotulados como itens ruins. Ou seja, 78,95% das questões de Física do ENEM puderam ser consideradas, segundo as análises realizadas, como não adequadas, por não atenderem a um conjunto de critérios de qualidade preconizados pela literatura psicométrica [5].

No material suplementar disponibilizado é possível conferir as características psicométricas de cada item de Física para cada edição analisada, que acarretaram nas classificações abaixo apresentadas.

Uma análise mais detalhada dos quantitativos, percentuais e de quais itens foram classificados em cada categoria, pode ser observada na Tabela 4.

O ano com mais itens de Física classificados como “bons” foi o de 2012. Mesmo assim, nessa edição, essas questões não chegaram nem a 50% do total. De modo a

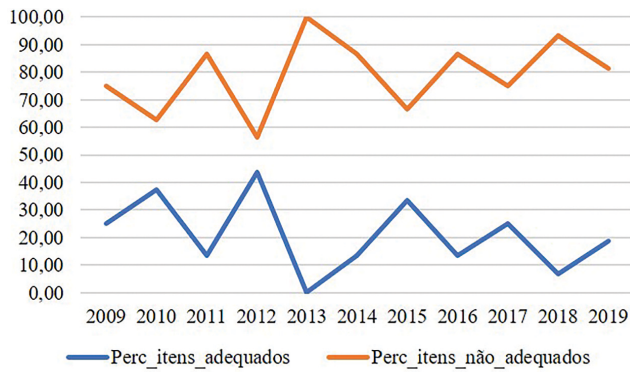


Gráfico 1: Relação entre itens considerados adequados e não adequados. Fonte: Autor.

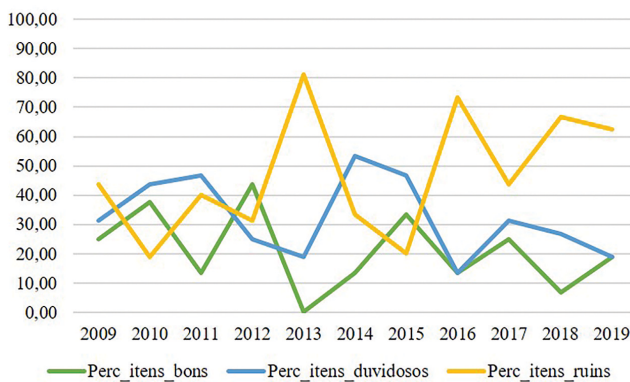


Gráfico 2: Relação de itens bons, duvidosos e ruins. Fonte: Autor.

apresentar à literatura um mapeamento da classificação dos itens de Física, as últimas 3 colunas da Tabela 4 mostram aonde cada item foi classificado. Já o Gráfico 1 apresenta, em função do tempo, uma observação longitudinal, que relaciona os itens considerados adequados (bons) e os não adequados (duvidosos e ruins).

No Gráfico 1 é possível concluir que, em todas as edições analisadas, o percentual de itens considerados não adequados superaram os 50%. As provas de 2013 e 2018 foram aquelas em que mais de 90% dos itens de Física foram classificados como não adequados. Ao todo, em ordem decrescente de itens considerados não adequados, a classificação se deu da seguinte forma: 2013 (100%); 2018 (93,33%); 2014 (86,67%); 2016 (86,67%); 2011 (86,66%); 2019 (81,25%); 2009 (75%); 2017 (75%); 2015 (66,67%); 2010 (62,50%) e 2012 (56,25%).

Ao analisar somente os itens classificados como duvidosos e ruins foi possível perceber que, na maior parte das edições analisadas, itens ruins foram mais numerosos que itens duvidosos. O Gráfico 2 auxilia a compreender essa relação longitudinal:

Destacam-se os itens da prova de 2013, pois, segundo o Gráfico 2, pode-se verificar que 100% deles foram classificados como duvidosos (18,75%) ou ruins (81,25%).

Após essa apreciação panorâmica, emerge o anseio de analisar cada um dos itens classificados como não adequados. No entanto, dada a quantidade de itens (78,95% do total), essa tarefa torna-se inviável para este manuscrito. Não obstante, de modo a ilustrar ao menos algumas questões, a próxima seção apresentará as questões que, segundo a análise realizada, puderam ser consideradas os “piores” itens de cada ano, ou seja, aquelas com os menores índices em cada indicador. Assim, serão apresentados 11 itens, um para cada edição do ENEM analisada.

4.3. Os 11 “piores” itens de Física do ENEM (de 2019 a 2009)

Em um primeiro momento o enunciado da questão será exibido. Posteriormente, uma série de figuras e gráficos apresentarão, nessa ordem: as informações psicométricas do item; o padrão das alternativas assinaladas; a relação entre desempenho e alternativas escolhidas pelos respondentes; e, por fim, as curvas de característica e informação do item. A apresentação será em ordem decrescente de edição do ENEM, iniciando assim, pelo ano de 2019.

4.3.1. ENEM 2019

O item 106 da prova azul do ENEM de 2019 (Figura 1) versa sobre decomposição de vetores aplicados ao estudo das forças. Ele foi elaborado conforme a competência 5 da área das Ciências da Natureza (entender métodos e procedimentos próprios das ciências naturais e aplicá-los em diferentes contextos) e a habilidade 17 (relacionar informações apresentadas em diferentes formas de linguagem e representação usadas nas ciências físicas, químicas ou biológicas, como texto discursivo, gráficos, tabelas, relações matemáticas ou linguagem simbólica).

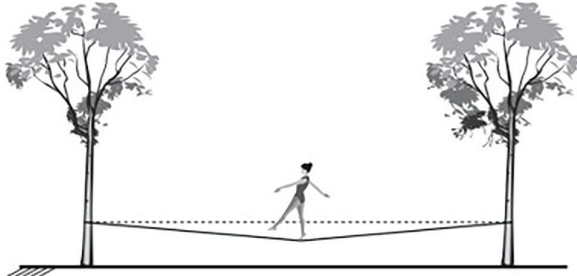
Ao analisar a Tabela 5, percebe-se que esse item foi considerado muito difícil, visto que apenas 12% dos respondentes acertaram-no. Segundo a correlação item total, o valor negativo desse indicador mostra que o item prejudica a prova de Ciências da Natureza no que tange a confiabilidade da medida. Essa condição pode ser corroborada pelo parâmetro “a” da TRI, o qual se ocupa de indicar o nível de discriminação do item.

No Gráfico 3, nota-se que as alternativas A, B e C foram mais assinaladas do que a alternativa correta (letra D).

Por fim, ao analisar a curva característica do item (linha azul), nota-se a qualidade da questão, segundo os 3 parâmetros da TRI. O comportamento observado mostra um item com um padrão fora do comum, em que participantes com maior proficiência possuem maior probabilidade de errar o item, ao invés de acertá-lo. Pode-se afirmar, com isso, que a questão não julga de maneira adequada o conhecimento do participante sobre o tema abordado. Ainda, ao observar a curva de informação do item (linha rosa), nota-se que a

Questão 106

Slackline é um esporte no qual o atleta deve se equilibrar e executar manobras estando sobre uma fita esticada. Para a prática do esporte, as duas extremidades da fita são fixadas de forma que ela fique a alguns centímetros do solo. Quando uma atleta de massa igual a 80 kg está exatamente no meio da fita, essa se desloca verticalmente, formando um ângulo de 10° com a horizontal, como esquematizado na figura. Sabe-se que a aceleração da gravidade é igual a 10 m s⁻², cos(10°) = 0,98 e sen(10°) = 0,17.



