



**UNIVERSIDADE FEDERAL DO CEARÁ**  
**CENTRO DE CIÊNCIAS**  
**DEPARTAMENTO DE ESTATÍSTICA E MATEMÁTICA APLICADA**  
**PROGRAMA DE PÓS-GRADUAÇÃO EM MODELAGEM E MÉTODOS**  
**QUANTITATIVOS**

**FERNANDO MARCIANO DE ALMEIDA**

**AVALIAÇÃO DA EQUIDADE DOS CADERNOS DE MATEMÁTICA DO ENEM 2017**

**FORTALEZA**

**2020**

FERNANDO MARCIANO DE ALMEIDA

AVALIAÇÃO DA EQUIDADE DOS CADERNOS DE MATEMÁTICA DO ENEM 2017

Dissertação apresentada ao Programa de Pós-Graduação em Modelagem e Métodos Quantitativos da Universidade Federal do Ceará, como requisito parcial à obtenção do título de Mestre em Modelagem e Métodos Quantitativos. Área de concentração: Interdisciplinar.

Orientador: Prof. Dr. André Jalles Monteiro.

FORTALEZA

2020

Dados Internacionais de Catalogação na Publicação  
Universidade Federal do Ceará  
Biblioteca Universitária  
Gerada automaticamente pelo módulo Catalog, mediante os dados fornecidos pelo(a) autor(a)

---

- A446a Almeida, Fernando Marciano de.  
Avaliação da equidade dos cadernos de matemática do Enem 2017 / Fernando Marciano de Almeida. –  
2019.  
110 f. : il. color.
- Dissertação (mestrado) – Universidade Federal do Ceará, Centro de Ciências, Programa de Pós-Graduação  
em Modelagem e Métodos Quantitativos, Fortaleza, 2019.  
Orientação: Prof. Dr. André Jalles Monteiro.
1. Equidade. 2. Enem . 3. Cadernos. I. Título.

CDD 510

---

FERNANDO MARCIANO DE ALMEIDA

AVALIAÇÃO DA EQUIDADE DOS CADERNOS DE MATEMÁTICA DO ENEM 2017

Dissertação apresentada ao Programa de Pós-Graduação em Modelagem e Métodos Quantitativos do Departamento de Estatística e Matemática Aplicada da Universidade Federal do Ceará, como requisito parcial à obtenção do título de Mestre em Modelagem e Métodos Quantitativos. Área de concentração: Interdisciplinar.

Aprovada em: \_\_\_/\_\_\_/\_\_\_\_\_.

BANCA EXAMINADORA

---

Prof. Dr. André Jalles Monteiro (Orientador)  
Universidade Federal do Ceará (UFC)

---

Prof. Dr. Rafael Bráz Azevedo Farias  
Universidade Federal do Ceará (UFC)

---

Prof. Dr. Wagner Bandeira Andriola  
Universidade Federal do Ceará (UFC)

A Deus.

Aos meus pais, a minha esposa Aline, aos meus filhos (Iuri e Ícaro), aos colegas da escola Alice Moreira e a todos que contribuíram de alguma forma para que este trabalho pudesse ser construído.

## **AGRADECIMENTOS**

“O presente trabalho foi realizado com apoio da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior – Brasil (CAPES) – Código de Financiamento 001”.

Ao Prof. Dr. André Jalles Monteiro, pela excelente orientação.

Aos professores participantes da banca examinadora, Dr. Rafael Bráz Azevedo Farias e Dr. Wagner Bandeira Andriola pelo tempo, pelas valiosas colaborações e sugestões.

Aos professores consultados, pelo tempo concedido e pelas contribuições em minha pesquisa e formação.

Aos colegas da turma de mestrado, pelas reflexões, críticas e sugestões recebidas.

“Devemos tratar igualmente os iguais e desigualmente os desiguais, na medida de sua desigualdade.”

**Aristóteles**

## RESUMO

O Enem é atualmente o exame de larga escala mais divulgado no país pelo fato de ser a porta de entrada principal em universidades públicas do Brasil. Em avaliações de larga escala são utilizados cadernos paralelos com o objetivo de evitar fraudes no processo avaliativo. No Enem há quatro tipos de cadernos (ROSA, AZUL, CINZA e AMARELO), esses cadernos devem promover a equidade de dificuldades para que os candidatos que recebem cadernos de cores diferentes não sejam prejudicados ou beneficiados na avaliação. Os especialistas em avaliação educacional utilizam os dados fornecidos pelo Enem para avaliar sistemas de ensino, falhas ou avanços no processo avaliativo e falhas na construção dos cadernos. Há várias formas de avaliar um sistema de ensino ou método avaliativo e nenhuma é completamente correta, vários pontos de vista podem ser construídos mas poucos levam em conta as condições educacionais oferecidas ao educando. Em um país com dimensões continentais, diferentes condições educacionais, diferentes climas, diferentes culturas e com diferentes condições socioeconômicas, torna-se muito difícil construir avaliações com equidade de dificuldade em todos os cadernos paralelos utilizados pelo exame, isso se deve ao fato de ser praticamente impossível considerar todos os fatores que causam viés na proficiência dos candidatos. Este trabalho utilizou o arquivo de Microdados do Enem 2017 com ênfase nos cadernos de Matemática para verificar a equidade do processo avaliativo. Foi construída uma argumentação baseada em estatísticas descritivas, ANOVA, DMS de Fischer, funcionamento diferencial do item (DIF), proporção de acertos e análise pedagógica de itens. Foi verificado que uma maior quantidade de acertos no caderno não determina uma maior proficiência, que alguns itens possuem sua dificuldade alterada em função de sua posição no caderno, que os cadernos que apresentam equidade são o AMARELO e CINZA, por outro lado os que apresentam maiores diferenças são ROSA e AMARELO, logo a equidade de dificuldade não existe entre todos os cadernos contrariando o que afirma o Inep.

**Palavras-chave:** Equidade. Enem. Cadernos.

## ABSTRACT

Enem is currently the most popular large-scale exam in the country because it is the main gateway to public universities in Brazil. In large-scale evaluations, parallel notebooks are used in order to avoid fraud in the evaluation process. In Enem there are four types of notebooks (PINK, BLUE, GRAY and YELLOW), these notebooks must promote equity of difficulties so that candidates who receive notebooks of different colors are not harmed or benefited in the evaluation. The specialists in educational evaluation use the data provided by Enem to evaluate teaching systems, flaws or advances in the evaluation process and flaws in the construction of the notebooks. There are several ways to evaluate a teaching system or evaluation method and none is completely correct, several points of view can be constructed but few take into account the educational conditions offered to the student. In a country with continental dimensions, different educational conditions, different climates, different cultures and with different socioeconomic conditions, it is very difficult to construct assessments with equity of difficulty in all the parallel books used by the exam, this is due to the fact that it is practically impossible to consider all the factors that cause bias in the candidates' proficiency. This work used the Enem Microdata 2017 file with an emphasis on Mathematics notebooks to verify the equity of the evaluation process. An argument was made based on descriptive statistics, ANOVA, Fischer DMS, differential item functioning (DIF), proportion of correct answers and pedagogical analysis of items. It was found that a greater number of correct answers in the notebook does not determine greater proficiency, that some items have their difficulty altered due to their position in the notebook, that the notebooks that present equity are YELLOW and GRAY, on the other hand those that present greater differences are PINK and YELLOW, therefore the difficulty equity does not exist among all the notebooks, contrary to what Inep affirms.

**Keywords:** Equity. Enem. Notebooks.

## LISTA DE FIGURAS

Figura 1	– Componentes do escore .....	27
Figura 2	– Distribuição da dificuldade por níveis .....	29
Figura 3	– Curva de proporção de acertos acumulado por nível de escore .....	32
Figura 4	– CCI para o modelo 3PL .....	34
Figura 5	– CCI de 4 itens diferentes .....	35
Figura 6	– Histórico de inscritos confirmados no Enem .....	42
Figura 7	– Possível distribuição dos cadernos do Enem por cor .....	43
Figura 8	– Partes constitutivas do item .....	46
Figura 9	– Tipos de DIF .....	49
Figura 10	– Distribuição das médias para três grupos .....	56
Figura 11	– Proporção de acertos por item e por caderno .....	68
Figura 12	– Item 1 .....	71
Figura 13	– Uma solução para o item 1 .....	72
Figura 14	– Item 2 .....	73
Figura 15	– Item 3 .....	74
Figura 16	– Item 4 .....	76
Figura 17	– Item 5 .....	77
Figura 18	– Item 6 .....	79
Figura 19	– Item 7 .....	81
Figura 20	– Termômetros com leituras .....	82
Figura 21	– Item 8 .....	83
Figura 22	– Caminho do pivô .....	84
Figura 23	– Item 9 .....	86
Figura 24	– Item 10 .....	88

Figura 25	– Item 11 .....	89
Figura 26	– Gráfico da função cosseno .....	90
Figura 27	– Item 12 .....	91
Figura 28	– Triângulo retângulo .....	92

## LISTA DE EQUAÇÕES

Equação 1 – Dificuldade do item .....	28
Equação 2 – Coeficiente de correlação bisserial .....	30
Equação 3 – Coeficiente de correlação bisserial por pontos .....	31
Equação 4 – Probabilidade de acerto para o modelo 4 PL .....	36
Equação 5 – Probabilidade de acerto para o modelo 3 PL .....	38
Equação 6 – Probabilidade de acerto para o modelo 2 PL .....	38
Equação 7 – Probabilidade de acerto para o modelo 1 PL .....	38
Equação 8 – Estatística alpha Mantel Haenszel .....	51
Equação 9 – Estatística qui-quadrado Mantel Haenszel .....	52
Equação 10 – Esperança de $A_j$ .....	53
Equação 11 – Variância de $A_j$ .....	53
Equação 12 – Teste t .....	55
Equação 13 – Soma dos quadrados total .....	57
Equação 14 – Soma dos quadrados de dentro .....	57
Equação 15 – Soma dos quadrados entre grupos .....	57
Equação 16 – Média quadrática entre grupos .....	57
Equação 17 – Média quadrática de dentro dos grupos .....	58
Equação 18 – Teste F .....	58
Equação 19 – DMS de Fischer .....	58

## LISTA DE GRÁFICOS

Gráfico 1 – Proporção de acertos (ordem normal) .....	61
Gráfico 2 – Proporção de acertos (ordem da prova AMARELA) .....	62
Gráfico 3 – DIF (AMARELA(REF)-AZUL(FO)) .....	63
Gráfico 4 – DIF (AZUL(REF)-CINZA(FO)) .....	63
Gráfico 5 – DIF (AZUL(REF)-ROSA(FO)) .....	64
Gráfico 6 – DIF (AMARELA(REF)-ROSA(FO)) .....	64
Gráfico 7 – DIF (CINZA(REF)-ROSA(FO)) .....	65
Gráfico 8 – DIF (AMARELA(REF)-CINZA(FO)) .....	65
Gráfico 9 – Classificação dos itens por nível de dificuldade .....	66

## LISTA DE TABELAS

Tabela 1 – Classificação de dificuldade .....	29
Tabela 2 – Classificação do índice de discriminação .....	30
Tabela 3 – Proporção de acertos do item 45(PROVA AZUL) .....	32
Tabela 4 – Tabela de frequências observadas de resposta a um item .....	51
Tabela 5 – Gabaritos por prova .....	60
Tabela 6 – Estatísticas descritivas .....	60
Tabela 7 – O teste ANOVA .....	60
Tabela 8 – DMS de Fischer .....	61
Tabela 9 – Itens por prova (ordenados pela prova AMARELA) .....	67
Tabela 10 – Quantidade de itens por dificuldade nos cadernos .....	67
Tabela 11 – Quantidade de itens com maior proporção de acerto .....	69
Tabela 12 – Posição do item e proporção de acerto por prova .....	70

## **LISTA DE ABREVIATURAS E SIGLAS**

ABNT	Associação Brasileira de Normas Técnicas
TCT	Teoria Clássica dos Testes
TRI	Teoria de Resposta ao Item
NBR	Norma Brasileira Regulamentar
UFC	Universidade Federal do Ceará
SIBI	Sistema Integrado de Bibliotecas
SISU	Sistema informatizado de seleção unificada
DIF	Funcionamento diferencial do item
LDB	Lei de diretrizes e bases
UFRGS	Universidade Federal do Rio Grande do Sul
REF	Grupo de referência
FO	Grupo focal
ANA	Avaliação Nacional da Alfabetização
ANRESC	Avaliação nacional do rendimento escolar

## LISTA DE SÍMBOLOS

$\theta$	Theta
$\alpha$	Alpha
$\pi$	Pi
$\chi^2$	Qui-quadrado
$\Delta$	Delta
$\exists$	Existe

