

# Análise da dificuldade e da discriminação de itens de Matemática do ENEM

## Analysis of the difficulty and discrimination of ENEM Mathematics items

Jean Piton-Gonçalves  
Universidade Federal de São Carlos (UFSCar), Departamento de Matemática  
São Carlos, SP, Brasil  
jpiton@ufscar.br

André Marcos Almeida  
Universidade Federal de São Carlos (UFSCar), Departamento de Matemática  
São Carlos, SP, Brasil  
andre\_tony147@hotmail.com

---

### Informações do Artigo



#### Histórico do Artigo

Submissão: 08 de junho de 2018.  
Aceite: 19 de setembro de 2018.

#### Palavras-chave

Testes Educacionais  
Teoria Clássica de Testes  
Exame Nacional do Ensino Médio

### Resumo

No contexto da Avaliação Educacional, um teste objetivo pode seguir a Teoria Clássica de Testes em que é possível verificar a qualidade de um item de um teste por meio (i) do Índice de Dificuldade e (ii) do Índice de Discriminação. O objetivo é fomentar discussões para a análise de itens, considerando aspectos quantitativos e qualitativos, partindo da amostra de 6.826 participantes do ENEM na edição de 2012. Os itens foram sistematicamente resolvidos e classificados de acordo com o PCN+ e os índices de dificuldade e de discriminação. Resultados mostram que 57,78% dos itens podem ser considerados, pelo menos, difíceis e 75,56% dos itens estão entre aceitável ou satisfatório. Destaca-se, ainda, que 6,67% poderiam ter sido excluídos por serem considerados ineficientes.

### Abstract

In the context of Educational Assessment, an objective test follows the Classical Test Theory where it is possible to check the quality of an item in a test by (i) the Difficulty Index and (ii) the Discrimination Index. The aim is to foster discussions for the analysis of items, considering quantitative and qualitative aspects, by a sample of 6,826 ENEM participants in the edition of 2012. The items were systematically solved and classified according to the PCN+ and difficulty and discrimination index. Results show that 57.78% of the items can be considered hard and 75.56% of the items are between acceptable or satisfactory. It is also noted that 6.67% could have been excluded because they were considered inefficient.

#### Keywords

Educational Tests  
Classical Test Theory  
National High School Exam

## 1 Introdução

A preocupação com a qualidade do ensino e da aprendizagem em ambientes educacionais vem crescendo ao longo das últimas décadas e, especificamente, a Avaliação Educacional tem um papel fundamental nos processos de ensino e aprendizagem, “de tal modo que, ao falar de avaliação, as pessoas a associam quase automaticamente à escola, ao ensino, de qualquer nível, de qualquer

área” (RIBEIRO, 2009, p. 58). Ludke (1987, p. 44) ressalta que “embora a avaliação acompanhe praticamente todas as atividades humanas, explicitamente ou não, é na educação e sobretudo, na escola elementar, que ela se exerce de maneira obrigatória e universal”.

Enquanto um dos instrumentos de avaliação, os Testes Objetivos buscam cada vez mais avaliações justas, precisas e rápidas, de acordo com métricas e critérios educacionais pré-estabelecidos e consolidados sistematicamente, em contrapartida aos testes que demandam resposta dissertativa (PITON-GONÇALVES, 2012). A elaboração e análise de um teste pode seguir a Teoria Clássica de Testes (TCT) (ERTHAL, 2001; PASQUALI, 2009) e/ou a Teoria de Resposta ao Item (TRI) (LORD, 1980; ANDRADE; TAVARES; VALLE, 2000). O maior exemplo de aplicação metodológica da TRI ocorre no Exame Nacional do Ensino Médio (ENEM), que tem como principal proposta avaliar o desempenho dos candidatos ao término da escolaridade básica de todo o país. Pasquali & Primi (2003) afirmam que é possível verificar a qualidade de um item de um teste por meio (i) do Índice de Dificuldade e (ii) do Índice de Discriminação.

No contexto das avaliações externas em larga escala, a nossa revisão de literatura mostrou que existem poucos estudos que tratem da TCT no contexto do ENEM. Nesse sentido, destacamos os trabalhos correlatos de Santos & Cortelazzo (2013) que, essencialmente, analisaram a frequência e a forma em que os itens de Biologia são apresentados nas provas de Ciências da Natureza e suas Tecnologias. Silva, Santiago & Santos (2013) estudaram os itens de matemática nas provas de Matemática e suas Tecnologias das edições de 2009, 2010 e 2011, com objetivo de analisar os números racionais nos seus diferentes significados e suas representações semióticas. Travitzki (2017) realizou uma meta-avaliação do ENEM de 2009 e 2011, aplicando técnicas psicométricas, comparando determinados indicadores relacionados à prova. Piton-Gonçalves, Souza & Lamonato (2017) analisam os itens que envolvem os conteúdos de logaritmos, trigonometria, matrizes e determinantes dos anos de 2009 a 2014 do ENEM.