Qual é a força que a fita exerce em cada uma das árvores por causa da presença da atleta?

- A 4,0 × 10² N
- B 4,1 × 10² N
- C 8,0 × 10² N
- D 2,4 × 10³ N
- E 4,7 × 10³ N

Figura 1: Item 106 de 2019. Fonte: INEP.

Tabela 5: Indicadores psicométricos para o item 106.

| | | | | | |
|-------------|-------|--------|--------|------------|---------|
| Item | 106 | Ano | 2019 | Respostas | 512.707 |
| Gabarito | | D | | Habilidade | 17 |
| Alt | A | B | C | D | E |
| Freq | 27,95 | 22,25 | 28,45 | 12,30 | 9,05 |
| BISS | 0,054 | 0,016 | -0,065 | -0,013 | 0,009 |
| | TCT | | TRI | | |
| Dificuldade | DP | CORR | a | b | g |
| 0,12 | 0,328 | -0,008 | -1,498 | -3684 | 0.113 |

Fonte: Autor.
 Legendas: DP: desvio-padrão; CORR: correlação item total; BISS: correlação bisserial; a: parâmetro “a” da TRI; b: parâmetro “b” da TRI; c: parâmetro “c” da TRI.

questão teria potencial de mensurar o conhecimento de participantes com proficiência em Ciências da Natureza abaixo da média do teste.

4.3.2. ENEM 2018

O item 123 da prova azul do ENEM de 2018 (Figura 2) aborda o tema “conversão de energia”, sendo elaborado a partir da competência 3 (associar intervenções que resultam em degradação ou conservação ambiental a processos produtivos e sociais e a instrumentos ou ações científico-tecnológicos) e da habilidade 9 (Compreender

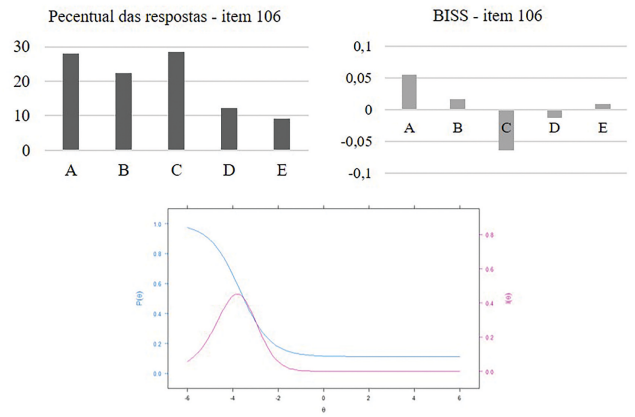
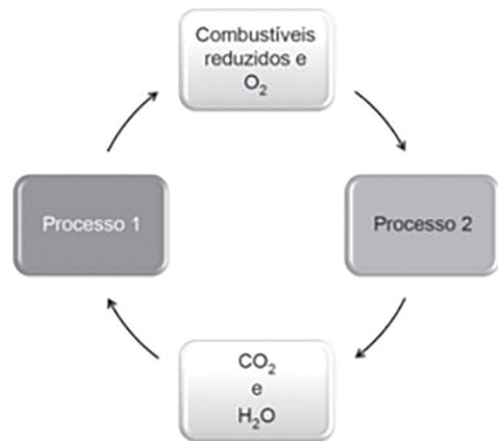


Gráfico 3: Análise gráfica do item 106. Fonte: Autor.

QUESTÃO 123

As células e os organismos precisam realizar trabalho para permanecerem vivos e se reproduzirem. A energia metabólica necessária para a realização desse trabalho é oriunda da oxidação de combustíveis, gerados no ciclo do carbono, por meio de processos capazes de interconverter diferentes formas da energia.



NELSON, D. L.; COX, M. M. *Lehninger: princípios de bioquímica*. São Paulo: Sarvier, 2002 (adaptado).

Nesse ciclo, a formação de combustíveis está vinculada à conversão de energia

- A térmica em cinética.
- B química em térmica.
- C eletroquímica em calor.
- D cinética em eletromagnética.
- E eletromagnética em química.

Figura 2: Item 123 de 2018. Fonte: INEP.

a importância dos ciclos biogeoquímicos ou do fluxo energia para a vida, ou da ação de agentes ou fenômenos que podem causar alterações nesses processos). As informações detalhadas do item podem ser observadas na Tabela 6.

Item considerado muito difícil, apenas 16% dos participantes assinalaram a alternativa correta. Conforme a

Tabela 6: Indicadores psicométricos para o item 123.

| Item | 123 | Ano | 2018 | Respostas | 543.571 |
|-------------|-------|--------|------------|-----------|---------|
| Gabarito | E | | Habilidade | | |
| Alt | A | B | C | D | E |
| Freq | 18,22 | 36,36 | 17,80 | 11,93 | 15,69 |
| BISS | 0,003 | 0,201 | -0,061 | -0,131 | -0,085 |
| TCT | | TRI | | | |
| Dificuldade | DP | CORR | a | b | g |
| 0,16 | 0,363 | -0,062 | -0,784 | -3,431 | 0,083 |

Fonte: Autor.

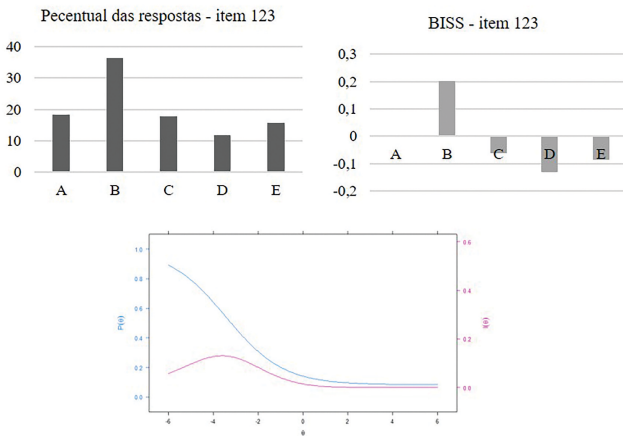


Gráfico 4: Análise gráfica do item 123. Fonte: Autor.

correlação item total, o valor negativo desse indicador mostra que esse item também prejudica a confiabilidade da prova de Ciências da Natureza. O parâmetro “a” da TRI indica que o item não discrimina aqueles que possuem daqueles que não possuem a proficiência aferida.

No Gráfico 4, observa-se que apenas um distrator foi menos assinalado pelos participantes do que a alternativa correta. Todas as outras alternativas incorretas (A, B e C) foram mais escolhidas do que a opção correta (letra E).

Ainda, a correlação biserial mostra que aqueles que tiveram maior desempenho na prova tenderam a assinalar mais a alternativa B do que a correta.

No que tange à curva característica do item, o padrão aqui encontrado segue a mesma tendência daquela observada na questão analisada anteriormente, o que confirma, por meio dos 3 parâmetros da TRI, a inadequação da questão. Sobre a curva de informação do item, essa, no que lhe concerne, contribui menos ainda para o teste do que o item anterior (o que pode ser visto ao comparar a extensão do eixo y da curva) pois, a partir do ponto que demarca a proficiência média de um participante, a curva assume um formato de reta.