## SUMÁRIO

<b>1</b>	<b>INTRODUÇÃO</b> .....	<b>17</b>
<b>2</b>	<b>AVALIAÇÃO EDUCACIONAL</b> .....	<b>21</b>
<b>3</b>	<b>TEORIAS PSICOMÉTRICAS</b> .....	<b>26</b>
<b>3.1</b>	<b>A Teoria Clássica dos Testes</b> .....	<b>27</b>
<b>3.1.1</b>	<b>Os postulados da TCT</b> .....	<b>27</b>
<b>3.1.2</b>	<b>A dificuldade do item</b> .....	<b>28</b>
<b>3.1.3</b>	<b>O índice de discriminação</b> .....	<b>30</b>
<b>3.1.4</b>	<b>A curva de proporção do item</b> .....	<b>31</b>
<b>3.2</b>	<b>A Teoria de Resposta ao Item</b> .....	<b>33</b>
<b>3.2.1</b>	<b>A curva característica do item</b> .....	<b>34</b>
<b>3.2.2</b>	<b>Os modelos logísticos</b> .....	<b>36</b>
<b>4</b>	<b>O ENEM</b> .....	<b>39</b>
<b>4.1</b>	<b>O Enem 2017</b> .....	<b>41</b>
<b>4.2</b>	<b>A construção do item</b> .....	<b>43</b>
<b>5</b>	<b>O FUNCIONAMENTO DIFERENCIAL DO ITEM (DIF)</b> .....	<b>48</b>
<b>5.1</b>	<b>Os métodos Mantel Haenszel</b> .....	<b>50</b>
<b>5.2</b>	<b>O cérebro em testes de rendimento</b> .....	<b>53</b>
<b>6</b>	<b>ANOVA</b> .....	<b>55</b>
<b>7</b>	<b>MATERIAL E MÉTODOS</b> .....	<b>59</b>
<b>7.1</b>	<b>Dos arquivos gerados</b> .....	<b>59</b>
<b>7.2</b>	<b>Dos resultados</b> .....	<b>64</b>
<b>8</b>	<b>ANÁLISE PEDAGÓGICA</b> .....	<b>71</b>
<b>8.1</b>	<b>A sequência de itens por nível de dificuldade</b> .....	<b>75</b>
<b>9</b>	<b>CONCLUSÃO</b> .....	<b>93</b>
	<b>REFERÊNCIAS</b> .....	<b>95</b>
	<b>APÊNDICE A – TABELA DE RESULTADOS DIF QUI-QUADRADO</b> .....	<b>102</b>
	<b>APÊNDICE B – TABELA DE RESULTADOS DIF ALPHA MH</b> .....	<b>104</b>
	<b>APÊNDICE C – TABELA DE RESULTADOS DIF QUI-QUADRADO</b> .....	<b>106</b>
	<b>APÊNDICE D – TABELA DE RESULTADOS DIF ALPHA MH</b> .....	<b>108</b>

## 1 INTRODUÇÃO

No mundo globalizado, no qual ocorrem diversas mudanças sociais, econômicas, políticas, tecnológicas e culturais, um dos grandes desafios é diminuir as desigualdades existentes entre as nações. Um dos caminhos mais percorridos para amenizar as diferenças é o da educação. “A educação é a arma mais poderosa que você pode usar para mudar o mundo”, (NELSON MANDELA, 2003, Apud GOMES, 2019).

O governo brasileiro vem promovendo melhorias na área educacional nas três últimas décadas. As novidades são positivas e trouxeram esperança à parcela da população que depende do sistema educacional público (BRASIL, 1988). As mudanças se iniciaram no país com a Constituição de 1988 e, na área da educação, as transformações passaram a ocorrer principalmente a partir de 1996 com a nova Lei de Diretrizes e Bases (LDB) (BRASIL, 1996), pois esta promoveu a universalização do ensino fundamental, a criação de fundos para manutenção e desenvolvimento da educação, a criação do ensino de jovens e adultos, a educação indígena e quilombola, a inclusão, entre outros fatores que passaram a fomentar a melhoria da qualidade do ensino.

A qualidade da educação é o objetivo primordial de qualquer sistema educacional e, para isso, é preciso analisar dados que sirvam como base de sustentação para políticas educacionais, estabelecer metas a serem alcançadas e disponibilizar recursos favoráveis a fim de que os objetivos possam ser alcançados.

No contexto brasileiro, a busca por uma melhor qualidade de ensino teve ênfase a partir dos anos 2000 devido a avaliações qualitativas e quantitativas no setor. Dentre essas avaliações, destaca-se o Enem (Exame Nacional do Ensino Médio), criado em 1998, com o objetivo de avaliar os discentes que concluem o ensino médio. Atualmente, um bom resultado obtido no Enem tornou-se meta a ser alcançada pela maior parte dos egressos do ensino médio. Isso se deve ao fato de as universidades brasileiras adotarem a nota do exame como critério de seleção de novos alunos.

Ao longo dos anos, o Enem foi ganhando notoriedade nos meios de comunicações devido aos bons resultados dos alunos das escolas particulares. Isso ocorreu porque tais resultados passaram a ser utilizados por essas escolas em propagandas com a finalidade de classificá-las entre si, promovendo, assim, os centros educacionais que contratam os serviços midiáticos.

O Inep afirma que os resultados repassados as escolas e disponíveis em seus domínios servem para que sejam estudados (pelas escolas, núcleo gestor e professores) e aproveitados

para implementação de melhorias no processo ensino-aprendizagem (INEP, 2018). O processo de ranqueamento pode esconder possíveis evoluções no processo educacional nas escolas e não leva em conta os recursos educacionais, financeiros, sociais e culturais ao qual os alunos e as escolas estão inseridos.

A nova atribuição do Enem é tema de discussão entre diversos especialistas em avaliação escolar, eles afirmam ser prejudicial ao exame uma dupla finalidade com vários objetivos distintos. Em seu documento básico de criação o Enem é descrito como avaliador do ensino básico, porém avaliar não se trata somente de classificar através de números obtidos em testes padronizados (ou não) de proficiência, avaliar é algo complexo e contínuo que busca tornar satisfatório o processo de ensino-aprendizagem (LUCKESI, 2011).

O Enem é aplicado por meio de quatro cadernos (ROSA, AZUL, CINZA E AMARELO) com itens iguais mas com posições diferentes entre os cadernos. As características dos itens são determinadas através da Teoria de resposta ao item (TRI) que garante equidade de dificuldade dos cadernos (ONU, 2010). A Teoria de resposta ao item é a teoria utilizada para seleção dos itens e estimação de seus parâmetros, é fortemente defendida por cientistas e pesquisadores por todo o mundo (INEP, 2011). O senso comum infere que deve haver diferentes dificuldades entre os cadernos do Enem, pois há fatores como a posição dos itens nos cadernos que podem prejudicar a verdadeira proficiência do candidato.

O fato de haver equidade (INEP, 2011) entre cadernos é relevante, pois atualmente uma pequena diferença de pontuação pode determinar o ingresso em uma universidade pública. Em 2014 a maior diferença registrada entre alunos cotistas e não cotistas em cursos muito concorridos na UFRGS (Universidade Federal do Rio Grande do Sul) era de até 180 pontos (G1, 2017). Em 2016 a diferença máxima entre os 10 cursos mais concorridos no SISU (sistema informatizado de seleção unificada) foi de no máximo 4% como afirmou o então ministro da educação Aluísio Mercadante (GGN, 2016).

Em 2019 para o curso de medicina da Fundação Universidade Federal de Ciências da Saúde de Porto Alegre (UFCSPA), foi observada uma diferença de 8,64 pontos entre a maior nota de um candidato cotista e a do último colocado não cotista, além disso, a diferença média entre as notas dos cotistas foi de 2,81 pontos para o mesmo curso, logo uma pequena diferença na pontuação determinou o ingresso do candidato em um curso concorrido (GGN, 2016). Essa realidade pode ser observada em vários outros cursos das universidades brasileiras.

Este trabalho tem como objetivo principal analisar a equidade de dificuldade entre os cadernos do Enem 2017, visto que a vaga em uma universidade pública é objeto de desejo de

milhões de concludentes do ensino médio que devem ter as mesmas oportunidades de ingresso. O estudo do item tratará de verificar o efeito de DIF (funcionamento diferencial do item) analisando como a dificuldade do item (nos cadernos) é afetada com a mudança de posição. Os distratores identificados serão avaliados, as sequências de itens com dificuldades semelhantes (com baixo ou alto grau de dificuldade) em pontos extremos (início ou fim) dos cadernos serão avaliadas.

O capítulo 1 apresenta de forma breve a importância da educação, algumas considerações sobre o Enem, o objetivo do trabalho, a metodologia a ser utilizada para verificar os resultados e as ideias discutidas nos capítulos. Este trabalho abordará conceitos sobre avaliação educacional, Psicometria, Enem (histórico e evolução), viés do item, teste ANOVA e a visão pedagógica sobre alguns itens. O trabalho está dividido em nove capítulos que se interligam por diversas vezes e resultam em conclusões no último capítulo.

No capítulo 2, apresentamos diferentes conceitos sobre avaliação educacional, que serviram de base para o presente trabalho. As posições abordadas sobre avaliação educacional, complementadas pelas ideias que serão posteriormente apresentadas, resultaram em uma sólida defesa do trabalho. O foco de discussão desse capítulo é a equidade educacional, cujo conceito contemplamos tradicionalmente de modo filosófico, bem como tratamos de sua variação no âmbito educacional.

O capítulo 3 abordará as Teorias Psicométricas com posições de vários cientistas sobre as vantagens e desvantagens de cada uma.

O histórico do Enem é retratado no quarto capítulo, no qual estudamos a sua elaboração, a escolha dos itens, a distribuição dos cadernos no dia de sua aplicação, as atribuições do exame e tratamos especificamente sobre o Enem 2017.

Discursamos, no capítulo 5, sobre o Funcionamento Diferencial do Item (DIF), no qual atentamos principalmente para a ideia sobre viés do item. Utilizamos procedimentos estatísticos da literatura para verificar a existência de DIF entre itens idênticos em cadernos diferentes. Desse modo, analisaremos se o senso comum está correto quando avalia que a mudança de posição do item causa diferença significativa de dificuldade no Enem.

No capítulo 6, apresentamos o teste ANOVA e definimos cada equação utilizada, assim como a forma de interpretar o referido teste.

No capítulo 7, apresentamos os motivadores de seleção dos dados, bem como os resultados obtidos por meio dos diversos testes. Nesse capítulo, tecemos comentários acerca dos arquivos gerados inicialmente, os métodos utilizados na geração dos resultados e realizamos uma breve análise sobre cada resultado. Dessa maneira, o leitor pode compreender

os resultados desse trabalho.

Discutiremos os 3 itens identificados com maiores DIF no capítulo 8, e analisaremos, de forma pedagógica, as duas sequências de itens que podem ter causado as diferenças de proficiência média entre candidatos, que responderam cadernos diferentes.

Por fim, no último capítulo apresentamos as conclusões, baseadas nos teóricos que fundamentam este trabalho.

## 2 AVALIAÇÃO EDUCACIONAL

O ato de avaliar é característica natural do ser humano. A todo instante estamos criando métodos baseando-se em diversos parâmetros com a finalidade de obter a melhor escolha para um determinado objetivo. Ao sair de casa podemos escolher o melhor caminho para um determinado destino, escolher o que comer ou simplesmente avaliar um produto pelo que conhecemos sobre o mesmo. “A avaliação tornou-se campo do conhecimento e prática social funcional ao modo de gestão estatal nas sociedades capitalistas avançadas nos últimos 60 anos” (SILVA, GOMES, 2018, p. 03).

O processo de avaliação pode ser aplicado em diversas áreas do conhecimento mostrando-se essencial ao surgimento de novas técnicas, novas tecnologias, modelos de gestão, desenvolvimento e surgimento de áreas do conhecimento e outros. Ao se avaliar algo é necessário responder algumas perguntas como: O que está sendo avaliado? Quais os parâmetros utilizados para avaliar? O método utilizado serve para avaliar todos os objetos (ou elementos) de mesmas características com a mesma precisão?

O ato de avaliar nas ciências do homem é uma atividade complexa e fundamenta-se no pensamento descritivo, crítico e analítico (VIANNA, 1992). Avaliar pode ser simples ou complexo como ocorre no âmbito educacional. “A avaliação educacional é um sistema de informações que tem como objetivos fornecer diagnóstico e subsídios para a implementação ou manutenção de políticas educacionais” (KLEIN, FONTANIVE, 1995, p. 01).

A avaliação deve ser um processo sistemático de coleta e análise de informações, que deve proporcionar pelo menos duas medidas: uma no início do processo de ensino e outra, no final, posto que sua finalidade última é a emissão de um juízo de valor acerca das mudanças de comportamento decorrentes do aprendizado. (ANDRIOLA, 2018, p. 68)

A avaliação educacional atingindo seus objetivos gerais contribui de forma significativa com o crescimento social, científico e tecnológico do país. Desenvolver o educando como pessoa e mostrar à sociedade a qualidade do processo educacional é função da avaliação (LUCKESI, 2008).