Embora a TRI tenha muitas vantagens metodológicas sobre a TCT (PASQUALI, 2009), a TRI utiliza-se de recursos da inferência estatística e da modelagem matemática que, do ponto de vista prático, são de difícil interpretação por não especialistas. Além disso, os processos de inferência estatística são computacionalmente custosos e trabalhosos, principalmente quando trata-se de um grande volume de dados (PANALOZA; VARGAS-PEREZ, 2017). Uma vez que as respostas dos participantes sejam conhecidas, a TCT pode ser aplicada, por exemplo, via planilhas de cálculo, como ocorreu em parte desta pesquisa.

Esta pesquisa instanciou o ENEM na edição de 2012, realizando a mineração, a classificação e a resolução de 45 itens de matemática com o objetivo de aplicarmos a TCT e relacioná-la de forma qualitativa com a resolução sistemática daquela edição.

## 2 Introdução à Teoria Clássica de Testes

Quando trata-se de avaliação educacional em larga escala, a expectativa é de que as medidas obtidas sejam confiáveis e precisas. Nesse contexto, podemos nos questionar em relação à qualidade de um teste, e de que forma é possível reduzir os erros envolvidos no processo de medição. Silveira (2015) afirma que é importante para a TCT o significado do que representa a soma dos itens respondidos corretamente, assim como a interpretação do resultado final do teste.

Como todo modelo teórico, sua conceituação deve estar fundamentada a partir de algumas hipóteses, que lhe darão sustentabilidade teórica e que possibilitarão descrever, da melhor forma possível, o comportamento de uma realidade (LEITE, 2009). A primeira hipótese parte do pressuposto de que o escore de um teste é, por natureza, falho e contém alguma parcela de erro. Dessa forma, o escore observado ( $T$ ) pode ser decomposto em dois componentes: o escore verdadeiro do participante ( $V$ ) e a parcela do erro associado ( $E$ ), que se relacionam da seguinte forma:  $T = V + E$ . Na segunda hipótese, assume-se que os erros de medição dos participantes em um teste não estão relacionados aos erros de medição em outro. Já a terceira hipótese, admite que não existe correlação entre os escores verdadeiros e os erros (de medição) em um mesmo teste.

Na análise dos itens que compõem um teste é preciso verificar a adequação de cada item à população ou amostra testada, que pode ser realizada por meio do Índice de Dificuldade (ID) e o Índice de Discriminação (IDS).

### 2.1 Índice de Dificuldade - ID

O nível de dificuldade (ou de facilidade) de um item em um teste pode ser medido por meio do Índice de Dificuldade (ID). Itens com baixa quantidade de acertos são considerados mais difíceis em relação a itens com um maior número de acertos. O ID do  $i$ -ésimo item é definido como:

$$ID_i = \frac{A}{n}, \quad (1)$$

em que  $A$  é o número de participantes que responderam corretamente ao item e  $n$  é o número de participantes que realizaram o teste e  $0 \leq ID_i \leq 1$ . Do ponto de vista dicotômico, se todos os participantes respondem incorretamente a um item, tem-se que  $ID_i = 0$ , representando assim um item com um índice máximo de dificuldade. Caso contrário, obtemos  $ID_i = 1$ , indicando um item com um índice máximo de facilidade. Para Erthal (2001), o objetivo deste índice é permitir a construção de testes heterogêneos, em que os itens possuam um aumento progressivo no grau de dificuldade, abrangendo os diferentes níveis ao longo do teste.

Seguindo a Equação (1), Cerdá (1978) propõe cinco classificações a partir de intervalos numéricos (Tabela 1) e seus percentuais (distribuição normal recomendada daquele ID em um teste), com a finalidade de se construir um teste que possui diferentes graus de dificuldade (heterogêneo).

Tabela 1 – Classificação do Índice de Dificuldade.

Intervalo do ID	Percentual	Classificação do item
[0, 75; 0, 95]	10%	Muito Fácil
[0, 55; 0, 75[	20%	Fácil
[0, 45; 0, 55[	40%	Normal (ou Moderado)
[0, 25; 0, 45[	20%	Difícil
[0, 05; 0, 25[	10%	Muito Difícil

Fonte: Cerdá (1978).

Outra forma para se obter o ID é por meio do Método Abreviado dos 27%. Neste considera-se apenas os participantes que encontram-se nos extremos da distribuição. Ou seja, o grupo inferior dos 27% e o grupo superior dos 27%, desconsiderando-se os 46% centrais. Detalhes são encontrados em Erthal (2001) e Pasquali (2017).