4.3.3. ENEM 2017

O item 131 da prova azul do ENEM 2017 (Figura 3) versa sobre Cinemática. Foi construído a partir da competência 6 (apropriar-se de conhecimentos da física

QUESTÃO 131

Um motorista que atende a uma chamada de celular é levado à desatenção, aumentando a possibilidade de acidentes ocorrerem em razão do aumento de seu tempo de reação. Considere dois motoristas, o primeiro atento e o segundo utilizando o celular enquanto dirige. Eles aceleram seus carros inicialmente a 1,00 m/s². Em resposta a uma emergência, freiam com uma desaceleração igual a 5,00 m/s². O motorista atento aciona o freio à velocidade de 14,0 m/s, enquanto o desatento, em situação análoga, leva 1,00 segundo a mais para iniciar a frenagem.

Que distância o motorista desatento percorre a mais do que o motorista atento, até a parada total dos carros?

- A** 2,90 m
- B** 14,0 m
- C** 14,5 m
- D** 15,0 m
- E** 17,4 m

Figura 3: Item 131 de 2017. Fonte: INEP.

Tabela 7: Indicadores psicométricos para o item 131.

| Item | 131 | Ano | 2017 | Respostas | 305.701 |
|-------------|-------|-------|------------|-----------|---------|
| Gabarito | E | | Habilidade | | |
| Alt | A | B | C | D | E |
| Freq | 23,51 | 20,74 | 17,42 | 28,99 | 9,33 |
| BISS | 0,03 | 0,08 | -0,022 | -0,07 | -0,017 |
| TCT | | TRI | | | |
| Dificuldade | DP | CORR | a | b | g |
| 0,09 | 0,29 | 0,004 | -0,072 | -34,259 | 0,016 |

Fonte: Autor.

para, em situações problema, interpretar, avaliar ou planejar intervenções científico tecnológicas) e da habilidade 20 (caracterizar causas ou efeitos dos movimentos de partículas, substâncias, objetos ou corpos celestes). A Tabela 7 aborda suas informações em detalhes.

É considerado um dos itens com maior índice de dificuldade, pois apenas 9% dos respondentes acertaram-no. A correlação item total mostra que, mesmo não tendo um valor negativo, o valor baixo desse indicador presume que o item pouco contribui para o teste. No Gráfico 5, nota-se que todos os distratores foram mais assinalados do que a alternativa correta (letra E).

Ainda sobre o Gráfico 5, é possível notar que os participantes de maior proficiência na prova tenderam a assinalar a alternativa B ao invés da alternativa correta.

É surpreendente observa-se as curvas de característica e informação do item, pois não há o padrão de curva que deve existir para realizar as aferições necessárias. Ou seja, o item, além de não possuir as características psicométricas adequadas, não mensura o conhecimento do participante em nenhuma faixa de proficiência.

4.3.4. ENEM 2016

O item 63 do ano de 2016 (Figura 4) também teve como assunto a Cinemática. Assim como no item 131

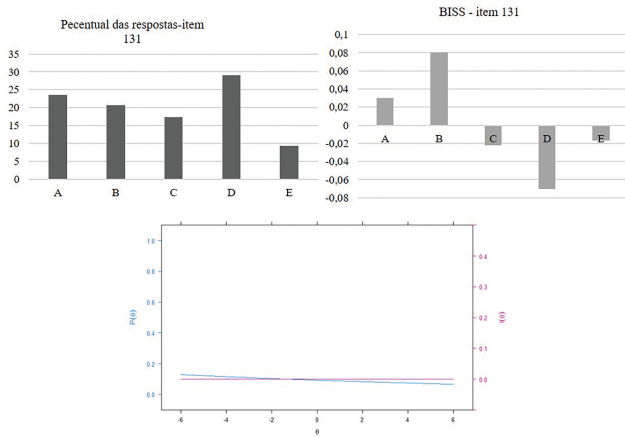


Gráfico 5: Análise gráfica do item 131. Fonte: Autor.

QUESTÃO 63

Dois veículos que trafegam com velocidade constante em uma estrada, na mesma direção e sentido, devem manter entre si uma distância mínima. Isso porque o movimento de um veículo, até que ele pare totalmente, ocorre em duas etapas, a partir do momento em que o motorista detecta um problema que exige uma freada brusca. A primeira etapa é associada à distância que o veículo percorre entre o intervalo de tempo da detecção do problema e o acionamento dos freios. Já a segunda se relaciona com a distância que o automóvel percorre enquanto os freios agem com desaceleração constante.

Considerando a situação descrita, qual esboço gráfico representa a velocidade do automóvel em relação à distância percorrida até parar totalmente?

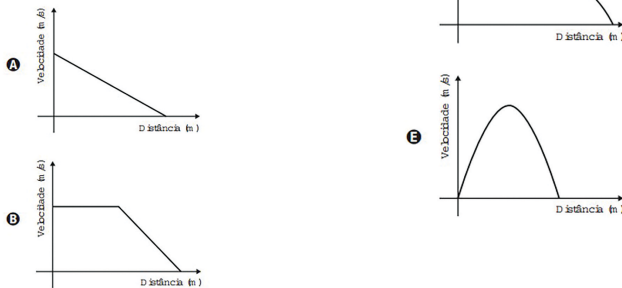


Figura 4: Item 63 de 2016. Fonte: INEP.

de 2017, esse item foi elaborado conforme a supracitada competência 6 e habilidade 20.

A Tabela 8 mostra que o item 63 é considerado também muito difícil e a correlação item total sinaliza uma baixa associação entre o item e a prova de Ciências da Natureza. Nessa questão, o percentual de respondentes que assinalaram a alternativa correta foi maior que aqueles que assinalaram as incorretas, conforme pode ser visto no Gráfico 6.

No entanto, ao analisar a correlação bisserial do item, nota-se que, mesmo sendo a alternativa correta (letra D) a mais assinalada de todas, os participantes com maior proficiência tenderam a assinalar mais a alternativa B.

Sobre as curvas de característica e informação do item, observa-se o mesmo padrão encontrado no item 131 de 2017, o que vai ao encontro dos demais parâmetros que atestam a classificação desse item como “ruim”.

Tabela 8: Indicadores psicométricos para o item 63.

| Item | 63 | Ano | 2016 | Respostas | 349.597 |
|-------------|--------|-------|--------|------------|---------|
| Gabarito | | | D | Habilidade | 20 |
| Alt | A | B | C | D | E |
| Freq | 7,65 | 8,86 | 24,06 | 32,56 | 26,88 |
| BISS | -0,113 | 0,2 | -0,098 | 0,006 | -0,098 |
| | TCT | | TRI | | |
| Dificuldade | DP | CORR | a | b | g |
| 0,235 | 0,424 | 0,009 | 0.035 | 148.393 | 0.230 |

Fonte: Autor.

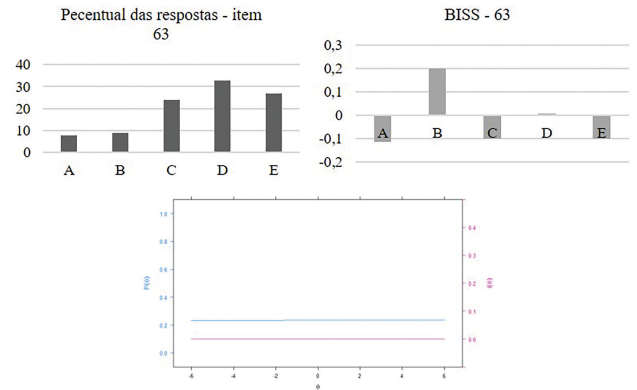


Gráfico 6: Análise gráfica do item 63. Fonte: Autor.