No campo educacional pode-se perceber que em geral as pessoas utilizam o conceito de avaliação educacional como forma de avaliar o rendimento do aluno, ou seja, utilizam a ideia de avaliação como algo pontual atrelado a uma nota, essa forma de pensar não é completamente incorreta mas carece de muitas informações que podem ser obtidas através de

uma análise mais complexa sobre os resultados. Para Gatti (2002), As pessoas são levadas a acreditar que avaliar é somente medir conhecimento devido ao destaque que a mídia faz sobre os exames de vestibular, provão e Enem.

Segundo Gatti (2002, p. 02),

A Avaliação Educacional hoje não é apenas um campo com teorias, processos e métodos específicos, mas também um campo abrangente que comporta subáreas, com características diferentes: avaliação de sistemas educacionais, de desempenho escolar em sala de aula, de rendimento escolar com objetivo de microanálises de programas, avaliação institucional e auto-avaliação.

No âmbito escolar o processo avaliativo pode ocorrer de diversas formas englobando comportamento de alunos, desempenho em testes padronizados, participação em atividades escolares, evolução frente a um conteúdo específico e outros (MACDONALD, 2003).

Os métodos de avaliação que ocorrem dentro da escola são englobados no processo conhecido como Microavaliação (PAYNE, 1974). A Microavaliação tem como foco o rendimento do aluno e o processo instrucional, é fundamental para o progresso do educando, pois possibilita ao docente identificar alunos com dificuldades de aprendizagem e elaborar estratégias pedagógicas para diminuir e (ou) solucionar os problemas observados (VIANNA, 2000).

A Macroavaliação pode ser compreendida como processo avaliativo e instrumento norteador educacional para sistemas educacionais. A avaliação desempenha um importante papel de mudança social, visto que é um instrumento de caráter político (CRONBACH, 1982). Para Vianna (2000), a Macroavaliação visa mensurar o desempenho de grupos (de alunos, professores, administradores, técnicos, etc.), sistemas de ensino, materiais pedagógicos, instituições e programas educacionais.

Para Helena, Ribeiro e Polato (2014, p. 04),

As avaliações em larga escala, portanto, têm sido utilizadas como um instrumento de regulação do processo de ensino-aprendizagem, objetivando um diagnóstico do estado atual do nível de proficiência adquiridos pelos estudantes e, conseqüentemente, do trabalho pedagógico da escola, ensejando um diagnóstico das redes de ensino, ou seja, trata-se de uma avaliação de sistema.

A Macroavaliação possui em sua essência o ato de avaliar os sistemas de ensino, mostrando suas deficiências e possíveis avanços. As duas formas de se avaliar apresentadas anteriormente são fundamentais e de certa forma se complementam, pois, a Microavaliação identifica e procura solucionar problemas educacionais no ambiente escolar, por outro lado, a Macroavaliação é capaz identificar problemas em sistemas educacionais, pode também comparar sistemas e métodos de ensino (entre escolas, municípios e estados), contribuindo para a melhoria do currículo (VIANNA, 2000).

As escolas no Brasil historicamente surgiram para preparação das elites, isso trouxe consigo a avaliação seletiva como cultura dominante no ambiente escolar. Nas escolas era natural a reprovação de alunos, em especial aos alunos de baixa renda e quase sempre o problema era tratado como natural e inquestionável (GATTI, 2002).

Na década de 60, uma grande massa de alunos, com o intuito de ingressar no ensino superior enfrentou o vestibular. O tradicional exame de ingresso no ensino superior era disponível a poucos, para se conseguir uma vaga em um curso superior o aluno deveria ter acesso à escola, poder permanecer na escola, ter concluído o ensino médio e ser aprovado no rigoroso vestibular. O acesso a todos esses benefícios eram para poucos em termos proporcionais com relação à população brasileira. Os exames de vestibular ao longo da história educacional no Brasil são excludentes e seletivos e refletem a cultura criada no país há décadas (GATTI, 2002).

Nos anos 1990, houve uma série de acontecimentos que favoreceram o processo educacional no país. Políticas de combate à evasão, correção de fluxo escolar, permanência na escola, programas de aceleração de aprendizagem e a criação da nova Lei de Diretrizes e Bases da Educação (LDB, Lei 9.394/96) contribuíram com o novo modelo educacional brasileiro (OBINO, 2011).

O processo de globalização exigiu de vários países, assim como do Brasil, a criação de métodos avaliativos que contribuíssem com a criação de indicadores e de resultados sobre os processos educacionais. Nesta época, iniciou-se um processo mais rigoroso e sofisticado de criação de cursos com a finalidade de formar especialistas em avaliação educacional.

A cultura de classificação em torno da avaliação é o que predomina historicamente no Brasil. Esse fato é percebido claramente quando nos deparamos com os resultados de avaliações educacionais de larga escala, as quais são submetidas a um grande número de alunos a fim de verificar, principalmente, os sistemas de ensino e os diversos grupos que de alguma forma contribuem para seu funcionamento, como os profissionais de educação, dentre os quais encontram-se os professores.

A atual expansão dos sistemas de ensino, unida com a ideia de democratização na educação, favorece o desenvolvimento educacional no país. Em contrapartida, a complexidade desse processo faz com que apenas ampliar a quantidade de escolas, professores e agentes educacionais não contribua de modo significativo para a melhoria da tão sonhada qualidade do ensino. Acreditamos que somente a avaliação, que é indicadora dos problemas educacionais a serem resolvidos, não trará melhorias à educação (VIANNA, 2003).

Atualmente, o poder público utiliza os resultados das avaliações de larga escala para fornecer indicadores aos sistemas de ensino a fim de melhorar o processo ensino-aprendizagem. Os resultados são baseados em objetivos pré-determinados e os questionamentos sobre os resultados são diversos, por exemplo: os objetivos das avaliações de larga escala estariam alinhados aos objetivos da escola? O núcleo gestor possui capacidade para manipular os dados para identificar problemas e construir ações pedagógicas corretivas? As avaliações de larga escala são construídas sobre a ótica democrática? Há equidade entre as provas? Os atuais modelos de correção das provas são os ideais para todos os sistemas de ensino do país?

O Sistema de Avaliação da Educação Básica (SAEB) teve sua primeira aplicação em 1990 e serviu como marco das aplicações de avaliação em larga escala no país (BECKER, 2010). No Brasil, houve um crescimento da quantidade de avaliações de larga escala e sua aplicação já abrange todos os níveis educacionais. No país, os exames mais conhecidos são: o SAEB, a Avaliação Nacional da Alfabetização (ANA); a Avaliação Nacional do Rendimento Escolar (ANRESC), mais conhecido como Prova Brasil; o Exame Nacional do Ensino Médio (Enem) e o Exame Nacional de Desempenho (ENADE).

A área educacional cresce constantemente, vários são os desafios a serem superados, como a falta de segurança nas escolas, a falta de insumos educacionais, os baixos salários dos professores da educação básica, a falta de equidade educacional, dentre outros.

O ambiente social em que os estudantes estão inseridos, em geral, influencia diretamente o seu comportamento, a dedicação, os sonhos, o interesse pelo estudo e outros aspectos de sua vida. O poder público oferece o ensino público gratuito a uma massa de estudantes que, devido ao contexto social onde estão inseridos, podem não conseguir absorver a maior parte do ensino proposto. Essa deficiência em conhecimentos pode ser percebida em exames de larga escala, em que a taxa de alunos com rendimento adequado é maior para os alunos de escolas particulares quando comparados aos de escolas públicas.

A equidade educacional se destaca dentre os problemas em educação e pode ser compreendida como um processo capaz de oferecer iguais oportunidades educacionais a diferentes sujeitos submetidos a um mesmo sistema de ensino.

Para Evaristo (2017, p. 04),

O conceito de justiça de Aristóteles (384-322 A.C), que, na obra *Ética a Nicômaco*, Livro V, afirmou que devemos “tratar igualmente os iguais e desigualmente os desiguais, na medida de sua desigualdade.”. Assim, o art. 5º, caput, da Constituição Federal é o dispositivo legal que ampara as ações afirmativas. Do mesmo modo, os incisos III e IV do art. 3º da Carta Magna, trazem os direitos sociais como expressões diretas do referido princípio, com destaque de que não se trata apenas da exigência de aplicação da lei pelos órgãos estatais e sim da efetividade do Princípio da Igualdade, caracterizado de forma genérica como direito à igualdade não só material, mas também substancial.

O ensino médio de escolas públicas, atualmente, é qualificado pela mídia e pelas secretarias de educação tendo como base a quantidade de alunos que conseguem alcançar boas notas em exames de larga escala e vagas em universidades públicas. A lógica desse tipo de análise é a de que quanto maior o número de alunos com boas notas nos exames melhor será a escola.

As escolas com taxas de rendimento acima das demais, geralmente, recebem mais recursos financeiros, culturais e educacionais, fazendo com que os discentes tenham maiores oportunidades de aprendizado que os estudantes de outras escolas. Esse fato é contraditório, pois acreditamos que se deve oferecer maiores oportunidades educacionais aos que mais necessitam.

[...médias de desempenhos, sem consideração às dificuldades de cada aluno, à relação entre desigualdade escolar e situações socioeconômicas familiares, devidas a territórios distintos ou a quaisquer condições de origem, não são resultados adequados. (RIBEIRO, 2013, p. 71)

Os candidatos cotistas (geralmente de escola pública) possuem médias inferiores aos candidatos de ampla concorrência no Enem. Apesar da diminuição dessa diferença ao longo dos anos, em muitos cursos concorridos a discrepância ainda é significativa. Se por um lado o Enem vem conseguindo diminuir as diferenças entre as duas classes de concorrentes por vaga, por outro ainda carece de métodos mais justos para estimar a proficiência dos candidatos.

### 3 TEORIAS PSICOMÉTRICAS

Medir e/ou avaliar é um ato que está bastante presente no cotidiano do ser humano. Medir é observar, registrar e avaliar qualidades ou quantidades de algum objeto (MARÔCO, 2009). Em diversas áreas do conhecimento humano, há métodos específicos para mensurar objetos, elementos da natureza e habilidades, os quais devem seguir um determinado padrão pré-estabelecido, confirmando a precisão e validade de suas medidas.

Um bom instrumento de medida deve ser invariante ao objeto mensurado, por exemplo, uma escala métrica deve mostrar a mesma precisão e validade ao medir o comprimento de uma mesa, altura de uma pessoa ou o tamanho de um livro. Para Thurstone (1928), um bom instrumento de medida não deve ser afetado pelo objeto mensurado.

Na área educacional, há teorias e métodos estudados e aplicados para avaliar estudantes, metodologias, programas educacionais e sistemas de ensino, que têm a Psicometria como a ciência responsável pela medição das variáveis latentes. Tais variáveis são habilidades que não podem ser observadas de forma direta (MARÔCO, 2009) e a Psicometria, devido ao seu conjunto de teorias e métodos matemáticos, é responsável por avaliá-las (MUÑIZ, 1996).

Para Pasquali (2003, p. 11),

A Psicometria representa a teoria e a técnica de medida dos processos mentais, especialmente aplicada na área da Psicologia e da Educação. Ela se fundamenta na teoria da medida em ciências em geral, ou seja, do método quantitativo que tem, como principal característica e vantagem, o fato de representar o conhecimento da natureza com maior precisão do que a utilização da linguagem comum para descrever a observação dos fenômenos naturais.

A Psicometria surgiu com o objetivo de quantificar aptidões humanas a fim de exercer determinadas tarefas (PASQUALI, 2003). Há duas principais teorias inseridas na Psicometria que são a Teoria Clássica dos Testes (TCT) e a Teoria de Resposta ao Item (TRI).

A TCT tem foco no resultado final, ou seja, ela tenta explicar o significado do escore obtido pelo respondente. Em contrapartida, a TRI tem foco em cada item do teste, verificando a probabilidade de que cada item tem de ser acertado ou errado pelo respondente (PASQUALI, 2003).

### 3.1 A Teoria Clássica dos Testes

Inicialmente, a Teoria Clássica dos testes (TCT) foi desenvolvida por psicólogos com orientação estatística, os quais eram influenciados pela visão materialista da época (início do século XIX) e buscavam analisar dados obtidos exclusivamente para aferir a inteligência humana (PASQUALI, 2003).