## 2.2 Índice de Discriminação - IDS

O Índice de Discriminação (IDS) indica se um item é capaz de diferenciar os participantes que obtiveram alto ou baixo escore no teste. Por exemplo, se um item foi respondido corretamente por todos os participantes com baixo escore, espera-se que um participante com alto escore também responda corretamente. Segundo Erthal (2001, p. 82), “para um teste discriminar bem o grupo de participantes testados na variável medida, é necessário que apresente um bom grau de consistência interna entre os itens e a totalidade do teste”. Ebel (1954) propõe uma classificação para a discriminação dos itens, mostrada na Tabela 2. Itens aceitáveis ou satisfatórios devem permanecer

no teste. Um item que necessita de revisão pode ser reincluído após ajustes. Um item ineficiente, recomenda-se sua eliminação do teste ou deve passar por uma profunda revisão.

Tabela 2 – Classificação do Índice de Discriminação.

Intervalo do IDS	Classificação do item
]0, 0; 0, 2[	Ineficiente
[0, 2; 0, 3[	Necessita revisão
[0, 3; 0, 4[	Aceitável
Maior ou igual a 0, 4	Satisfatório

Fonte: Ebel (1954).

Em termos de cálculo do IDS e tomando a correlação item total, o IDS pode ser determinado de duas formas (PASQUALI, 2017, p. 135):

**Correlação Bisserial.** É indicada quando as variáveis correlacionadas são contínuas, porém quando uma delas foi dicotomizada (resposta correta ou incorreta):

$$r_{bis} = \frac{\bar{x}_p - \bar{x}_t}{s_t} \cdot \frac{pq}{y} \quad (2)$$

em que  $\bar{x}_p$  é a média no teste dos examinados que acertaram o item,  $\bar{x}_t$  é a média total do teste,  $s_t$  é o desvio padrão populacional do teste,  $p$  é a proporção de examinados que acertaram o teste e  $y$  é a ordenada na curva normal no ponto de divisão dos segmentos que contêm as proporções  $p$  e  $q = 1 - p$ . Observa-se que a Equação (2) poderá resultar em valores maiores do que um, quando a distribuição não for normal e, neste caso, recomenda-se o uso da Correlação Bisserial por Pontos, descrita a seguir.

**Correlação Bisserial por Pontos.** É indicada quando a distribuição dos dados é dicotômica e pode ser calculada por:

$$r_{bisp} = \frac{\bar{x}_p - \bar{x}_t}{s_t} \cdot \sqrt{\frac{p}{q}} \quad (3)$$

Além das correlações supracitadas, pode-se utilizar a Correlação ( $\Phi$ ) e a Correlação Tetracórica, que são detalhadas em Pasquali (2017).

### 2.3 ENEM

Em 1998, o Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira (INEP), uma autarquia federal vinculada ao Ministério da Educação (MEC), desenvolveu o Exame Nacional do Ensino Médio (ENEM), com a principal proposta de avaliar o desempenho dos estudantes ao término da escolaridade básica de todo o país. Trata-se de uma importante avaliação que permite o acesso a cursos profissionalizantes, pós-médios e à Educação Superior (BRASIL, 2018).

Desde 2009, a parte objetiva (múltipla escolha) do exame é composta por 180 itens agrupados em quatro grandes áreas, distribuídos em 45 itens por área em dois dias de aplicação. Quanto à logística de aplicação, no primeiro dia são aplicadas as provas de (i) Linguagens, Códigos e suas Tecnologias, Redação e (ii) Ciências Humanas e suas Tecnologias. No segundo dia, são aplicadas as provas de (iii) Ciências da Natureza e suas Tecnologias e (iv) Matemática e suas Tecnologias. As provas são organizadas em cores, sendo que no primeiro dia são rosa, azul, amarela e branca; e no segundo dia, são as mesmas cores do primeiro dia, porém a branca é substituída pela cor cinza.

Em termos de modelagem para a elaboração e cálculo das notas (score) dos candidatos, o ENEM utiliza a Teoria de Resposta ao Item (TRI). Em linhas gerais, “a TRI está ligada a aspectos psicométricos do teste e propõe uma modelagem estatístico-matemática para as características latentes do indivíduo<sup>1</sup> e para os parâmetros associados aos itens” (PITON-GONÇALVES; ALUISIO, 2015). Além dos itens conterem os parâmetros estimados de discriminação e dificuldade, há também o acerto aleatório ou “chute” do candidato. Apesar das excelentes características da TRI, Andrade, Laros & Gouveia (2010) observam que o trabalho nesta área depende dos avanços da computação e que há determinada complexidade da interpretação e nos cálculos inerentes à esta área. Detalhes da TRI são encontrados nos trabalhos de Lord (1980) e Andrade, Tavares & Valle (2000).

### 3 Resultados

Buscando aplicar a TCT nas provas do ENEM e, com o objetivo de descrever os índices de dificuldade e discriminação somados à análise do conteúdo matemático; realizamos a mineração dos microdados do INEP, a classificação numérica e qualitativa e a resolução sistemática de 45 itens do ENEM da edição de 2012.