4.3.5. ENEM 2015

O item 70 da prova de 2015 (Figura 5) aborda o assunto “força elástica”. Foi construído segundo a competência 2 (Identificar a presença e aplicar as tecnologias associadas às ciências naturais em diferentes contextos) e habilidade 6 (relacionar informações para compreender manuais de instalação ou utilização de aparelhos, ou sistemas tecnológicos de uso comum).

Conforme pode ser observado na Tabela 9, esse item, embora tenha apresentado uma dificuldade média, ainda assim manteve parâmetros como correlação item total e discriminação abaixo dos valores recomendados. Em suma, esses dois últimos indicadores sinalizam que: o item não contribui para a prova de Ciências da Natureza; e, o item não tem capacidade de diferenciar respondentes com maior e menor proficiência no assunto.

Nessa questão, uma grande parcela dos respondentes também assinalou a alternativa correta (letra B). Mas, mesmo assim, aqueles com maior proficiência tenderam a assinalar a alternativa D, de acordo com a correlação bisserial.

Conforme observado no Gráfico 7, o padrão observado na curva característica do item corresponde a um comportamento anômalo (já visto nos itens analisados nas provas de 2019 e 2018), no qual participantes com maior proficiência tendem a ter menos probabilidade de acertar o item. Já a curva de informação do item segue o mesmo padrão dos itens de 2017 e 2016, nos quais

QUESTÃO 70 ◇◇◇◇◇

Um garoto foi à loja comprar um estilingue e encontrou dois modelos: um com borracha mais “dura” e outro com borracha mais “mole”. O garoto concluiu que o mais adequado seria o que proporcionasse maior alcance horizontal, D , para as mesmas condições de arremesso, quando submetidos à mesma força aplicada. Sabe-se que a constante elástica k_d (do estilingue mais “duro”) é o dobro da constante elástica k_m (do estilingue mais “mole”).

A razão entre os alcances $\frac{D_d}{D_m}$, referentes aos estilingues

com borrachas “dura” e “mole”, respectivamente, é igual a

- A** $\frac{1}{4}$.
- B** $\frac{1}{2}$.
- C** 1.
- D** 2.
- E** 4.

Figura 5: Item 70 de 2015. Fonte: INEP.

Tabela 9: Indicadores psicométricos para o item 70.

| Item | 70 | | Ano 2015 | | Respostas 336.720 |
|-------------|----------|--------|----------|--------|-------------------|
| | Gabarito | | B | | |
| Alt | A | B | C | D | E |
| Freq | 17,77 | 40,66 | 7,09 | 28,44 | 6,05 |
| BISS | -0,116 | -0,053 | -0,045 | 0,216 | -0,055 |
| | TCT | | TRI | | |
| Dificuldade | DP | CORR | a | b | g |
| 0,406 | 0,491 | -0,06 | -0,172 | -3,353 | 0,070 |

Fonte: Autor.

o item não auxilia a mensurar nenhuma informação do conhecimento aferido pela prova.

4.3.6. ENEM 2014

O item 46 (Figura 6) aborda o assunto “oscilações”, sendo construído conforme a competência 1 (compreender as ciências naturais e as tecnologias a elas associadas como construções humanas, percebendo seus papéis nos processos de produção e no desenvolvimento econômico e social da humanidade) e habilidade 1 (reconhecer características ou propriedades de fenômenos ondulatórios ou oscilatórios, relacionando-os a seus usos em diferentes contextos).

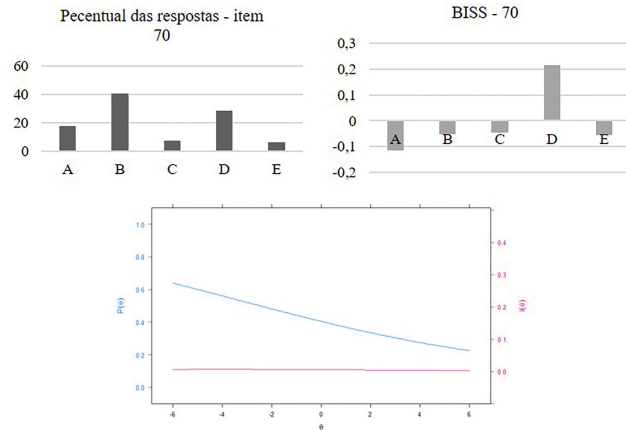


Gráfico 7: Análise gráfica do item 70. Fonte: Autor.

QUESTÃO 46 ◇◇◇◇◇

Christiaan Huygens, em 1656, criou o relógio de pêndulo. Nesse dispositivo, a pontualidade baseia-se na regularidade das pequenas oscilações do pêndulo. Para manter a precisão desse relógio, diversos problemas foram contornados. Por exemplo, a haste passou por ajustes até que, no início do século XX, houve uma inovação, que foi sua fabricação usando uma liga metálica que se comporta regularmente em um largo intervalo de temperaturas.

YODER, J. G. *Unrolling Time: Christiaan Huygens and the mathematization of nature*. Cambridge: Cambridge University Press, 2004 (adaptado).

Desprezando a presença de forças dissipativas e considerando a aceleração da gravidade constante, para que esse tipo de relógio realize corretamente a contagem do tempo, é necessário que o(a)

- A** comprimento da haste seja mantido constante.
- B** massa do corpo suspenso pela haste seja pequena.
- C** material da haste possua alta condutividade térmica.
- D** amplitude da oscilação seja constante a qualquer temperatura.
- E** energia potencial gravitacional do corpo suspenso se mantenha constante.

Figura 6: Item 46 de 2014. Fonte: INEP.

Segundo as suas características psicométricas, observadas na Tabela 10, é um item considerado muito difícil. Sua correlação item total sugere que a questão está pouco associada à prova. Já a sua discriminação, caracterizada pelo parâmetro “a”, mostrou-se adequada. No entanto, conforme pode-se notar no Gráfico 8, a inadequação desse item, além da grande dificuldade e da baixa correlação dele com a prova de Ciências da Natureza, está no fato de que outras duas alternativas (letras D e E) tiveram maior percentual de respondentes do que a alternativa correta (letra A).

Não obstante, ao analisar o coeficiente bisserial do item, observa-se que os respondentes que obtiveram maior desempenho na prova tenderam a assinar a alternativa D, mostrando que há possivelmente

Tabela 10: Indicadores psicométricos para o item 46.

| Item | 46 | Ano | 2014 | Respostas | 1.030.138 |
|-------------|-------|--------|------------|-----------|-----------|
| Gabarito | A | | Habilidade | | |
| Alt | A | B | C | D | E |
| Freq | 14,94 | 7,11 | 14,05 | 37,00 | 26,90 |
| BISS | 0,028 | -0,121 | -0,097 | 0,155 | -0,043 |
| TCT | | TRI | | | |
| Dificuldade | DP | CORR | a | b | g |
| 0,15 | 0,356 | 0,05 | 3.970 | 2.494 | 0.139 |

Fonte: Autor.