Os cientistas psicométricos buscavam compreender o significado do desempenho de um indivíduo (soma dos resultados de todas as tarefas) em uma série de tarefas onde atribuíam o valor 1 a cada acerto e zero a cada erro. Pasquali (2003) afirma que a TCT tinha foco no escore e não na habilidade. O pensamento científico mensurava que o desempenho do candidato representava aproximadamente a magnitude de sua habilidade, em outras palavras, era a habilidade que o teste queria aferir no candidato. Todo instrumento de medida contém erros, logo, o desempenho (ou escore) não pode representar puramente a habilidade de um sujeito em um teste (PASQUALI, 2003).

#### 3.1.1 Os postulados da TCT

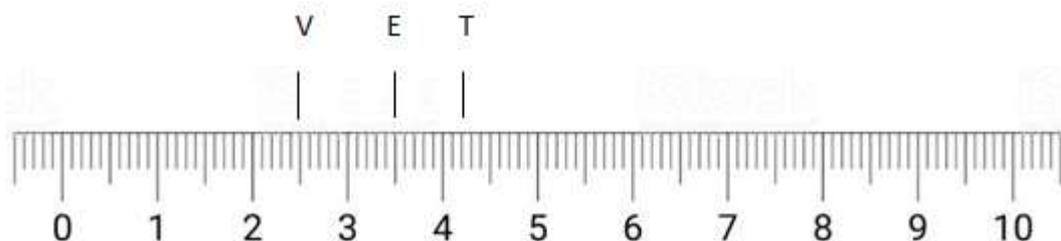
Os postulados básicos da TCT foram desenvolvidos por Spearman (1904, 1907 e 1913, p. 60 apud PASQUALI, 2003, p.15).

D)  $T=V+E$ ;

T= escore medido no teste; V= escore verdadeiro; E= erro de medida associado ao teste.

A figura abaixo mostra que o escore verdadeiro é diferente do escore medido pelo teste pois o erro de medição sempre existe.

Figura 1 - Componentes do escore



Fonte: Autor (2019)

O desafio é controlar os erros, esses podem ser causados por vários fatores do próprio teste, viés do candidato, fatores ambientais (ou históricos), socioeconômicos e outros (VIANNA, 2000). O escore verdadeiro do indivíduo é impossível de ser calculado pela TCT, pois não há como calcular o erro cometido por cada indivíduo.

Um procedimento muito utilizado para minimizar os erros da TCT é aplicar o teste a um número muito grande de indivíduos, e após sua aplicação, é necessário utilizar estatísticas apropriadas (distribuição de frequências de T, V e E) para reduzir os erros de medição. Ao considerar o erro como evento casual, podemos aplicar a esperança estatística ( $E'$ ) na equação  $T=V+E$ , daí teremos  $E'(T)=E'(V)+E'(E)$ . Sabe-se que  $E'(E)=0$ , logo  $E'(T)=E'(V)$ , como V é constante teremos  $E'(T)=V$ .

II) Não há correlação entre V e E. ( $r_{V,E}=0$ )

$r$ = coeficiente de correlação.

“Não há razão para afirmar que escores verdadeiros maiores terão maiores erros e que escores verdadeiros menores terão menores erros” (PASQUALI, 2003, p. 68)

III) Os erros cometidos em um teste não estão correlacionados com erros em testes paralelos (Testes que estão medindo a mesma habilidade), os erros são aleatórios.

### 3.1.2 A dificuldade do item

Os itens pela TCT podem ser classificados de acordo com seu nível de dificuldade ( $d$ ). Se em um mesmo teste para dois itens diferentes tivermos  $d_1 < d_2$ , então houve uma maior quantidade de acertos no item 2 quando comparado ao item 1, assim o item 2 é considerado menos difícil que o item 1. Se o valor de  $d$  se aproxima de zero então o item será mais difícil, e se o valor se distancia de zero, o item será mais fácil. É válido lembrar que o valor de  $d$  não pode ser maior que 1 pois neste trabalho os dados utilizados são Dicotômicos (cada acerto vale 1 e cada erro vale 0(zero)). A fórmula para se calcular o valor da dificuldade do  $i$ -ésimo item é dada pela equação abaixo.

$$d_i = \frac{p}{N}, \text{ onde:}$$

Equação 1

$p$ =quantidade de indivíduos que responderam corretamente ao item  $i$ ;

$N$ =quantidade de pessoas que responderam ao item  $i$ ;

$d_i$ =dificuldade do item  $i$ .

O índice de dificuldade permite que sejam construídos testes heterogêneos, os itens que irão compor o teste podem ter diferentes valores de dificuldade abrangendo diferentes níveis de dificuldade dentro do teste (CERDÁ, 1978). Baseando-se na equação definida anteriormente, cinco classificações por intervalos numéricos com seus respectivos percentuais foram definidos (CERDÁ, 1978).

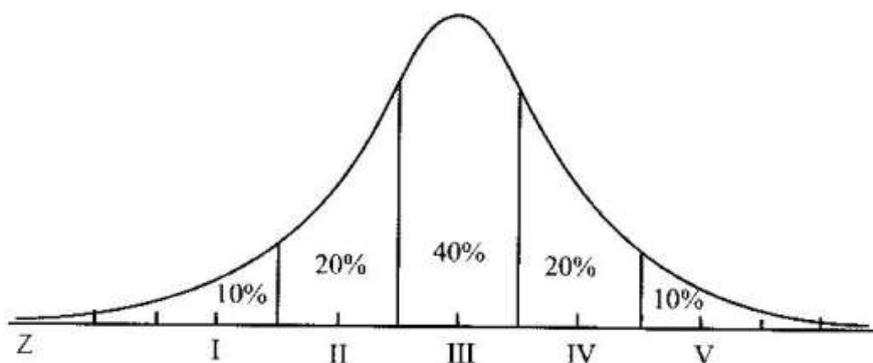
Tabela 1 - Classificação da dificuldade

INTERVALO d	PERCENTUAL	CLASSIFICAÇÃO
[0,75; 0,95]	10%	MUITO FÁCIL
[0,55; 0,75[	20%	FÁCIL
[0,45; 0,55[	40%	NORMAL(MODERADO)
[0,25; 0,45[	20%	DIFÍCIL
[0,05; 0,25[	10%	MUITO DIFÍCIL

Fonte: Cerdá (1978)

Os valores dispostos na tabela acima forma propostos por Cerdá (1978), uma vez que representava uma aproximação para distribuição normal para um teste. Pasquali (2003) afirma que um bom teste deve conter itens com distribuição em níveis de dificuldade (I, II, III, IV e V) seguindo uma forma próxima da curva normal.

Figura 2 - Distribuição da dificuldade por níveis



Fonte: Pasquali (2003)

### 3.1.3 O índice de discriminação

O índice de discriminação (Id) é uma estatística capaz de diferenciar candidatos (ou participantes) que obtiveram alto ou baixo escore em um teste. Se um grupo de candidatos com baixo escore consegue responder corretamente a um determinado item, então é natural supor que o mesmo item também seja respondido corretamente por um grupo de candidatos de alto escore. Para 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”.

Uma classificação do Id foi proposta por Ebel (1954), na qual ele estabelecia critérios para que um item fosse incluído, revisado ou eliminado de um teste. Em sua classificação Ebel (1954) sugere que apenas itens aceitáveis ou satisfatórios devam estar no teste e os itens revisados podem ser incluídos após serem ajustados.

Tabela 2 - Classificação do índice de discriminação

INTERVALO Id	CLASSIFICAÇÃO DO ITEM
]0; 0,2[	INEFICIENTE
[0,2; 0,3[	REVISÃO
[0,3; 0,4[	ACEITÁVEL
Id ≥0,4	SATISFATÓRIO

Fonte: Ebel (1954)

Pasquali (2017) mostra duas formas de cálculo da discriminação na TCT: a Correlação Bisserial e a Correlação Bisserial por pontos.

#### I) Correlação Bisserial

Deve ser utilizada quando as variáveis correlacionadas são contínuas (uma das variáveis deve ser dicotômica). A equação do coeficiente é definida por:

$$r_{bis} = \frac{\bar{X}_p - \bar{X}_t \cdot pq}{S_t \cdot y}$$

Equação 2

$\bar{X}_p$  = média dos participantes que acertaram o item;

$\bar{X}_t$  = média total do teste;

$S_t$  = desvio padrão populacional do teste;

P = proporção dos candidatos que responderam corretamente ao item;

y = ordenada na curva normal no ponto de divisão dos segmentos que contém as proporções p e 1-p. Se surgir valores maiores que um (distribuição não normal), recomenda-se a Correlação Bisserial por pontos.

II) Correlação Bisserial por pontos.

$$r_{\text{bisp}} = \frac{\bar{x}_p - \bar{x}_t}{S_t} \cdot \sqrt{\frac{p}{q}}$$

Equação 3

$r_{\text{bisp}}$  = coeficiente de Correlação Bisserial por pontos.

Outras formas para o cálculo de Id podem ser vistas em Pasquali (2017).

### 3.1.4 A curva de proporção de acertos do item

A curva de proporção do item é uma ferramenta clássica para avaliar a proporção de candidatos com habilidade no máximo igual a  $\theta$ . A curva é construída com base na proporção de acertos para cada score, que de forma acumulativa para um determinado item irá gerar pontos para a construção do gráfico.

Em um teste, é solucionado um item qualquer para obter os dados de construção da curva, todos os candidatos que responderam ao item são separados com base em seus scores alcançados no teste e, para cada score, é calculado a proporção de acertos. Os pontos no gráfico (habilidade x proporção de acertos) são calculados da seguinte maneira: a proporção para o menor score observado é plotado diretamente (gráfico: score x proporção de acertos), sempre que o score vai aumentando, a proporção de acertos do score calculado inicialmente será adicionada do valor calculado para o ponto anterior, e assim um novo ponto será plotado na figura, assim, a curva construída será monotonicamente crescente. Como exemplo, temos uma amostra de 500 candidatos selecionados de forma aleatória, esses responderam ao item 45 (correspondente a questão 180) do caderno AZUL para a prova de Matemática.

Tabela 3 - Proporção de acertos ao item 45 (prova azul)

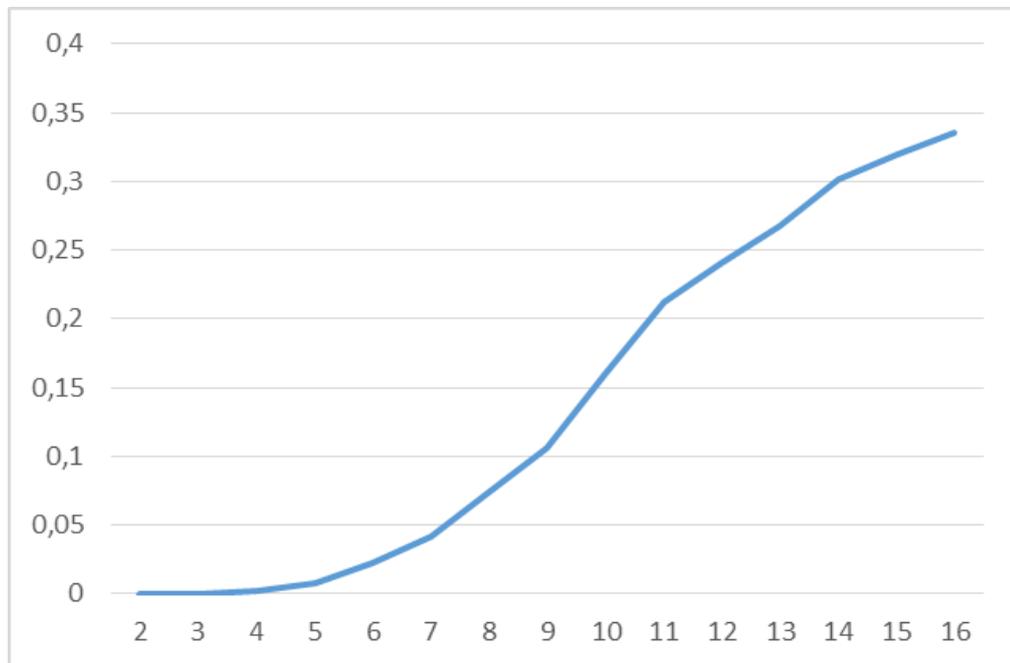
ESCORES															
	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
PA	0	0	0	0,01	0,01	0,02	0,03	0,03	0,05	0,05	0,03	0,03	0,03	0,02	0,02
PAC	0	0	0	0,01	0,02	0,04	0,07	0,11	0,16	0,21	0,24	0,27	0,3	0,32	0,34

Fonte: Autor (2019)

PA-proporção de acertos

PAC-proporção de acertos acumulado

Figura 3 - Curva de proporção de acertos acumulado por nível de escore  
(Item 45 da prova azul)



Fonte: Autor (2019)

Essa ideia de curva foi trabalhada por Tucker (1946) e por Lord (1952), os quais construíram uma nova curva baseada em modelos logísticos dicotômicos, que serviria como base para a TRI (PASQUALI, 2003). Esse novo modelo de curva não precisa ser necessariamente em forma de “S”, como geralmente ocorria na TCT, visto que ela pode assumir outras formas ao longo da habilidade medida.