#### 3.1 Materiais e métodos

Os microdados do INEP (BRASIL, 2012) apuraram a participação de 5.791.065 candidatos para a edição de 2012. Esse número é elevado quando consideramos a capacidade computacional de processamento. Para viabilizar a análise dos dados, extraímos uma amostra que foi selecionada pelo seguinte critério: participantes que realizaram a prova na cidade de São Carlos/SP. O tratamento e a mineração dos 3.8GB (Gigabytes) de dados<sup>2</sup> foram realizados pelos autores desta

---

<sup>1</sup>Habilidade, competência e/ou proficiência são exemplos de traços latentes de um indivíduo.

<sup>2</sup><<http://inep.gov.br/microdados>>.

pesquisa, a partir do desenvolvimento de algoritmos elaborados em linguagem C++<sup>3</sup>. Uma vez minerados, foram classificados 6.826 participantes<sup>4</sup> da cidade de São Carlos/SP. As provas rosa, azul e branca foram mapeadas para a amarela<sup>5</sup>.

Por meio da planilha de cálculo LibreOffice Calc<sup>6</sup>, foram calculados os índices de dificuldade (que seguiu a Equação (1)) e o de discriminação (que seguiu a Correlação Bisserial por Pontos). Foram extraídos dos dados a “distribuição das respostas” que indicam a proporção (em porcentagem) de participantes que assinalaram aquela opção de resposta, de (A) a (E). A extração deu-se a partir do desenvolvimento de algoritmos em linguagem C++.

Resolvemos e analisamos o conteúdo dos 45 itens da prova de Matemática da edição de 2012, o que culminou na classificação dos conteúdos da prova. Enquanto recorte para este artigo, selecionamos, a partir das Tabelas 1 e 2, três itens a saber: (i) item 139, que foi difícil e satisfatório, (ii) item 172, que foi o mais difícil da amostra e ineficiente e (iii) item 164, que foi o segundo mais difícil e necessita de revisão. Os resultados desta pesquisa estão organizados em duas partes, descritos nas Seções 3.2 e 3.3.

### 3.2 Resultado 1: Classificação dos conteúdos

Os conteúdos matemáticos foram classificados de acordo com PCN+ Ensino Médio<sup>7</sup> (BRASIL, 2002), que sistematiza a Matemática do Ensino Médio em três eixos: Números e Funções (eixo A), Geometria e Medidas (eixo B) e Análise de Dados e Probabilidade (eixo C). Dessa forma, associamos os conteúdos curriculares sugeridos no PCN+ Ensino Médio, culminando com o que denominamos de *Grupos de Conhecimento*, sumarizados na Tabela 3.

---

<sup>3</sup>O volume de dados foi grande e tínhamos uma capacidade computacional de processamento limitada, o que impossibilitou o uso de linguagens como R e SAS, por exemplo.

<sup>4</sup>Observação: ser um participante significa que este realizou a prova. Participantes inscritos não necessariamente comparecem à uma prova.

<sup>5</sup>Em todas as cores de provas aparecem os mesmos itens, porém com ordenação trocada.

<sup>6</sup><<http://libreoffice.org>>

<sup>7</sup>Orientações Educacionais complementares aos Parâmetros Curriculares Nacionais. Ciências da natureza, matemática e suas tecnologias.

Tabela 3 – Grupos de Conhecimento da prova amarela do ENEM 2012.

Eixo - Grupo de Conhecimento	Numeração do itens	Representatividade
(A) – Análise de Gráficos	140, 143, 145, 158, 159, 179	13,33%
(A) – Aritmética e Álgebra	142, 144, 146, 153, 155, 156, 157	28,89%
(B) – Geometria Espacial	141, 147, 154, 166, 176	11,11%
(B) – Geometria Plana	148, 149, 151, 152, 162	11,11%
(B) – Grandezas, Unidades de Medida e Escalas	137, 139, 161, 163, 167, 168	13,33%
(C) – Combinatória e Probabilidade	136, 138, 164, 173, 174	11,11%
(C) – Estatística	150, 170, 171, 172, 175 160, 165, 169, 177, 178, 180	11,11%

Fonte: Dados da pesquisa.

### 3.3 Resultado 2: Análise de itens

As próximas seções trazem os itens selecionados para este artigo.

#### 3.3.1 Item 172

Figura 1 – Item 172 da prova amarela

**QUESTÃO 172**

Um produtor de café irrigado em Minas Gerais recebeu um relatório de consultoria estatística, constando, entre outras informações, o desvio padrão das produções de uma safra dos talhões de sua propriedade. Os talhões têm a mesma área de 30 000 m<sup>2</sup> e o valor obtido para o desvio padrão foi de 90 kg/talhão. O produtor deve apresentar as informações sobre a produção e a variância dessas produções em sacas de 60 kg por hectare (10 000 m<sup>2</sup>).

A variância das produções dos talhões expressa em (sacas/hectare)<sup>2</sup> é

- A 20,25.
- B 4,50.
- C 0,71.
- D 0,50.
- E 0,25.

Fonte: Brasil (2012).