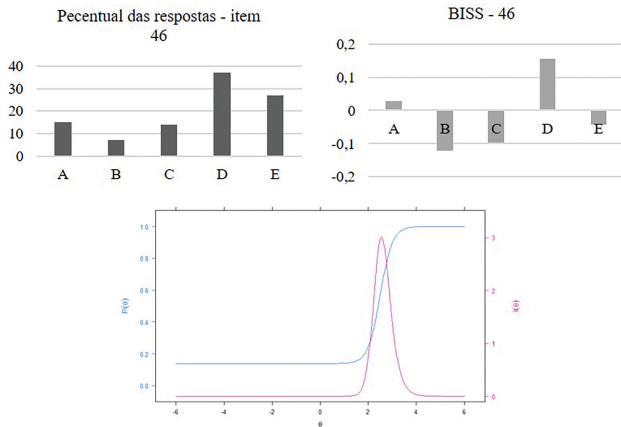


Gráfico 8: Análise gráfica do item 46. Fonte: Autor.

impedimentos, seja conceitual, seja na elaboração do item, que esteja causando este padrão.

Ao analisar as curvas de característica e de informação do item observa-se uma questão com um grau de dificuldade alto (curva azul alinhada à direita), mas com os 3 parâmetros relativamente adequados. A curva de informação mostra que o item auxilia a mensurar a proficiência exigida para uma parcela delimitada no intervalo de 2 a 4 na escala do gráfico, o que representa no mínimo dois desvios padrão acima da média, ou seja, o item poderia ajudar na medida em pessoas com níveis de proficiência muito altos, caso não tivesse as inadequações observadas na correlação bisserial.

4.3.7. ENEM 2013

O item 79 (Figura 7) aborda o tema “corrente elétrica”, criado com base na competência 1 (compreender as ciências naturais e as tecnologias a elas associadas como construções humanas, percebendo seus papéis nos processos de produção e no desenvolvimento econômico e social da humanidade) e na habilidade 3 (confrontar interpretações científicas com interpretações baseadas no senso comum, ao longo do tempo ou em diferentes culturas).

De acordo com a Tabela 11, a questão em foco, de grande dificuldade também, apresenta baixa correlação item total, ou seja, indica não contribuir para a prova na totalidade. O parâmetro “a”, de discriminação, também aponta que o item não tem capacidade de discernir

QUESTÃO 79

Um circuito em série é formado por uma pilha, uma lâmpada incandescente e uma chave interruptora. Ao se ligar a chave, a lâmpada acende quase instantaneamente, irradiando calor e luz. Popularmente, associa-se o fenômeno da irradiação de energia a um desgaste da corrente elétrica, ao atravessar o filamento da lâmpada, e à rapidez com que a lâmpada começa a brilhar. Essa explicação está em desacordo com o modelo clássico de corrente.

De acordo com o modelo mencionado, o fato de a lâmpada acender quase instantaneamente está relacionado à rapidez com que

- A o fluido elétrico se desloca no circuito.
- B as cargas negativas móveis atravessam o circuito.
- C a bateria libera cargas móveis para o filamento da lâmpada.
- D o campo elétrico se estabelece em todos os pontos do circuito.
- E as cargas positivas e negativas se chocam no filamento da lâmpada.

Figura 7: Item 79 de 2013. Fonte: INEP.

Tabela 11: Indicadores psicométricos para o item 79.

| Item | 79 | Ano | 2013 | Respostas | 1.011.596 |
|-------------|-------|-------|------------|-----------|-----------|
| Gabarito | D | | Habilidade | | |
| Alt | A | B | C | D | E |
| Freq | 18,23 | 12,52 | 13,36 | 20,74 | 35,15 |
| BISS | 0,043 | 0,165 | -0,045 | 0,017 | -0,127 |
| TCT | | TRI | | | |
| Dificuldade | DP | CORR | a | b | g |
| 0,21 | 0,405 | 0,014 | 0.089 | 21.476 | 0.088 |

Fonte: Autor.

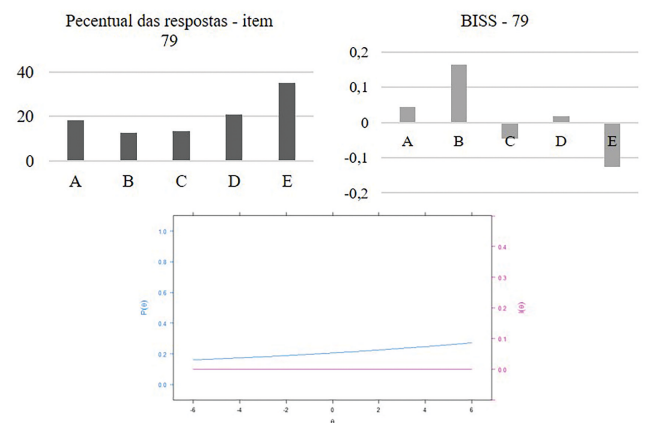


Gráfico 9: Análise gráfica do item 79. Fonte: Autor.

aqueles que têm daqueles que não têm a proficiência. O padrão de respostas, no Gráfico 9, aponta que a alternativa mais assinalada foi o distrator localizado na letra E.

Ao mesmo tempo, o Gráfico 9 mostra, por meio do coeficiente bisserial, que os respondentes que obtiveram

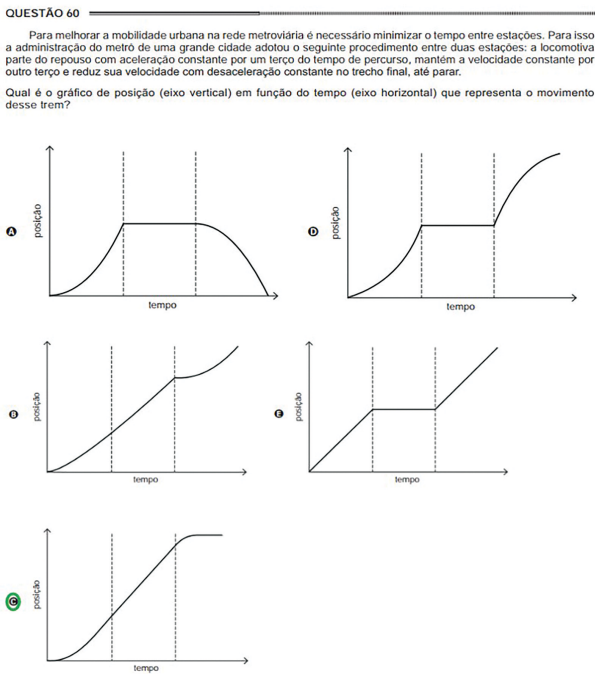


Figura 8: Item 60 de 2012. Fonte: INEP.

maior proficiência na prova de Ciências da Natureza tenderam a escolher a alternativa B como resposta. Tais inadequações são confirmadas ao analisar-se as curvas do item. A sua curva característica confirma o item como sendo de discriminação irrisória (conforme inclinação da linha azul), enquanto a curva de informação (linha rosa) aponta que, em nenhuma faixa de proficiência o item consegue contribuir para a prova no ato de mensurar qualquer informação sobre o constructo.

Tal inadequação nos parâmetros desse item foi objeto de análise em estudos como os de Rodrigues [31] e Rinaldi [32], nos quais os autores ressaltam aspectos pedagógicos da questão, mostrando sua fragilidade teórica e empírica, corroborando os resultados aqui expostos.

4.3.8. ENEM 2012

A questão 60 (Figura 8) aborda gráficos sobre Cinemática e foi elaborada de acordo com a competência 5 e habilidade 17, supracitadas no item 106 da prova de 2019.

Analisando as suas características psicométricas na Tabela 12, pode-se observar um item difícil e com discriminação adequada, conforme o parâmetro “a”, mas que, mesmo assim, tem o indicador correlação item total apontando para uma pouca correlação do item para com a prova de Ciências da Natureza. Ou seja, o item pouco contribui para mensurar o constructo desejado.