### 3.2 A Teoria de Resposta ao Item

Os modelos de Teoria de resposta ao Item (TRI) surgiram na década de 30 com o trabalho de Richardson (1936), sendo sucedido pelos trabalhos de Lawley (1943), Tucker (1946) e Lord (1952) (COUTO, PRIMI, 2011). Os primeiros modelos de TRI consideravam que apenas uma única habilidade de apenas um grupo estava sendo medidos pelo teste. Os itens eram corrigidos como certo ou errado, isto é, de forma dicotômica.

Os primeiros modelos de TRI eram baseados na função de ogiva normal que, posteriormente, foi substituída por modelos de função logística (ANDRADE; TAVARES & VALLE, 2000). O modelo unidimensional de 2 parâmetros foi desenvolvido por Lord (1952), baseando-se na função de ogiva normal. Lord percebeu que o modelo de 2 parâmetros não era capaz de fornecer informações relevantes sobre o acerto casual. Então, ele propôs um novo modelo que avaliasse o acerto ao acaso e, assim, surgiu o modelo de 3 parâmetros.

Os modelos de Lord eram baseados na função de ogiva normal, mas foram substituídos por modelos logísticos propostos por Birnbaum (1968). Os modelos logísticos são mais convenientes pelo fato de não envolver integração e dependerem somente dos parâmetros do item e da habilidade (ANDRADE; TAVARES & VALLE, 2000).

Em 1960, Rasch propôs o modelo unidimensional de 1 parâmetro baseado na função de ogiva normal, em 1968, Wright descreveu o mesmo modelo baseando-se na função logística (MUÑIZ, 1990).

Os estudos sobre a TRI continuaram, grandes contribuições e novas abordagens foram surgindo como o modelo de resposta gradual de Samegima (ANDRADE; TAVARES & VALLE, 2000). O modelo proposto por Samegima buscava obter maior quantidade de informação sobre as respostas dos participantes quando comparado aos modelos que consideravam somente acerto e erro.

Muitos cientistas voltaram seus olhares para a TRI buscando contribuir com a melhoria da avaliação das variáveis latentes. Bock (1972), Andrich (1978), Masters (1982) e Muraki (1992) construíram modelos para mais de duas categorias de resposta e com diferentes estruturas entre categorias (ANDRADE; TAVARES & VALLE, 2000). Em 1997, surgiram modelos de 1, 2 e de 3 parâmetros logísticos para duas ou mais populações. Uma vantagem desses modelos, é o de poder comparar o rendimento entre populações distintas, que foram submetidas a testes com alguns itens em comum.

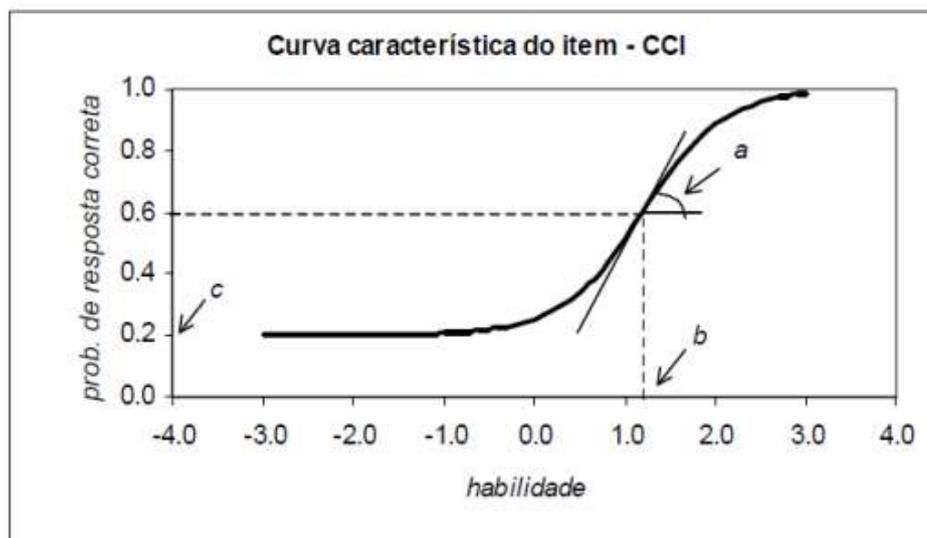
A estimativa dos parâmetros nos modelos de TRI é um processo sofisticado, que necessita de recursos computacionais (via estimativa dos parâmetros dos itens) e pode ser

considerado como a parte mais complicada na TRI. Com a evolução do poder estatístico e computacional no âmbito da TRI, é possível estimar as habilidades de todos os respondentes, que participaram efetivamente do teste, inclusive os que erraram ou acertaram todos os itens (ANDRADE; TAVARES & VALLE, 2000). Atualmente, no Brasil, a TRI vem sendo amplamente aplicada no campo educacional, como no SAEB (Sistema de avaliação da educação básica) desde 1995, porém o uso da TRI é amplamente conhecido quando se fala de Enem.

### 3.2.1- A curva característica do item (CCI)

A proporção das respostas corretas no item  $i$ , dadas no conjunto de indivíduos com habilidade  $\theta_j$ , representa a probabilidade  $P(U_{ij}=1|\theta_j)$  (ANDRIOLA, 1998). Esse modelo acrescenta que indivíduos de maior habilidade tem maior probabilidade de acerto ao item, podendo, essa relação, ser não linear (ANDRADE; TAVARES & VALLE, 2000). A CCI é uma curva monotônica que geralmente tem forma de “S”.

Figura 4-CCI para o modelo 3PL



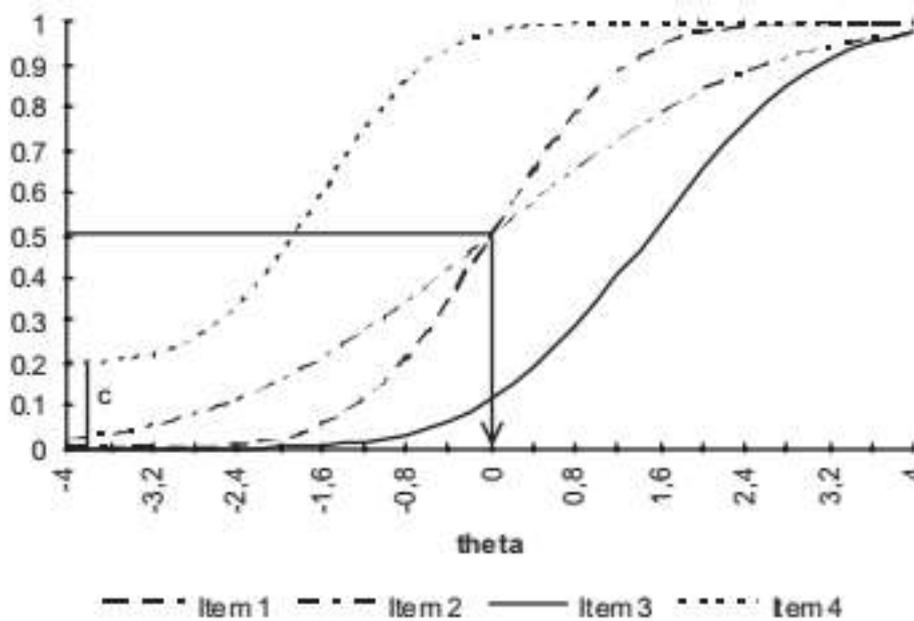
Fonte: Gleiber, Primi (2011)

“No Enem, esta relação monotônica é uma função logística de 3 parâmetros desenvolvida por Birbaum em 1968” (INEP, 2011, p. 03). O modelo 3PL assegura que, entre os indivíduos avaliados, os de maior habilidade possuem maior probabilidade de responder corretamente um item qualquer e esse modelo é cumulativo, ou seja, se um candidato de

habilidade  $\theta_j$  consegue responder corretamente a um item com uma dificuldade  $x$ , então ele tem alta probabilidade de responder corretamente a todo item com dificuldade menor ou igual a  $x$ , caso a resposta ao item não tenha sido um acerto ao acaso.

Por meio da visualização da CCI (Figura 04), é possível inferir sobre o comportamento do item. Um modo simples de compreender o comportamento dos parâmetros é verificando o ponto de inflexão, a taxa de crescimento no ponto de inflexão e o acerto ao acaso.

Figura 5- CCI de 4 itens diferentes



Fonte: Gleiber, Primi (2011)

A CCI deve ser interpretada da esquerda para a direita, pois assim cresce a dificuldade do item dentro da escala de habilidade. De acordo com a figura 02, o item mais fácil é o item 4, enquanto o mais difícil é o item 3. A inclinação está relacionada com a discriminação do item, pois quanto mais inclinado (com relação ao eixo das abcissas) maior será a discriminação, que é o parâmetro responsável por separar candidatos de alta habilidade de candidatos de baixa habilidade.

Na figura 02, o item 1 é o menos discriminativo e o item 2, o mais discriminativo. O acerto ao acaso (como no gráfico) pode ser compreendido como a probabilidade de acerto ao item para candidatos com a menor habilidade na escala utilizada. Como exemplo, podemos visualizar no gráfico acima, os itens 1 e 2, que possuem baixa probabilidade de acerto ao item para candidatos com habilidade ( $\theta$ ) menor ou igual a zero. Para o item 3, a probabilidade

de acerto ao chute é baixa para candidatos com habilidade menor que 1,6 e alta para maiores valores de habilidade.

Os parâmetros dos itens na TRI, geralmente, seguem uma nomenclatura pré-definida. O parâmetro discriminação é representado pela letra  $a$ , o parâmetro dificuldade é representado por  $b$  e o acerto ao acaso é representado pela letra  $c$  (ANDRADE; TAVARES & VALLE, 2000).

“Existem vários modelos de TRI, dependendo do tipo de função matemática adotada e dos parâmetros dos itens que se queira investigar” (COUTO, PRIMI, 2011, p. 05). Os modelos de TRI dependem, principalmente, da natureza do item (se são dicotômicos ou não), da quantidade de populações (uma ou mais de uma) e do número de variáveis latentes (ANDRADE; TAVARES & VALLE, 2000).

Os modelos de TRI representam a probabilidade de um candidato em um teste informar uma resposta correta a um item, dependendo apenas dos parâmetros do item e de sua habilidade. Dessa forma, é de se esperar que quanto maior for a habilidade do candidato maior será a probabilidade de acerto a um item.

Neste trabalho, chamaremos o modelo de 1 parâmetro logístico de 1PL, o de dois parâmetros logísticos de 2PL, o de três parâmetros logísticos de 3PL e o de quatro parâmetros logísticos de 4PL. Os modelos logísticos exigem que os dados sejam dicotômicos ou dicotomizados, isso quer dizer que, se a resposta a um determinado item foi a resposta correta, então o item receberá o valor 1 e receberá 0 (zero), caso contrário.

### 3.2.2 Os modelos logísticos

#### O modelo 4 PL

O modelo 4PL procura controlar erros aleatórios com relação a construção dos itens (MUÑIZ, 1990). Se houve problema na construção do item, é de se esperar que candidatos com grandes habilidades também respondam incorretamente ao item. A expressão do modelo é:

$$P_i(\theta) = c_i + (Y_i - c_i) * \frac{e^{Da_i(\theta - bi)}}{1 + e^{Da_i(\theta - bi)}}, \text{ onde:}$$

Equação 4

$P_i(\theta)$  = Probabilidade de acerto ao item  $i$  em função da habilidade  $\theta$ ;

$\theta$  = variável (habilidade) a ser medida pelo item;

$a_i$  = discriminação do item  $i$ ;

$b_i$  = dificuldade do item  $i$ ;

$c_i$  = acerto ao acaso;

$D$  = constante que em geral recebe valor 1, quando utilizada para aproximar aos valores da curva normal acumulada (ogiva normal) seu valor será 1,7;

$e \cong 2,72$  (base dos logaritmos neperianos);

$Y_i$  = erro de construção do item.

A discriminação do item ( $a$ ) é o parâmetro que consiste na medida de quanto o item consegue separar o grupo de candidatos com alto grau de habilidade do grupo de candidatos com baixo grau de habilidade. A discriminação corresponde o valor proporcional de inclinação da CCI no ponto  $b_i$  (ANDRADE; TAVARES & VALLE, 2000).

A dificuldade do item ( $b$ ) que representa a habilidade mínima para que o candidato tenha probabilidade de acerto igual a  $(1+c)/2$ , logo quanto menor o valor de  $b$  (podendo ser até mesmo negativo) mais fácil será o item, o contrário também é verdade (dificuldade negativa não implica em discriminação negativa).

O acerto ao acaso ( $c$ ) é o parâmetro que representa a probabilidade que indivíduos com baixa habilidade tem de responder corretamente o item (ANDRADE; TAVARES & VALLE, 2000).