O item 172 (Figura 1) é do grupo de conhecimento Estatística e obteve um  $ID = 0,14$ , sendo classificado como *Muito Difícil* e, além disso, foi o item mais difícil da amostra. Um  $IDS = 0,15$  considerou-o como um item *Ineficiente*. A distribuição das respostas foi de (A) 17,30%, (B) 24,80%, (C) 21,61%, (D) 21,66% e (E) 14,40%.



Resolução: utilizando o desvio padrão  $s$  fornecido e realizando as devidas transformações, obtemos que  $s = \frac{90kg}{talhao} = \frac{90kg}{30.000m^2} = \frac{30kg}{10.000m^2} = \frac{30kg}{hectare} = \frac{0,5saca}{hectare}$ . Logo a variância em  $(sacas/hectares)^2$  é  $s^2 = (0,5)^2 = 0,25$ ; indicando (E) como opção correta. Tomamos agora as duas opções mais assinaladas. A (D) não expressa um erro matemático previsto pelo item, mas provavelmente, está associada à consideração, pelo participante, do desvio padrão (sendo que o pedido foi a variância) enquanto resposta. Resultados da pesquisa de Haubert (2009) apontam que estudantes do Ensino Médio apresentam dificuldades na compreensão do significado do desvio padrão.

De acordo com o IBGE (2010), 96% da população da cidade de São Carlos/SP é urbana e, com isso, os termos “saca” e “talhão” podem não ser do cotidiano dos participantes. Para Alvarenga, Andrade & Santos (2016, p. 34) “muitas vezes, a dificuldade não está na matemática, mas em compreender o enunciado de um problema”. Ainda em Smole & Diniz (2001), a não compreensão de um enunciado (ou texto-base) é um fator associado ao erro (de um item, por exemplo).

O teor do texto-base, somado à necessidade de cálculos que envolvem o desvio padrão e variância, podem ter influenciado o item a ser classificado como muito difícil e ineficiente.

### 3.3.2 Item 164

Figura 2 – Item 164 da prova amarela  
QUESTÃO 164

Em um *blog* de variedades, músicas, mantras e informações diversas, foram postados “Contos de Halloween”. Após a leitura, os visitantes poderiam opinar, assinalando suas reações em: “Divertido”, “Assustador” ou “Chato”. Ao final de uma semana, o *blog* registrou que 500 visitantes distintos acessaram esta postagem.

O gráfico a seguir apresenta o resultado da enquete.



O administrador do *blog* irá sortear um livro entre os visitantes que opinaram na postagem “Contos de Halloween”.

Sabendo que nenhum visitante votou mais de uma vez, a probabilidade de uma pessoa escolhida ao acaso entre as que opinaram ter assinalado que o conto “Contos de Halloween” é “Chato” é mais aproximada por

- A** 0,09. **B** 0,12. **C** 0,14. **D** 0,15. **E** 0,18.

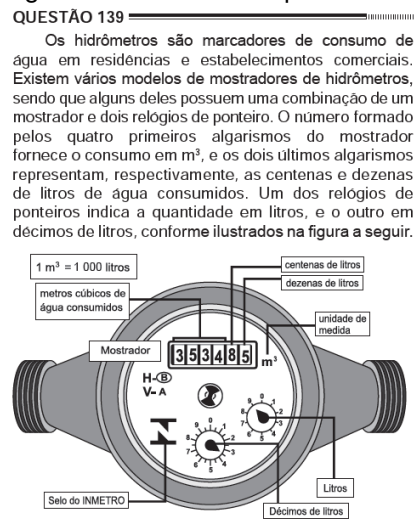
Fonte: Brasil (2012).

O item 164 (Figura 2) é do grupo de conhecimento Combinatória e Probabilidade; e obteve um  $ID = 0,15$ , sendo classificado como *Muito Difícil*. Um  $IDS = 0,23$  considerou-o como um item que *Necessita de Revisão*. A distribuição das respostas foi (A) 16,32%, (B) 53,00%, (C) 8,90%, (D) 15,30% e (E) 6,69%. Observa-se que, mais da metade dos participantes optaram pelo distrator (B), sendo que a opção correta é a opção (D).

Resolução: uma importante informação está no seguinte trecho do enunciado “*O administrador do blog irá sortear um livro entre os visitantes que opinaram na postagem Contos de Halloween*”. Assim retirando os 21% que não opinaram do espaço amostral, obtemos  $100\% - 21\% = 79\%$  e, considerando os 12% que responderam “Chato” na pesquisa como casos favoráveis, a probabilidade é calculada como  $P(E) = \frac{12\%}{79\%} \approx 0,1518$  o que nos indica a opção (D) como correta. Possivelmente, a falta de atenção à leitura do enunciado levou pouco mais da metade dos participantes a não eliminarem os que “não opinaram” do espaço amostral, calculando então a probabilidade como  $P(E) = \frac{12\%}{100\%} \approx 0,12$ , selecionando a opção de resposta (B) (incorreta), conclusão corroborada por Smole & Diniz (2001).