Outro indicador que chama atenção é o parâmetro “c” da TRI, pois mostra que a chance de alguém sem a proficiência desejada acertar o item por acaso é de 22,3%. Destaca-se que os valores recomendados para esse parâmetro devem ser menores que 20%.

Tabela 12: Indicadores psicométricos para o item 60.

| | | | | | |
|-------------|----------|--------|------------|-----------|---------|
| Item | 60 | Ano | 2012 | Respostas | 776.140 |
| | Gabarito | C | Habilidade | 17 | |
| Alt | A | B | C | D | E |
| Freq | 44,53 | 9,39 | 24,13 | 6,21 | 15,73 |
| BISS | 0,068 | -0,051 | 0,062 | -0,057 | -0,082 |
| | TCT | | TRI | | |
| Dificuldade | DP | CORR | a | b | g |
| 0,24 | 0,428 | 0,078 | 4.115 | 2.174 | 0.223 |

Fonte: Autor.

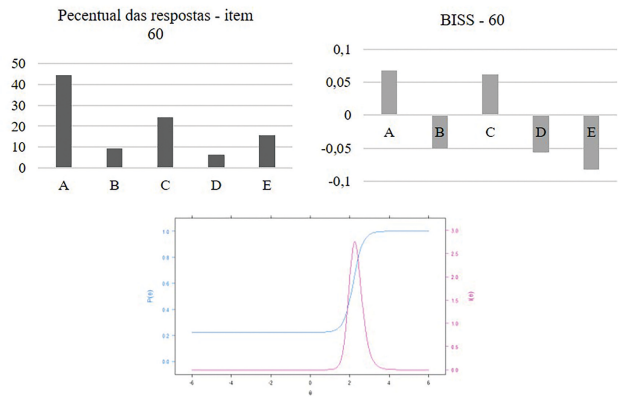


Gráfico 10: Análise gráfica do item 60. Fonte: Autor.

Ao notar o padrão de acertos, observa-se que a alternativa A foi a mais escolhida entre todos os participantes. Sobre essa questão é importante ressaltar que, embora o gabarito aponte a letra C como a alternativa correta, o item que pode ser questionado, uma vez que como os três intervalos de tempo são iguais, as acelerações as etapas final e inicial somente diferem de sinal. Sendo assim, o gráfico deveria mostrar a mesma variação da velocidade no mesmo intervalo de tempo, o que não acontece na opção C. O Gráfico 10 permite verificar esse e outros parâmetros.

As curvas de característica e informação do item assemelham-se ao padrão observado no item 46 da prova de 2014. No item 60, nota-se que ele tem potencial de aferir a informação sobre a proficiência do respondente apenas em níveis acima de 2 desvios padrão da média, enquanto mostra-se um item de grande dificuldade (curva mais à direita) e com elevado potencial de acerto ao acaso (acima de 20%), conforme discutido nos parágrafos anteriores.

Os aspectos pedagógicos do item 60 foram analisados no trabalho de Silveira [33], no qual demonstra que, tendo em vista as alternativas apresentadas, essa questão deveria ser anulada, uma vez que não apresenta uma alternativa correta.

4.3.9. ENEM 2011

O item 74 da prova de 2011 (Figura 9) versa sobre o tema “ondas” e foi construído segundo a competência 1 e habilidade 1, já referida no item 46 de 2014.

Questão 70

Observe a tabela seguinte. Ela traz especificações técnicas constantes no manual de instruções fornecido pelo fabricante de uma torneira elétrica.

Especificações Técnicas

| Modelo | Torneira | | | |
|----------------------------|--------------------|--------------------|-------------------|-------------------|
| | 127 | | 220 | |
| Tensão Nominal (Volts~) | (Frio) | | Desligado | |
| Potência Nominal (Watts) | 2 800 | 3 200 | 2 800 | 3 200 |
| | (Morno) | | | |
| | 4 500 | 5 500 | 4 500 | 5 500 |
| | (Quente) | | | |
| Corrente Nominal (Ampères) | 35,4 | 43,3 | 20,4 | 25,0 |
| Fiação Mínima (Até 30 m) | 5 mm ² | 10 mm ² | 4 mm ² | 4 mm ² |
| Fiação Mínima (Acima 30 m) | 10 mm ² | 16 mm ² | 6 mm ² | 6 mm ² |
| Disjuntor (Ampères) | 40 | 50 | 25 | 30 |

Disponível em: http://www.cardal.com.br/manualprod/Manuais/Torneira%20Suprema/-Manual_Torneira_Suprema_roo.pdf

Considerando que o modelo de maior potência da versão 220 V da torneira suprema foi inadvertidamente conectada a uma rede com tensão nominal de 127 V, e que o aparelho está configurado para trabalhar em sua máxima potência. Qual o valor aproximado da potência ao ligar a torneira?

- A 1.830 W
- B 2.800 W
- C 3.200 W
- D 4.030 W
- E 5.500 W

Figura 10: Item 70 de 2010. Fonte: INEP.

Tabela 14: Indicadores psicométricos para o item 70.

| Item | 70 | Ano | 2010 | Respostas | 727.174 |
|-------------|--------|--------|--------|------------|---------|
| Gabarito | | A | | Habilidade | 5 |
| Alt | A | B | C | D | E |
| Freq | 13,40 | 25,68 | 25,42 | 13,09 | 22,40 |
| BISS | -0,033 | -0,059 | 0,026 | -0,044 | 0,108 |
| | TCT | | TRI | | |
| Dificuldade | DP | CORR | a | b | g |
| 0,13 | 0,34 | -0,022 | -0,822 | -4,954 | 0,113 |

Fonte: Autor.

4.3.11. ENEM 2009

O item 27 da prova de 2009 (Figura 11) versa sobre conceitos pertencentes às Leis de Newton e foi elaborado segundo a competência 6 e habilidade 20, já exibidas no item 131 da prova de 2017.

Na Tabela 15 pode-se notar a alta dificuldade da questão, bem como, os valores inapropriados de correlação item total e discriminação. Assim como na maior parte dos itens aqui apresentados, esse também não contribui para a prova de Ciências da Natureza e não tem capacidade de discriminar pessoas com e sem proficiência.

Observando o padrão de respostas, percebe-se que 3 distratores (A, B e E) foram mais assinalados que a

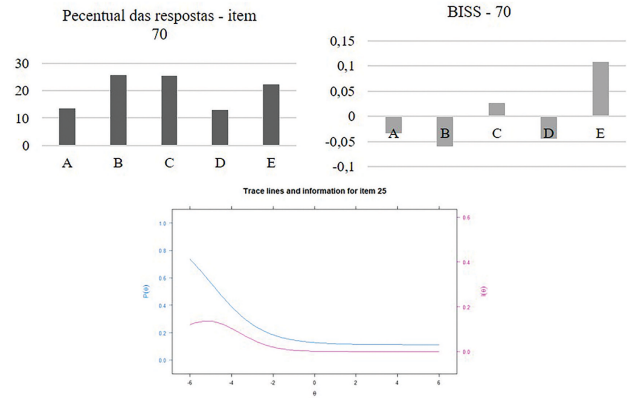
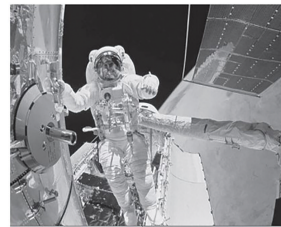


Gráfico 12: Análise gráfica do item 70. Fonte: Autor.