$Y_i$  representa o erro ao construir os itens e os valores são próximos mas inferiores a 1. Ainda não há vantagem comprovada do modelo com relação ao modelo 3PL (GLEIBER, PRIMI, 2011).

### **O modelo 3 PL**

O modelo 3PL é um dos modelos de TRI mais utilizados na última década, sua aplicação é mais conhecida entre estudantes e pesquisadores brasileiros quando o assunto é o Enem, esse modelo é capaz de estimar a proficiência para todos os candidatos que resolveram a prova, inclusive aos que erraram todos os itens.

O modelo 3PL possui os mesmos parâmetros do modelo anterior apenas trocando-se o parâmetro de erro de construção do item ( $Y_i$ ) por 1.

A expressão do modelo 3PL é representada por:

$$P_i(\theta) = C_i + (1 - C_i) * \frac{e^{D a_i(\theta - b_i)}}{1 + e^{D a_i(\theta - b_i)}}$$

Equação 5

### O modelo 2 PL

O modelo é descrito por:

$$P_i(\theta) = \frac{e^{D a_i(\theta - b_i)}}{1 + e^{D a_i(\theta - b_i)}}$$

Equação 6

$a_i$ =discriminação do item  $i$ ;

$b_i$ =dificuldade do item  $i$ .

Para o modelo 2PL, o conceito de discriminação é o mesmo da TCT (COUTO, PRIMI, 2011), em outras palavras, é uma medida que visa diferenciar o grupo de maior habilidade do grupo de menor habilidade. “A dificuldade do item corresponde ao valor da habilidade em que a probabilidade de acerto ao item é de 50%” (COUTO, PRIMI, 2011).

### O modelo 1 PL

O modelo proposto por Rasch e adaptado por Wright propõe que a probabilidade de acerto ao item depende somente da dificuldade do item e da habilidade do candidato. “Conhecendo a dificuldade do item e a habilidade do sujeito, é possível prever qual é a probabilidade desse sujeito acertar o item” (GLEIBER, PRIMI, 2011, p. 05). Sua expressão matemática é representada por:

$$P_i(\theta) = \frac{e^{D(\theta - b_i)}}{1 + e^{D(\theta - b_i)}}, \text{ onde,}$$

Equação 7

A dificuldade ( $b$ ) deste modelo é a mesma utilizada no modelo 2PL.

## 4 O ENEM

O Enem é uma avaliação de larga escala que permite a comparação do rendimento dos alunos entre escolas, estados, regiões e serve como indicador da qualidade educacional no Brasil (AFONSO, 2005). O Enem foi criado em 1998 e pode ser compreendido como uma avaliação que mensura a capacidade do aluno ao fim do ensino médio com relação às competências fundamentais ao pleno exercício da cidadania.

A proposta inicial para o exame (Enem) era a de promover a interdisciplinaridade entre os conteúdos, valorizar o pensamento crítico, interpretar informações, solucionar problemas e inferir sobre a qualidade do ensino no Brasil. Em seus primeiros dez anos de criação, o exame continha 63 questões objetivas e uma redação. As provas objetivas ocorriam em cadernos de cores diferentes (AZUL, AMARELO, VERDE e ROSA). Os itens dispostos na prova eram iguais mas diferiam na posição do item de acordo com a cor do caderno.

O Enem utilizou de 1998 a 2008 a Teoria Clássica dos Testes (TCT) para definir a proficiência dos candidatos participantes do certame. A TCT abrange uma série de modelos matemáticos para classificar uma variável psicológica (PASQUALI, 2003). A TCT contém problemas dentre os quais se destacam: ter os parâmetros dos itens dependentes da amostra, não poder diferenciar candidatos que acertaram o mesmo número de itens e de poder ser fortemente afetado pelo objeto do estudo (THURSTONE, 1928).

Em 2009, o Enem passou por várias mudanças, por exemplo: o aumento da quantidade de itens objetivos, o aumento da quantidade de dias de aplicação das provas e, em especial, a mudança na teoria utilizada na correção dos gabaritos.

A TRI substituiu a TCT em 2009 por apresentar vantagens em relação a teoria outrora utilizada. Segundo o Inep (2011, p. 03), “A TRI não busca substituir a Teoria Clássica dos Testes (TCT), pelo contrário, é importante que se busque utilizar os avanços oferecidos em cada uma delas”.

“A TRI é considerada a forma mais avançada de se mensurar um traço-latente, no caso, o conhecimento” (INEP, 2011, p. 05). A TRI é composta por um conjunto de modelos matemáticos que avalia o traço latente (habilidade) do candidato utilizando-se dos parâmetros dos itens estimados através de recursos computacionais. Cronbach (1963) ressaltou a importância de se estudar os testes de medida em processos complexos dando grande importância ao estudo dos itens. Para Andrade; Tavares & Valle (2000), a TRI compreende um conjunto de modelos matemáticos, que representam a probabilidade de um indivíduo informar uma resposta correta a um item como função dos parâmetros do item e da habilidade

(ou habilidades) do respondente.

A construção dos cadernos de prova do Enem passa por um processo sofisticado, que exige o intermédio de um grande número de técnicos responsáveis por sua criação, seleção, estimação dos parâmetros dos itens, logística e outros (INEP, 2011). Para que seja possível comparar resultados, o modelo utilizado no exame educacional deve garantir que os itens apresentem as mesmas características (mesmos parâmetros) para os diversos grupos avaliados (SOARES; GENOVEZ & GALVÃO, 2005). A TRI afirma que os parâmetros dos itens não dependem da habilidade dos respondentes, ou seja, não variam para diferentes grupos (BAKER, 2001).

“[...] talvez o aspecto mais importante da nova teoria é a promessa de fornecer medidas invariantes do desempenho cognitivo, que não dependem dos itens que compõem a prova ou das pessoas investigadas na amostra”

(FLETCHER, 1994 Apud ANDRIOLA, 1998, p. 10)

Para a Organização das Nações Unidas (ONU) (2010, p.03):

Em relação ao TRI, que uma de suas vantagens metodológicas é a possibilidade da elaboração de provas diferentes para o mesmo exame, que podem ser aplicadas em qualquer período do ano, com o mesmo grau de dificuldade, permitindo assim sua comparabilidade no tempo.

Os modelos de TRI dependem da natureza do item, das populações estudadas e da quantidade de traços latentes estimados (ANDRADE; TAVARES & VALLE, 2000). O Enem utiliza o modelo de três parâmetros logísticos (3PL), que avalia apenas um traço latente, e esse modelo exige a unidimensionalidade do teste. Apesar de haver muitas habilidades que influenciam nas respostas em um teste, a TRI considera que há apenas uma habilidade determinante no conjunto de respostas dadas aos itens em que esta será medida pelo teste. A unidimensionalidade exigida na TRI, é contrária aos pensamentos de Cronbach (1963) e Vianna (2000), eles afirmam que o ser humano é multidimensional e não pode ter suas habilidades representadas por uma única variável. A unidimensionalidade implica em independência local.

A independência local afirma que, mantidas constantes as habilidades que afetam o exame, a resposta dada a um item não influencia na resposta apontada a outro item, isso quer dizer que as respostas são independentes do ponto de vista estatístico. A independência local

admite que as respostas apontadas aos itens são independentes ou não correlacionadas. Por outro lado, Pasquali (2003) afirma que, se houver correlação entre as respostas, esse fato não teria origem com a habilidade medida no teste, mas sim sobre outra(s) habilidade(s). Se os fatores causadores de correlação forem controlados, teremos apenas uma habilidade dominante, caracterizando a unidimensionalidade exigida pela TRI (LORD, 1980).

Os pressupostos da TRI são intrigantes por fazer afirmações que podem ser questionadas por meio de argumentos simples. Vários estudos comprovam que, em exames de larga escala, a posição do item, assim como a troca de um item por outro semelhante, pode levar a resultados completamente diferentes (CARLOS, 2004). Andriola (2002) afirma que alguns itens podem ter comportamento diferente para diferentes grupos, fazendo com que a validade de um teste seja ao mínimo questionada. Alguns itens podem apresentar um viés por conta de sua localização no teste, assim, a resposta indicada a um item pode ter sido influenciada pela resposta apresentada a outro item, contrariando o pressuposto da independência local (KENNETH, 2016).

Os exames de larga escala podem fornecer informações divergentes das informações reais para diferentes grupos participantes. Nessa perspectiva, algumas perguntas são necessárias: a TRI realmente traz validade e fidedignidade ao Enem? A posição do item pode alterar a dificuldade da prova? Há correlação entre os itens? O Enem 2017 é de fato unidimensional? Será que realmente existe equidade nas provas?

Apesar das vantagens apontadas pela TRI frente a TCT, é necessário a continuidade de seus estudos, pois a TRI possui pontos questionáveis. O Brasil por ser um país com dimensões continentais, multicultural e com diferenças econômicas entre regiões, pode apresentar resultados distintos para os diferentes grupos que participam do Enem.

#### **4.1 O Enem 2017**

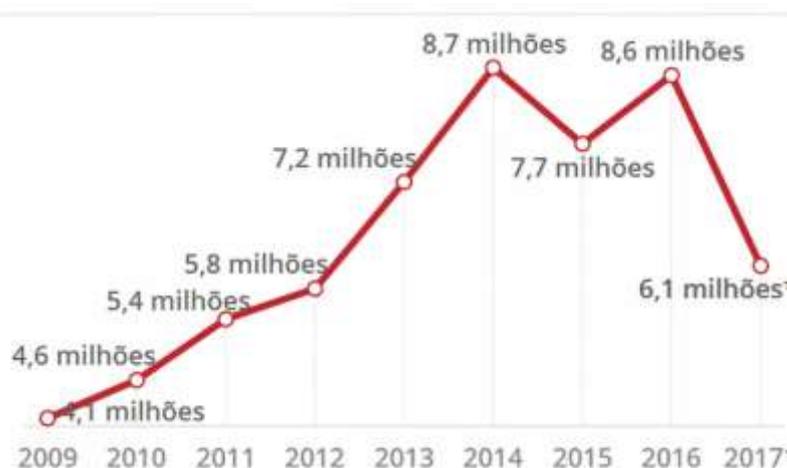
O sonho de conseguir uma vaga em uma universidade pública, bolsa de estudos em instituições de ensino superior privada ou mesmo financiamento de curso universitário pode ser concretizado por um bom desempenho no Enem. Para conseguir a tão sonhada vaga, é preciso muito estudo, bom controle emocional, estar bem preparado fisicamente, assim como um pouco de sorte.

Os candidatos aprovados no Enem podem se candidatar a vagas por todo país por meio do Sisu (Sistema informatizado do Ministério da Educação, que oferta vagas a candidatos participantes do Enem em instituições de ensino superior pública). Esse modelo de seleção

permite que o candidato possa ver sua classificação diariamente, possibilitando a avaliação de suas reais chances de conseguir a vaga em um determinado curso escolhido previamente.

O Enem foi ganhando credibilidade e a quantidade de inscritos foi crescendo, bem como a quantidade de universidades que passaram a adotar o exame como forma de ingresso ao ensino superior. Ao longo dos anos, o número de inscritos cresceu e somente nos exames dos anos anteriores a 2015 e 2017 é que houve uma menor quantidade de inscritos.

Figura 6 - Histórico de inscritos confirmados no enem 2017



Fonte: Inep (2017)

Há vários fatores que influenciaram o decréscimo da quantidade de inscritos em 2017, dentre as quais destacamos a mudança sobre os critérios de isenção, o fato da nota mínima de aprovação no Enem não poder ser utilizada como conclusão de ensino médio e o valor cobrado pelo exame.

Em 2017 ocorreram aproximadamente 2,3 milhões de ausências por motivos diversos implicando em um prejuízo enorme aos cofres públicos (G1, 2017), como o fato era recorrente então uma nova medida foi tomada, a isenção não seria mais concedida a todos os solicitantes. “O objetivo é possibilitar uma aplicação consciente do recurso público, garantindo a isenção aos que realmente necessitam, além de criar uma cultura mais ética do uso desse benefício” (INEP, 2017, p. 01).

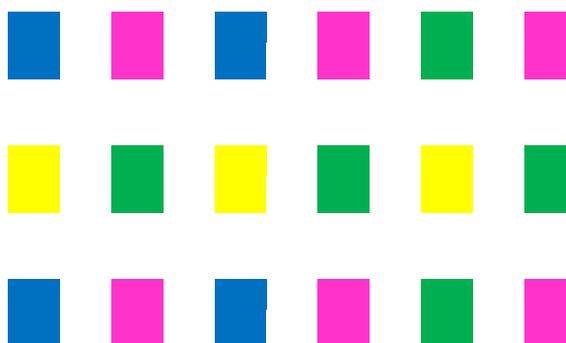
A quantidade de faltosos no Enem 2017 foi de aproximadamente 2,2 milhões e os candidatos presentes representavam praticamente o dobro da quantidade de faltosos (4,4 milhões). O número aproximado de 6,6 milhões não está corretamente representado na figura 01 pelo fato do gráfico ter sido gerado antes do término do segundo processo de isenções.