### 3.3.3 Item 139

Figura 3 – Item 139 da prova amarela



Considerando as informações indicadas na figura, o consumo total de água registrado nesse hidrômetro, em litros, é igual a

- A 3 534,85.
- B 3 544,20.
- C 3 534 850,00.
- D 3 534 859,35.
- E 3 534 850,39.

Fonte: Brasil (2012).

O item 139 (Figura 3) é do grupo de conhecimento Grandezas, Unidades de Medida e Escalas; e apresentou um  $ID = 0,35$ , sendo classificado como *Difícil*. Um  $IDS = 0,47$  considerou-o como um item *Satisfatório*, indicando que o item foi capaz de distinguir os participantes que obtiveram melhor rendimento. A distribuição das respostas foi (A) 16,26%, (B) 4,85%, (C) 34,15%, (D) 34,80% e (E) 10,21%.

De acordo com o hidrômetro, notamos que o mostrador indica  $3.534m^3 = 3.534.000l$ , 8 centenas igual a 800l e 5 dezenas igual a 50l. O ponteiro dos litros indica o número 9 e, por fim, o ponteiro dos décimos de litros indica  $3,5 = 0,35l$ . Assim, o consumo total em litros é dado por  $3.534.000 + 800 + 50 + 9 + 0,35 = 3.534.859,35l$ .

Partindo da distribuição das respostas, observamos que as opções (C) (incorreta) e (D) (correta) estão muito próximas em frequência. Provavelmente, a maior parte dos participantes que marcaram a (C) desconsideraram a leitura dos ponteiros do hidrômetro. Em resumo, a Tabela 4 sumariza os índices dos três itens analisados neste artigo.

Tabela 4 – Sumarização dos itens analisados. AC significa que a alternativa correta foi a mais selecionada.

Item	ID	Classificação	ID	IDS	Classificação	IDS	AC
139	0,35	Difícil		0,47	Satisfatório		Sim
164	0,15	Muito Difícil		0,23	Necessita de Revisão		Não
172	0,14	Muito Difícil		0,15	Ineficiente		Não

Fonte: Dados da pesquisa.

### 3.4 Discussão

Observamos, a partir dos resultados desta pesquisa, que a disposição, a qualidade de escrita (do texto-base e enunciado) e a interpretabilidade esperada do participante podem afetar diretamente a eficiência do item. Com base em Smole & Diniz (2001), algumas respostas incorretas podem ocorrer devido a não compreensão de um enunciado. Como exemplo, citamos o item 172 que apresentou a maior dificuldade da amostra, mas com baixa discriminação, indicando ser um item ineficiente.

Nesse contexto, notamos que não há relação direta entre os índices de dificuldade e de discriminação. Para justificar tal posicionamento, destacamos os itens 164 e 139, com seus índices e interpretações na Tabela 5. Ainda na mesma tabela, o item 140 é *Muito Fácil* porém *Necessita de Revisão*.

Tabela 5 – Resultados totais. Ordenados pelo IDS.

Item	ID	Classificação do ID	IDS	Classificação do IDS
173	0,18	Muito Difícil	0,13	Ineficiente
<b>172</b>	<b>0,14</b>	<b>Muito Difícil</b>	<b>0,15</b>	<b>Ineficiente</b>
163	0,22	Muito Difícil	0,19	Ineficiente
174	0,29	Difícil	0,20	Necessita de Revisão
<b>164</b>	<b>0,15</b>	<b>Muito Difícil</b>	<b>0,23</b>	<b>Necessita de Revisão</b>
175	0,16	Muito Difícil	0,23	Necessita de Revisão
176	0,20	Muito Difícil	0,23	Necessita de Revisão
165	0,20	Muito Difícil	0,24	Necessita de Revisão
162	0,47	Normal	0,25	Necessita de Revisão
<b>140</b>	<b>0,90</b>	<b>Muito Fácil</b>	<b>0,27</b>	<b>Necessita de Revisão</b>
177	0,26	Difícil	0,28	Necessita de Revisão
146	0,43	Difícil	0,31	Aceitável
178	0,27	Difícil	0,31	Aceitável
148	0,32	Difícil	0,32	Aceitável
149	0,28	Difícil	0,33	Aceitável
150	0,48	Normal	0,33	Aceitável
161	0,28	Difícil	0,33	Aceitável
179	0,41	Difícil	0,33	Aceitável
138	0,38	Difícil	0,36	Aceitável
147	0,40	Difícil	0,38	Aceitável
169	0,49	Normal	0,40	Satisfatório
141	0,79	Muito Fácil	0,42	Satisfatório
170	0,55	Fácil	0,42	Satisfatório
166	0,31	Difícil	0,43	Satisfatório
167	0,75	Muito Fácil	0,43	Satisfatório
142	0,73	Fácil	0,44	Satisfatório
154	0,61	Fácil	0,45	Satisfatório
157	0,63	Fácil	0,45	Satisfatório
160	0,79	Muito Fácil	0,45	Satisfatório
180	0,29	Difícil	0,45	Satisfatório
151	0,41	Difícil	0,46	Satisfatório
<b>139</b>	<b>0,35</b>	<b>Difícil</b>	<b>0,47</b>	<b>Satisfatório</b>
143	0,67	Fácil	0,47	Satisfatório
152	0,38	Difícil	0,48	Satisfatório
153	0,41	Difícil	0,49	Satisfatório
144	0,70	Fácil	0,50	Satisfatório
171	0,53	Normal	0,50	Satisfatório
159	0,56	Fácil	0,52	Satisfatório
145	0,65	Fácil	0,53	Satisfatório
136	0,40	Difícil	0,54	Satisfatório
137	0,40	Difícil	0,55	Satisfatório
155	0,46	Normal	0,55	Satisfatório
168	0,39	Difícil	0,55	Satisfatório
156	0,53	Normal	0,56	Satisfatório
158	0,59	Fácil	0,56	Satisfatório