Questão 27

O ônibus espacial *Atlantis* foi lançado ao espaço com cinco astronautas a bordo e uma câmera nova, que iria substituir uma outra danificada por um curto-circuito no telescópio *Hubble*. Depois de entrarem em órbita a 560 km de altura, os astronautas se aproximaram do *Hubble*. Dois astronautas saíram da *Atlantis* e se dirigiram ao telescópio. Ao abrir a porta de acesso, um deles exclamou: "Esse telescópio tem a massa grande, mas o peso é pequeno."



Considerando o texto e as leis de Kepler, pode-se afirmar que a frase dita pelo astronauta

- A se justifica porque o tamanho do telescópio determina a sua massa, enquanto seu pequeno peso decorre da falta de ação da aceleração da gravidade.
- B se justifica ao verificar que a inércia do telescópio é grande comparada à dele próprio, e que o peso do telescópio é pequeno porque a atração gravitacional criada por sua massa era pequena.
- C não se justifica, porque a avaliação da massa e do peso de objetos em órbita tem por base as leis de Kepler, que não se aplicam a satélites artificiais.
- D não se justifica, porque a força-peso é a força exercida pela gravidade terrestre, neste caso, sobre o telescópio e é a responsável por manter o próprio telescópio em órbita.
- E não se justifica, pois a ação da força-peso implica a ação de uma força de reação contrária, que não existe naquele ambiente. A massa do telescópio poderia ser avaliada simplesmente pelo seu volume.

Figura 11: Item 27 de 2009. Fonte: INEP.

Tabela 15: Indicadores psicométricos para o item 27.

| Item | 27 | Ano | 2009 | Respostas | 571.247 |
|-------------|-------|--------|--------|------------|---------|
| Gabarito | | D | | Habilidade | 20 |
| Alt | A | B | C | D | E |
| Freq | 41,22 | 20,05 | 6,80 | 14,03 | 17,90 |
| BISS | 0,168 | -0,045 | -0,125 | -0,029 | -0,057 |
| | TCT | | TRI | | |
| Dificuldade | DP | CORR | a | b | g |
| 0,14 | 0,347 | -0,021 | -0,499 | -7,178 | 0,113 |

Fonte: Autor.

alternativa correta (letra D). Essa distribuição pode ser melhor percebida ao observar o Gráfico 13.

Ao mesmo tempo, o Gráfico 13 mostra que a alternativa A foi a que os participantes com maior proficiência tenderam a escolher, sinalizando um potencial problema de validade ou elaboração do item, o que pode ser corroborado pelas curvas de característica e informação da questão, com padrão semelhante às curvas inadequadas vistas em itens anteriores.

Aspectos pedagógicos do item 27 são debatidos Barroso, Rubini e Silva [17], que corroboram que a CCI não se comporta de acordo com a previsão do modelo.

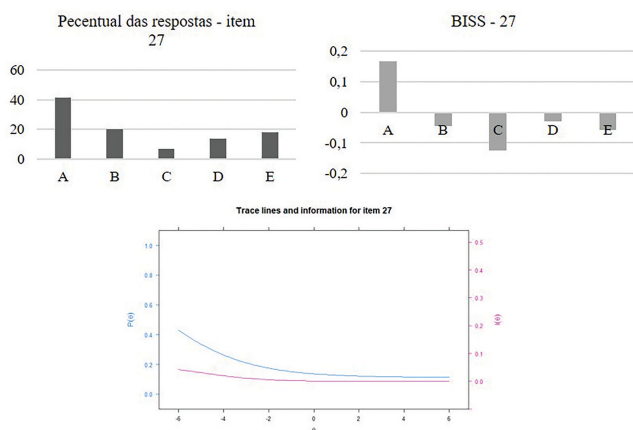


Gráfico 13: Análise gráfica do item 27. Fonte: Autor.

De acordo com os autores, uma discriminação negativa indica que participantes com maior aptidão possuem uma chance menor de acertar tal item.

5. Um Breve Olhar Sobre a Qualidade do ENEM

É importante salientar que apenas alguns indicadores psicométricos foram considerados para as análises aqui expostas, como discriminação, confiabilidade e os parâmetros dos itens. No entanto, aspectos referentes à validade dos mesmos não foram objeto de estudo, sendo essa lacuna, outra sugestão para estudos futuros.

Dessa maneira, pode ser que itens considerados duvidosos, por exemplo, possam ter tido importância para a prova na totalidade, visto que a distribuição dos assuntos abordados deve ser considerada para a prova, em geral.

É importante destacar que muitos dos itens classificados como os “piores” nessa pesquisa já foram analisados em estudos anteriores, conforme foram sendo mencionados em cada análise. Tais estudos abordaram a problemática através de outros critérios com aspectos ligados ao Ensino de Física, ao que é avaliado em cada item, ao conhecimento prévio dos participantes, etc. Ou seja, salienta-se que, após a apresentação dos 11 piores itens, empregar aspectos unicamente psicométricos para aferir a qualidade das questões tem suas limitações, sendo as considerações feitas a partir dos resultados obtidos limitadas ao recorte teórico e metodológico escolhido para a presente pesquisa. Daí o constante incentivo para que não somente os 11 “piores” itens, mas todos aqueles considerados “duvidosos” ou “ruins” sejam objetos de uma análise qualitativa em estudos futuros.

Por fim, reflete-se que em 2009 – quando o ENEM passou por uma série de mudanças teórico-metodológicas, dentre outras coisas, as modificações possibilitaram realizar a validação dos itens e a estimativa dos desempenhos por meio da TRI. A inovação produziu um melhoramento com ganhos significativos para a qualidade da prova e dos resultados de desempenho.

Ao refletir sobre os resultados obtidos nessa investigação e considerando o rigor que a TRI e os demais procedimentos psicométricos já utilizados pelo INEP podem proporcionar para a qualidade de uma prova, questiona-se: quais aspectos do ENEM necessitam de maior atenção para entender como melhorar a qualidade das questões de Física? Será que as melhorias devem acontecer na fase de elaboração dos itens? Na revisão pedagógica? Na validação de conteúdo? No processo de calibração? Enfim, as possibilidades de reflexões e investigações são diversas, sendo sugestões potencialmente profícuas para pesquisadores da área do Ensino de Física voltarem suas atenções.

Destaca-se o esforço de autarquias como o INEP, as quais realizam um trabalho imprescindível ao disponibilizar os Microdados e diversos outros documentos para a população. Essa relação de transparência torna possível a realização de inúmeras pesquisas na área da Educação, com vistas a analisar, a questionar e a refletir sobre o papel, a importância, a intencionalidade e os desafios das avaliações em larga escala existentes hoje no Brasil. Os resultados advindos de tais investigações podem atuar como subsídios teóricos para fundamentar políticas públicas que auxiliem na garantia de que exames da magnitude do ENEM possam cumprir, com qualidade e equidade, o seu objetivo fim.

Enfatiza-se a importância que o ENEM tem para cada cidadão que deseja ingressar em um curso do Ensino Superior. Investigações com o foco em avaliar a qualidade da prova não têm como função desacreditar o Exame e os seus referenciais metodológicos, mas sim, buscar reconhecer a sua importância e magnitude e promover práticas que visem estimular o seu melhoramento e a reflexão sobre como atuar frente aos desafios de se elaborar uma avaliação em larga escala com validade e confiabilidade. Afinal, conforme abordado nesse texto, um desempenho satisfatório nessa avaliação, pode trazer recompensas extremamente importantes para a população, seja o ingresso no Ensino Superior em IES públicas, seja o financiamento e bolsas de estudo em IES privadas, entre outros. Por isso, deseja-se que o exame possa cumprir o seu papel com qualidade, sendo que, para isso, mostra-se fundamental estudar tais aspectos.