O Enem ocorrido no ano de 2017 foi atípico, visto que houve falta de energia elétrica

em três municípios (Teresina, no Piauí; Uruaçu, em Goiás; e Olinda, em Pernambuco), prejudicando 3500 candidatos participantes do primeiro dia de prova (redação, ciências humanas e linguagens). Assim, uma nova data foi marcada e uma nova prova foi aplicada no mês de dezembro (12 /12 /2017). Segundo a ONU (2010) e o Inep (2011), não houve prejuízos aos candidatos que participaram da segunda aplicação em 2017, pois, a TRI permite a construção de cadernos diferentes com o mesmo grau de dificuldade.

Os cadernos do Enem são distribuídos de tal forma que os candidatos mais próximos recebam cadernos de cores diferentes. Logo, a forma de distribuição dos cadernos tende a diminuir possíveis fraudes e a famosa “cola” na aplicação.

Figura 7 - Possível distribuição dos cadernos do Enem por cor



Fonte: Autor (2019)

#### 4.2 A construção do item

O Enem é elaborado e aplicado pelo Inep (Instituto nacional de estudos e pesquisa educacional) e, como avaliação de larga escala, permite que se faça juízos de valor e se proponha melhorias no processo ensino-aprendizagem em um ambiente muito maior que em uma escola.

Para o Inep (2010, p.05),

“As avaliações em larga escala distinguem-se das avaliações internas, na medida em que estas são feitas pelo professor ou pela própria instituição de ensino para fazer julgamentos de valor e propor alternativas no âmbito da sala de aula ou da instituição, enquanto as avaliações em larga escala são elaboradas por um órgão externo às escolas, com a finalidade de fazer juízos de valor e propor alternativas em âmbito mais amplo que o da instituição de ensino.”

Por meio das análises dos resultados obtidos pelos testes educacionais (como o Enem), é possível estabelecer juízos de valor sobre os participantes. O principal objetivo é identificar as maiores dificuldades apresentadas pelos educandos e inferir sobre o atual modelo educacional.

Para que os testes sejam exequíveis e válidos, o Inep criou o BNI (Banco Nacional de Itens). O BNI compõe uma extensa quantidade de itens das diversas áreas do conhecimento com qualidade comprovada (INEP, 2010).

Segundo o Inep, (2010, p.05)

O BNI define-se, portanto, como uma coleção de itens de testes de natureza específica – organizada segundo determinados critérios – disponíveis para a construção de instrumentos de avaliação. A manutenção do BNI depende da entrada constante de itens de qualidade. Para tanto, educadores e pesquisadores da educação brasileira são chamados a colaborar nessa construção, elaborando itens que possam fazer parte desse banco. Em atividades dessa natureza, a experiência docente é de fundamental importância para que se possam elaborar itens em consonância com o contexto educacional.

O Inep criou um guia de elaboração e revisão de itens para avaliações estruturado por definições, item de múltipla escolha, etapas de elaboração do item, especificidade de apresentação do item, etapas de validação e protocolo de revisão.

Os itens são construídos com base na Matriz de Referência construída pelo Inep, a qual se estrutura por meio de competências e habilidades esperadas dos candidatos do teste em uma etapa da educação básica. O currículo é muito mais amplo que a Matriz de Referência e os dois não devem ser confundidos. A matriz é importante porque garante transparência ao processo avaliativo, orienta a preparação para o processo e fornece subsídios confiáveis para validação dos resultados.

A competência é a capacidade de organizar ideias diversas, estabelecendo relações entre sujeito e objeto, visando resolver ou enfrentar situações (problemas) propostas. Nos itens do Enem, é muito comum verificar o desafio feito a um candidato para mobilizar muitos recursos cognitivos em um texto-problema para que o mesmo tome decisões que contribuam de forma simples para uma solução do problema. O conceito de habilidade é uma consequência da competência, que pode ser a rápida resolução e de forma simples um determinado problema proposto. As habilidades decorrem das competências adquiridas e se referem ao plano imediato do “saber fazer” (INEP, 2005, p. 17)

O item é a unidade básica dos testes, é a célula do instrumento fornecedor de dados. O nome popular para item é questão em avaliações educacionais. Nenhum instrumento de medida é perfeito (VIANNA, 1982), porém sua utilidade é fundamental para medir o desempenho acadêmico (INEP, 2010).

No Enem, os itens são objetivos (de resposta orientada) e cada item possui cinco (5) opções de resposta. Dessa maneira, o candidato tem a opção de verificar se sua resposta é idêntica a uma das opções propostas e, finalmente, escolher uma delas. Os itens objetivos permitem verificar memorização, compreensão, análise, síntese de ideias e avaliação.

Os itens objetivos em um teste trazem muitas vantagens, entre elas destacam-se a rapidez de correção e ampliação de cobrança por conteúdo (ANASTASI, 1977). Esses itens são mais confiáveis para validade dos resultados e são recomendados sempre que o número de avaliados for muito grande ou quando há pressão com relação aos resultados (VIANNA, 1982).

A indução ao erro na construção de itens deve ser evitada, isso se deve ao fato de exigir atenção (por parte do candidato) a detalhes, que apontam itens incorretos como resposta, mesmo que o candidato domine o conteúdo. Assim, um juízo de valor incorreto sobre o candidato é construído.

Na maioria dos itens do Enem, é proposta uma situação-problema. Entendemos que uma situação-problema é um desafio proposto pelo item, que exige reflexão, organização de ideias e construção de soluções por parte do examinado. O Inep orienta que toda situação-problema deve ser contextualizada, permitindo ao candidato incorporar conhecimentos adquiridos ao longo de sua vida seja no ambiente escolar ou fora dele.

Para o Inep (2010, p.08),

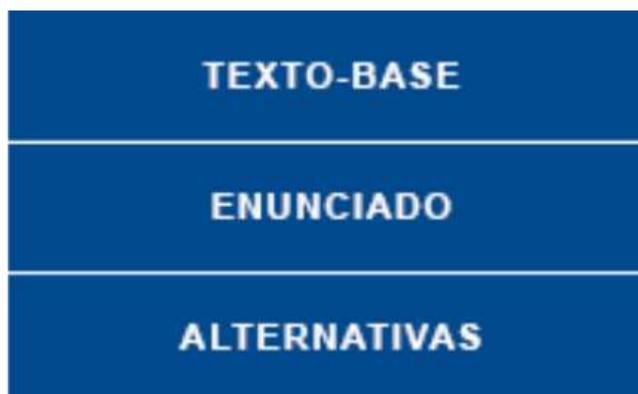
“Uma situação-problema não deve se restringir a uma parte específica do item, mas deve permear toda a sua estrutura, ao longo de todo o processo de composição, a começar pela escolha do texto-base, passando pela construção de todas as partes que compõem um item”

A contextualização do item busca fazer com que o examinado pense no problema e compare-o a algo que possa acontecer ou que tenha acontecido em seu cotidiano. Os itens de múltipla escolha exigem que o elaborador tenha amplo domínio da área abordada, dos objetivos avaliados, das características dos candidatos e dos procedimentos técnicos sobre a elaboração de itens. O avaliador deve também propor itens que exigem um raciocínio criativo

para serem solucionados (VIANNA, 1982).

Os itens selecionados são revisados várias vezes por profissionais distintos na tentativa de identificar possíveis erros de elaboração. O item no Enem é dividido em três partes (INEP, 2010), como mostramos na figura a seguir.

Figura 8 - Partes constitutivas do item



Fonte: Inep (2010)

O Inep afirma que o item deve contemplar apenas uma habilidade na Matriz de Referência, porém qualquer questão de um exame de proficiência exige mais de uma habilidade para que a resposta apontada seja correta (CRONBACH, 1963).

Segundo o Inep (2010, p.09),

O item deve ser estruturado de modo que se configure uma unidade de proposição e contemple uma única habilidade da Matriz de Referência. Para tanto, devem ser observadas a coerência e a coesão entre suas partes (texto-base, enunciado e alternativas), de modo que haja uma articulação entre elas e se explicita uma única situação-problema e uma abordagem homogênea de conteúdo.

O texto-base deve compor ou motivar a situação-problema por meio de imagens, tabelas, gráfico, infográficos, quadros, textos verbais e outros. Os textos-base podem ser construídos pelo próprio autor do item, ou por publicação pública referenciada.

O enunciado do item é composto por orações, que não apresentam informações adicionais ao texto-base e deve mostrar, de forma simples e objetiva, qual ou quais são os objetivos do item.

As alternativas são compostas por itens em que são incluídos o gabarito e os distratores. O gabarito é a única resposta correta e os distratores são as respostas incorretas, que parecem estar corretas. Os candidatos que escolhem os distratores, em geral, não

desenvolveram a habilidade necessária que o item exige (INEP, 2010).

O distrator plausível deve retratar hipóteses e raciocínio utilizadas na busca da solução da situação-problema apresentada. Como consequência, se esse distrator retrata uma dificuldade real do participante com relação à habilidade, não devem ser criadas situações capazes de induzi-lo ao erro. (INEP, 2010, p.11)

Os erros comuns, que são observados no cotidiano escolar em itens semelhantes, são levados em consideração na construção do item, o que provoca um aumento da quantidade de distratores. Nas alternativas também existem opções que fogem muito dos padrões de resposta, logo, nos itens também há alternativas que aumentam a probabilidade de acerto.

As justificativas de alternativa devem ser formuladas pelos elaboradores e fazem parte do protocolo de apresentação do item (INEP, 2010). Os revisores possuem autonomia para modificar ou sugerir mudanças nos itens e suas justificativas devem informar exatamente os motivos pelos quais cada uma das alternativas representa, ou não, a opção correta de resposta, de modo que não serão aceitas justificativas tautológicas (INEP, 2010, p. 12).

## 5 O FUNCIONAMENTO DIFERENCIAL DO ITEM (DIF)

O Estudo sobre DIF teve início nos Estados Unidos (1958-1960) em uma época em que a discussão sobre os direitos civis, desigualdades de oportunidades educacionais entre etnias, desigualdades de oportunidades de emprego eram temas recorrentes no país (COLE, 1993).

Geralmente, a presença de DIF em um teste está associada ao viés do item. Entende-se como viés do item a diferença significativa entre parâmetros de um mesmo item (discriminação ou dificuldade) para grupos diferentes, ou seja, os valores referentes a discriminação e (ou) dificuldade são diferentes para grupos diferentes submetidos a um mesmo teste ou teste paralelo.

Alguns autores, como Camilli e Shepard (1994), assumem que a presença de DIF não implica em viés do item, contrariando Cole e Moss (1989), os quais afirmam que o viés do item é causado pela presença de DIF. A análise de DIF permite que se extraia padrões de resposta diferenciados das respostas dadas ao teste.

Em um teste, há DIF quando é possível identificar sujeitos com mesma habilidade latente, que possuam diferentes probabilidade de acerto a um mesmo item (ANDRIOLA, 2001). A ideia básica de comparar, de acordo com o nível de habilidade, nos remete ao conceito de igualdade e pode ser confirmada pelo paradoxo de Simpson, segundo o qual devemos comparar aquilo que é comparável (ANDRIOLA, 2002). Para Fidalgo (2012, p.02), “se um item tem DIF, necessariamente, apresentaria diferentes propriedades estatísticas entre os grupos comparados”.

Os indícios sobre o viés do item foram percebidos já em 1910 por Alfred Binet quando, em seus testes psicológicos, alguns grupos de indivíduos de condições socioeconômicas distintas obtinham resultados muito diferentes. Hambleton (1997) afirma que a presença de DIF está ligada a padronização dos testes educacionais no âmbito da TRI. “A presença de DIF num teste é um fator que torna o processo avaliativo injusto, sendo imperativo conhecer o conceito de DIF para compreender-se tal afirmação” (ANDRIOLA, 2002, p.06).

Segundo Fidalgo (2012), pode acontecer de existir estatísticas diferentes entre grupos e não haver DIF. Isso ocorre devido ao impacto (*impact*). Supondo que os homens tenham maior habilidade em resolver problemas matemáticos do que as mulheres, considerando que isso seja verdade, então podemos esperar que geralmente, em média, os homens consigam maiores proficiências em testes de Matemática do que as mulheres. Dessa forma, tais testes

mostrariam apenas diferenças reais nessa habilidade e não podemos considerar tal diferença como DIF, mas como impacto.