Fonte: Dados da pesquisa.

De acordo com a Tabela 6, 57,78% dos itens foram classificados como difíceis e muito difíceis e, por outro lado, 28,89% como fáceis e muito fáceis. Portanto, podemos afirmar que o nível geral da prova analisada foi considerado difícil.

Tabela 6 – Taxa do ID dos 45 itens desta pesquisa.

<b>Classificação</b>	<b>Taxa</b>
Muito Difícil	15,56%
Difícil	42,22%
Normal	13,33%
Fácil	20,00%
Muito Fácil	8,89%
Total	100,0%

Fonte: Dados da pesquisa.

Por outro lado, a Tabela 7 nos indica que três itens (6,67%) poderiam ter sido excluídos, tomando como base a Tabela de Ebel (1954). O item 172, por exemplo, indicou que a alternativa menos assinalada foi a correta, o que pode ser uma preocupação em relação aos resultados de um teste.

Tabela 7 – Taxa do IDS dos 45 itens desta pesquisa.

<b>Classificação</b>	<b>Taxa</b>
Aceitável	20,00%
Satisfatório	55,56%
Necessita de Revisão	17,78%
Ineficiente	6,67%
Total	100,0%

Fonte: Dados da pesquisa.

#### 4 Considerações finais

Os resultados quantitativos despertaram o interesse dos pesquisadores na investigação dos motivos educacionais e de conteúdo matemático dos itens envolvidos nesta pesquisa e, com isso, verificar o porquê dos participantes escolherem uma opção de resposta em detrimento de outras. E, ainda, por-que em muitos casos a opção mais assinalada não é a correta. O índice de dificuldade sozinho não nos fornece informações necessárias para a análise do item, sendo também necessário o índice de discriminação.

Consideramos que a análise de um item em um teste deve contemplar aspectos estatísticos, de conteúdo específico e educacionais. Nesse sentido, esta pesquisa traz subsídios e uma discussão para a análise de itens via índices de dificuldade e discriminação interpretados em cenários educacionais, contemplando o conteúdo matemático. Além disso, espera-se que os resultados das Tabelas 5 e 3 contribuam para a realização de outras pesquisas acadêmicas, uma vez que os itens da prova do ENEM 2012 são disponibilizados publicamente, possibilitando outras conclusões do ponto de vista educacional. Enquanto trabalho futuro, tem-se como objetivo analisar outras edições do ENEM e desenvolver outros algoritmos com o objetivo de minerar os dados populacionais.

## Referências

- ALVARENGA, K. B.; ANDRADE, I. D.; SANTOS, R. de J. Dificuldades na resolução de problemas básicos de matemática: um estudo de caso do agreste sergipano. **Amazônia - Revista de Educação em Ciências e Matemática**, v. 12, n. 24, p. 39–52, 2016.
- ANDRADE, D. F.; TAVARES, H. R.; VALLE, R. da C. **Teoria de Resposta ao Item: Conceitos e Aplicações**. [S.l.]: Associação Brasileira de Estatística, São Paulo, 2000. 154 p.
- ANDRADE, J. M. d.; LAROS, J. A.; GOUVEIA, V. V. O uso da teoria de resposta ao item em avaliações educacionais: diretrizes para pesquisadores. **Avaliação Psicológica**, scielopepsic, v. 9, p. 421 – 435, 12 2010.
- BRASIL. **PCN+ Ensino Médio: Orientações Educacionais complementares aos Parâmetros Curriculares Nacionais. Ciências da natureza, matemática e suas tecnologias**. [S.l.]: Secretaria de Educação Média e Tecnológica, MEC, SEMTEC, Brasília, 2002.
- BRASIL. **Prova de Matemática e suas Tecnologias - ENEM 2012**. INEP - Brasília - DF, 2012.
- BRASIL. **ENEM - Documento Básico**. INEP - Brasília - DF: [s.n.], 2018. Disponível em: <<<http://portal.inep.gov.br>>>. Acesso em: 10 set. 2018.
- CERDÁ, E. **Psicometria General**. 2. ed. [S.l.]: Editorial Herder, 1978. 190 p.
- EBEL, R. L. Procedures for the analysis of classroom tests. **Educational and Psychological Measurement**, v. 14, n. 2, p. 352–364, 1954.
- ERTHAL, T. C. **Manual de Psicometria**. 7a. ed. [S.l.]: Jorge Zahar Ed., 2001. 152 p.
- HAUBERT, M. C. **Uma Medida Saudável: Uma abordagem de Educação Estatística no Ensino Médio associada à área da Saúde**. 234 p. Dissertação (Mestrado) — UFRGS, Universidade Federal do Rio Grande do Sul, Porto Alegre, Brasil, 2009.
- IBGE. **Censo da Cidade de São Carlos-SP - 2010**. 2010. Disponível em: <<<https://cidades.ibge.gov.br>>>. Acesso em: 15 out. 2016.