6. Considerações

Apresentaram-se os resultados de uma investigação que analisou as características psicométricas das questões de Física do ENEM.

Conforme a análise, 78,95% dos itens de Física das edições de 2009 até 2019 foram classificados como não adequados, por apresentarem diferentes distorções nos indicadores de qualidade elegidos para as análises.

Essa conclusão vai ao encontro de trabalhos da área que, por meio de outras análises que não a psicométrica, chegaram a conclusões semelhantes sobre a qualidade dos itens, com destaque para a pesquisa de Marcom

e Kleinke [13], na qual verificaram que as alternativas incorretas mais assinaladas demonstram dificuldades em comum em processos como o uso de unidades de medida, raciocínio indutivo, resolução de problemas, análise de imagens, etc.

Salienta-se que uma análise psicométrica nos moldes realizados neste artigo, mesmo com toda robustez do método aplicado, analisa apenas um recorte dos dados com base nas informações oficiais disponibilizadas. Isso quer dizer que, devido às limitações metodológicas, epistêmicas e práticas, uma avaliação mais integradora deve considerar também outros aspectos, muitos deles qualitativos, como em uma análise pedagógica das questões, por exemplo. Por isso, esse estudo não tem como meta encerrar a discussão acerca da qualidade dos itens do ENEM, e sim, incentivar que mais estudos com esse escopo possam ser realizados nas mais diferentes áreas, com os mais diversos instrumentos, métodos e referenciais teóricos. Por fim, como perspectivas para estudos futuros, recomenda-se analisar a qualidade psicométrica dos itens de Física das edições de 2020 e 2021.

Material suplementar

O seguinte material suplementar está disponível online: Tabela S1 – Características psicométricas dos itens de Física do ENEM de 2009 a 2019.

Referências

- [1] R. Travitzki, Est. Aval. Educ. **28**, 256 (2017).
- [2] T.E.A. Soares, D.J.M. Soares e W. Santos, Jornal int. Est. Educ. Mat. **14**, 119 (2021).
- [3] C.S. Hutz, D.R. Bandeira e C.M. Trentini, *Psicometria* (Artmed, Porto Alegre, 2015), p. 188.
- [4] L. Pasquali, *Psicometria: teoria dos testes na psicologia e na educação* (Vozes, Petrópolis, 2017), p. 392.
- [5] American Educational Research Association, American Psychological Association e National Council on Measurement in Education, *Standards for educational and psychological testing* (AERA, Washington, 2014), p. 230.
- [6] R. Travitzki. *ENEM: limites e possibilidades do Exame Nacional do Ensino Médio enquanto indicador de qualidade escolar*. Tese de Doutorado, Universidade de São Paulo, São Paulo (2013).
- [7] J.A.F. Pontes Junior, A.G. Silva, E.D. Tavare, L.A. Sousa, F.A.C. Bastos, F.N.I. Cruz e L.A. Silva, *Motricidade* **12**, 12 (2016).
- [8] C.M.A. Gomes, H.F. Golino e A.J.S. Peres, *Psico* **51**, e31145 (2020).
- [9] J. Piton-Gonçalves e A.M. Almeida, *Rev. Eletr. Mat.* **4**, 38 (2018).
- [10] W.P. Gonçalves Junior e M.F. Barroso, *Rev. Bras. Ensino Física* **36**, 1402 (2014).
- [11] J.M.F. Bassalo, *Cad. Bras. Ensino Física*, **28**, 325 (2011).
- [12] J.C. Lopes, *As questões de física do Enem 2011*. Dissertação de Mestrado, Universidade Federal do Rio de Janeiro, Rio de Janeiro (2015).
- [13] G.S. Marcom e M.U. Kleinke, *Cad. Bras. Ensino Física* **33**, 72 (2016).
- [14] G.S. Marcom e M.U. Kleinke, em: *XXII Simpósio Nacional de Ensino de Física (São Carlos, 2017)*.
- [15] P.R.O. Carlos, *Uma análise do desempenho dos estudantes no exame nacional do ensino médio e as contribuições para o ensino-aprendizagem de física*. Dissertação de Mestrado, Universidade Federal de Juiz de Fora, Juiz de fora (2016).
- [16] A.S. Lemos e J.S. Hernandez, *Mult. Sci. Journal* **1**, 21 (2018).
- [17] M.F. Barroso, G. Rubini e T. Silva, *Rev. Bras. Ensino Física* **40**, e4402 (2018).
- [18] L.P. Duarte, W.P. Gonçalves Junior e M.F. Barroso, em: *XX Simpósio Nacional de Ensino de Física* (São Paulo, 2013).
- [19] M.P. Rodrigues, *A taxonomia de Bloom aplicada à questões de física*. Dissertação de Mestrado, Universidade Federal de Viçosa, Viçosa (2018).
- [20] M.M. Nascimento, C.C. Cavalcanti e F. Ostermann, *Rev. Bras. Ensino Física* **40**, e3402 (2018).
- [21] F.L. Silveira, M.C.B. Barbosa e R. Silva, *Rev. Bras. Ensino Física* **37**, 1101 (2015).
- [22] F.L. Silveira, J.F. Stilck e M.C.B. Barbosa, *Cad. Bras. Ensino Física* **31**, 473 (2014).
- [23] G.M. Rubini, *O que o Enem revela sobre a Aprendizagem em Física na Educação Básica*. Tese de Doutorado, Universidade Federal do Rio de Janeiro, Rio de Janeiro (2019).
- [24] A.P.L. Vilarinho, *Uma proposta de análise de desempenho dos estudantes e de valorização da primeira fase da OBMEP*. Dissertação de Mestrado, Universidade de Brasília, Brasília (2015).
- [25] F.A. Santos, *Rev. Elet. Ciên. Adm.* **17**, 192 (2018).
- [26] J.V.L. Robaina, R.S. Fenner, L.A.M. Martins, R.A. Barbosa e J. R. Soares (org.), *Fundamentos teóricos e metodológicos da pesquisa em educação em ciências* (Editora Bagai, Curitiba, 2021).
- [27] A.C. Gil, *Métodos e técnicas de pesquisa social* (Atlas, São Paulo, 2008), p. 220.
- [28] R Core Team, *R: a language and environment for statistical computing* (R Foundation for Statistical Computing, Vienna, 2018), p. 10.
- [29] R.P. Chalmers, *Journal Stat. Soft.* **48**, 1 (2012).
- [30] W. Revelle, *Psych: Procedures for Personality and Psychological Research. R package version 1.4.3. 2014*, disponível em: <https://cran.r-project.org/web/packages/psych/psych.pdf>
- [31] R.C. Rodrigues, *Proposta de sequência didática sobre circuitos elétricos para o ensino fundamental*. Dissertação de Mestrado, Universidade Federal do Rio de Janeiro, Rio de Janeiro (2020).
- [32] B.B. Rinaldi, *Estudo das questões de física do ENEM 2013*. Dissertação de Mestrado, Universidade Federal do Rio de Janeiro, Rio de Janeiro (2017).
- [33] F.L. Silveira, *Questões de física na prova de ciências da natureza no ENEM 2012*, disponível em: https://if.ufrgs.br/~lang/Textos/Quest_Fisica.pdf, acessado em 23/06/2022.