- LEITE, R. H. **Uso da Teoria Clássica dos Testes - TCT e da Teoria de Resposta ao Item - TRI na avaliação da qualidade métrica de testes de seleção**. 325 p. Dissertação (Doutorado) — Universidade Federal do Ceará, Fortaleza, Brasil, 2009.
- LORD, F. M. **Application of Item Response Theory to Practical Testing Problems**. first. Hillsdale, New Jersey, EUA: Lawrence Erlbaum Associates, 1980. 274 p.
- LUDKE, M. A caminho de uma sociologia da avaliação escolar. **Educação e Seleção**, n. 16, p. 351–371, 1987.
- PANALOZA, J. L.; VARGAS-PEREZ, C. Big-data and the challenges for statistical inference and economics teaching and learning. **Multidisciplinary Journal for Education, Social and Technological Sciences**, MUSE, v. 4, p. 64–87, 2017.
- PASQUALI, L. Psicometria. **Revista da Escola de Enfermagem da USP**, v. 43, n. spe, Dec. 2009.
- PASQUALI, L. **Psicometria: Teoria dos Testes na Psicologia e na Educação**. [S.l.]: Vozes, 2017. 392 p.
- PASQUALI, L.; PRIMI, R. Fundamentos da teoria de resposta ao item - tri. **Avaliação Psicológica**, v. 2, p. 99–110, 2003.
- PITON-GONÇALVES, J. **Desafios e perspectivas da implementação computacional de testes adaptativos multidimensionais para avaliações educacionais**. Tese (Tese de Doutorado) — Instituto de Ciências Matemáticas e de Computação, Universidade de São Paulo, São Carlos, São Carlos, 2012.
- PITON-GONÇALVES, J.; ALUISIO, S. M. A. Teste adaptativo computadorizado multidimensional com propósitos educacionais: princípios e métodos. **Ensaio: Avaliação e Políticas Públicas em Educação**, scielo, v. 23, p. 389–414, 06 2015. ISSN 0104-4036.
- PITON-GONÇALVES, J.; SOUZA, E. R.; LAMONATO, M. Logaritmos, trigonometria, matrizes e determinantes: uma análise de itens do enem do ponto de vista curricular. **Educação Matemática em Revista - RS**, n. 18, 2017. ISSN 1518-8221.
- RIBEIRO, J. L. L. de S. Avaliação educacional desatando e reatando nós. In: \_\_\_\_\_. [S.l.]: EDUFBA. SciELO Books, 2009. cap. A avaliação como uma política pública: aspectos da implementação do SINAES, p. 91–114.
- SANTOS, J. S.; CORTELAZZO, A. L. Os conteúdos de biologia celular no exame nacional do ensino médio - enem. **Avaliação: Revista da Avaliação da Educação Superior (Campinas)**, scielo, v. 18, p. 591 – 612, 11 2013. ISSN 1414-4077.
- SILVA, F.; SANTIAGO, M.; SANTOS, M. dos. Análise de itens da prova de matemática e suas tecnologias do enem que envolvem o conceito de números racionais à luz dos seus significados e representações. **Revemat: Revista Eletrônica de Educação Matemática, Edição Especial**, v. 8, 2013.
- SILVEIRA, P. M. **Criação de um índice de satisfação com a vida por meio da teoria da resposta ao item e fatores associados em trabalhadores brasileiros**. 115 p. Dissertação (Mestrado) — Universidade Federal de Santa Catarina, Florianópolis, Brasil, 2015.

SMOLE, K. S.; DINIZ, M. I. **Ler, escrever e resolver problemas: habilidades básicas para aprender matemática**. Porto Alegre: Artmed, 2001.

TRAVITZKI, R. Avaliação da qualidade do enem 2009 e 2011 com técnicas psicométricas. **Estudos em Avaliação Educacional**, v. 28, n. 67, p. 256–288, jan-abr 2017.