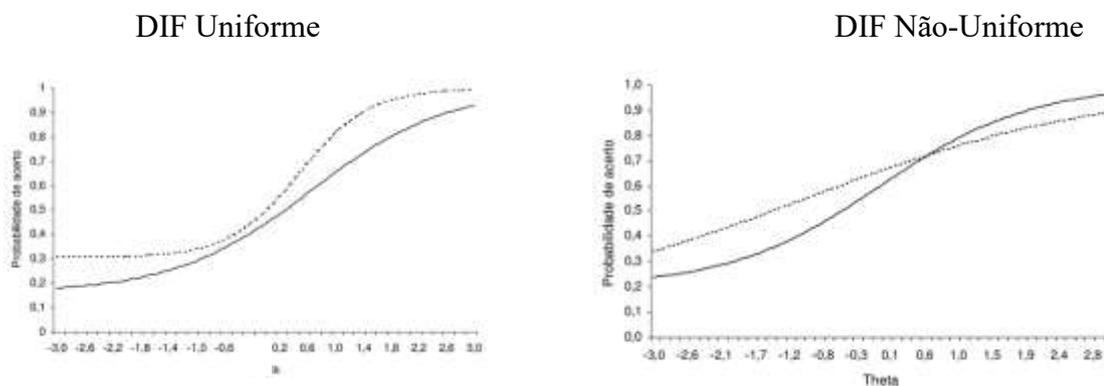
Matematicamente um item não terá DIF se e somente se  $P(X | g, z) = P(X | z)$ . Ao se estudar o conceito de DIF estamos comparando Grupos ( $g$ ) de acordo com a variável  $Z$  (que representa a habilidade  $\theta$ ).

- $X$ = pontuação obtida no item;
- $g$ = valor obtido de acordo com a variável  $G$ ;
- $z$ = valor obtido com relação a variável  $Z$ ;

Assim, temos que a esperança  $E(X | g, \theta) = E(X | \theta)$  para todo  $g$  e  $\theta$  (ANDRIOLA, 2006). Em se tratando dos modelos dicotômicos, ou mesmo dicotomizados como no caso do Enem, a esperança será obtida por meio das probabilidades de acerto ao item. Ou seja,  $P(1 | g, \theta) = P(1 | \theta)$  de acordo com o grupo e seu nível de habilidade. Os modelos dicotômicos são os modelos que recebem o valor 1 (um) quando o item foi respondido corretamente e recebem 0 (zero) quando o item foi respondido de forma incorreta. Nos cadernos do Enem, os itens possuem 5 opções de resposta (a, b, c, d ou e) em que apenas um deles é o correto.

No âmbito da TRI, a forma mais fácil de se perceber uma possível presença de DIF é por meio das curvas características dos itens (CCI) para diferentes grupos. Há vários tipos de DIF (BOCK, 2003) e, neste trabalho, consideraremos dois grupos: o primeiro chamado de grupo referência (GR) e outro de grupo focal (GF). Da mesma maneira, apresentamos dois tipos de DIF que são a DIF Uniforme e a DIF Não-Uniforme.

Figura 9 -TIPOS DE DIF



Fonte: Gleiber, Primi (2011)

Há DIF Uniforme quando não ocorre cruzamento entre as curvas características dos itens para GR e GF ao longo da habilidade latente medida, caso contrário teremos DIF Não-Uniforme. A linha tracejada (figura 09) se refere ao GR, enquanto que a linha contínua representa GF. Devido ao fato de a linha contínua estar completamente abaixo da linha tracejada (DIF Uniforme), podemos concluir que o item é mais fácil para GF em relação a GR ao longo de toda habilidade Theta. A presença de DIF Uniforme, nesse caso, indica que o GF foi beneficiado com relação ao item estudado.

Para o gráfico DIF Não-Uniforme, o item é mais fácil para GF desde  $\Theta = -3$  até o cruzamento das curvas e a partir desse ponto torna-se mais fácil para GR. Nesse exemplo, a questão de conceituar o grupo mais beneficiado com a presença de DIF é discutível, pois o item é benéfico a um determinado grupo até uma certa habilidade e a partir dela, o item beneficiaria outro grupo. Em contrapartida, o item exigiria uma maior habilidade (Theta) para que a probabilidade de seu acerto fosse menor. A identificação de DIF em um teste mostra que as diferenças existem em vários níveis de conhecimento; e a presença de DIF exige a utilização de métodos pedagógicos de recuperação de aprendizagem aos grupos que não conseguiram alcançar os objetivos almejados pelos sistemas de ensino.

### **5.1 Os Métodos Mantel-Haenszel (MH)**

Existem muitas estatísticas disponíveis sobre DIF para itens dicotômicos, ou para itens politômicos (FIDALGO, 2002). Andriola (2018) descreve alguns métodos de detecção de DIF e cita como principais métodos os seguintes: o método do cálculo entre áreas de CCI's; o método da comparação entre probabilidades; o método de comparação entre CCI's; o método da comparação entre parâmetros dos itens; o método Qui-quadrado de Lord; o método Qui-quadrado de Sheuneman; o método Qui-quadrado de Pearson; o método da Regressão Logística; o método Logístico interativo e os métodos de Mantel-Haenszel (MH).

#### **O método $\alpha$ MH**

O método  $\alpha$ MH (alfa Mantel-Haenszel) foi criado em 1959 por N. Mantel e W.Haenszel e, em 1998, P.W.Holland e D.Thayer aplicaram o método ao estudo do DIF (HOLLAND, WAINER, 1993). O método  $\alpha$ MH trabalha com as frequências (de acertos) esperadas e observadas para o grupo de referência e para o grupo focal, segundo os níveis de habilidade de interesse. As frequências podem ser organizadas em uma tabela de acordo com

os grupos de referência e o grupo focal. A seguir, mostramos um exemplo:

Tabela 4 - Tabela de frequências observadas de resposta a um item

GRUPOS	ACERTOS(1)	ERROS(0)	TOTAL
REFERÊNCIA	$A_j$	$B_j$	$N_{r,j}$
FOCAL	$C_j$	$D_j$	$N_{f,j}$
TOTAL	$M_{1j}$	$M_{0j}$	$T_j$

Fonte: Autor (2019)

Para comparar as frequências, a proposta de N.Mantel e W.Haenszel foi definida da seguinte forma:

$$\alpha_{MH} = \frac{\sum_{j=1}^{j=S} \frac{A_j D_j}{T_j}}{\sum_{j=1}^{j=S} \frac{B_j C_j}{T_j}}$$

Equação 8

$A_j$ = frequência observada das respostas corretas no grupo de referência de acordo com as pontuações escolhidas;

$B_j$ = frequência observada das respostas incorretas no grupo de referência de acordo com as pontuações escolhidas;

$C_j$ = frequência observada das respostas corretas do grupo focal de acordo com as pontuações escolhidas;

$D_j$ = frequência observada das respostas incorretas do grupo focal de acordo com as pontuações escolhidas;

$T_j$ = Número de participantes de cada grupo de acordo com as pontuações escolhidas;

$m_{1j}$ =soma dos acertos no nível de pontuação  $j$ ;

$m_{0j}$ =soma dos erros no nível de pontuação  $j$ .

$N_{r,j}$  =quantidade de respondentes do item  $j$  no GR;

$N_{f,j}$  =quantidade de respondentes do item  $j$  no GF.

A comparação entre GR e GF com base no teste  $\alpha_{MH}$  pode ser interpretada da seguinte maneira (ANDRIOLA, 2018):

$\alpha_{MH}=1,0$  implica em item sem DIF;

$\alpha_{MH}< 1,0$  implica em DIF benigno ao GR;

$\alpha_{MH} > 1,0$  teremos DIF benigno ao GF;

$$\Delta_{MH} = \ln(\alpha_{MH})$$

### O método $\chi^2_{MH}$

O método qui-quadrado Mantel-Haenszel ( $\chi^2_{MH}$ ), como os demais testes Mantel Haenszel, utiliza o total de acertos (o escore total) como estimador da variável a ser medida no teste (Fidalgo, 2012). Em outras palavras, o escore será utilizado para comparações entre grupos. O procedimento utiliza-se de tabelas de contingência para cada nível de pontuação, como a tabela 03. O teste  $\chi^2_{MH}$  é um caso especial da estatística generalizada de Mantel Haenszel, assim como a estatísticas  $\alpha_{MH}$ . Os gabaritos dicotomizados devem ser dispostos em D tabelas de contingência 2x2. D será a quantidade de intervalos de divisões dos escores (1,2,3,...,j,...,D). Para cada nível de pontuação j, haverá uma tabela de contingência com os dados do grupo (referência/ focal) em uma das entradas e o valor obtido na resposta (0 ou 1) na outra.

As hipóteses para  $\chi^2_{MH}$  são:

$$\left\{ \begin{array}{l} H_0: \left( \frac{A_j}{B_j} \right) = \mu \left( \frac{C_j}{D_j} \right) \quad ; \\ \mu = 1, \text{ não há DIF;} \\ H_1: H_0: \left( \frac{A_j}{B_j} \right) = \mu \left( \frac{C_j}{D_j} \right) \\ \mu \neq 1, \text{ em algum } j \text{ então há DIF.} \end{array} \right.$$

A estatística proposta por Holland e Thayer (1988) para testar  $H_0$  com relação ao DIF é dada por:

$$\chi^2_{MH} = \frac{(|\sum_1^D A_j - \sum_1^D E(A_j)| - 0,5)^2}{\sum_1^D Var(A_j)}$$

Equação 9

$E(A_j)$  = Esperança de  $A_j$ ;

$$E(A_j) = (n_{1j} * n_{1j}) / N_j$$

Equação 10

$Var(A_j)$  = Variância de  $A_j$ ;

$$Var(A_j) = \frac{n_{1j} * n_{1j} * n_{1j} * n_{1j} * n_{1j} * n_{1j}}{n_{1j}^2 * (n_{1j} - 1)}$$

Equação 11

A  $\chi^2_{MH}$  tem distribuição  $\chi^2$  com um grau de liberdade, o item terá DIF caso  $\chi^2_{MH} > \mu$   $\chi^2_{1^2}$  para uma significância  $\mu$ . O método de se calcular a estatística  $\chi^2_{MH}$  também pode ser visto em Fidalgo (2005).

## 5.2 O cérebro em testes de rendimento

O cérebro humano é complexo e dinâmico. A repetição de exercícios acelera a troca de informação entre neurônios, ajuda a fixar conhecimentos e ativar a memória. A interligação entre conhecimentos já adquiridos faz com que as ideias sobre algum assunto seja rapidamente processada, reorganizando ideias e construindo novos conhecimentos. Para Barbara Oakley, novos conhecimentos potencializam a capacidade criativa do cérebro, ampliando o desempenho em testes de rendimento (BBC, 2019).

Em testes de proficiência, é comum aos candidatos iniciarem resolvendo itens difíceis, o que pode ser um erro. Em entrevista para a revista guia do estudante, Daniel Perry (coordenador do Anglo SP, 2018) afirma que o ideal para o resolver o Enem é o seguinte:

“como o Enem é corrigido pela TRI, que avalia a coerência das respostas em termos pedagógicos e visa evitar o chute, a melhor estratégia é o aluno resolver primeiro as questões que ele considera mais fáceis, de resolução imediata, ou seja, após receber a prova, busque identificar essas questões e resolvê-las. E aí, depois de resolvidas, se dedique às questões médias e, depois, às mais difíceis” (SPAGNA, 2018).

Bárbara Oakley (BBC, 2019) afirma que, em exames com grande número de itens (como no Enem), o ideal é que os itens com grande diferença de dificuldade sejam resolvidos de forma alternada, fazendo com que o cérebro trabalhe em parte de forma focada (atento) e em parte de forma difusa (modo relaxado).

Se o cérebro trabalha de forma focada, então, ao encontrar uma barreira e não conseguir superá-la, fará com que os neurônios (com grande proximidade) percam a conexão e deixem de fornecer e criar estratégias de solução para o problema. Por outro lado, a forma difusa permite que diversas partes do cérebro se comuniquem de modo suave, influenciando o pensamento criativo e a ligação entre ideias. Se o cérebro está trabalhando de forma focada em um problema sem conseguir grandes avanços, então mudar o foco (trabalhar de forma difusa) pode trazer vantagens, assim como afirma a pesquisadora Barbara Oakley em entrevista (IOETA, 2019). As duas formas de funcionamento do cérebro são essenciais e devem trabalhar de forma alternada.

Em exames de larga escala com provas paralelas, pode haver sequências de itens difíceis, de itens fáceis ou até mesmo de itens com dificuldade alternada. No Enem, há quatro tipos de provas paralelas que, segundo o MEC, possuem equidade de dificuldade, pois se utilizam da TRI para a construção dos cadernos.

A dificuldade de um item em um teste pode variar de acordo com a posição, pois é comum aos candidatos responderem aos itens seguindo a sequência dos itens na prova. No decorrer da prova, é comum haver problemas como esgotamento físico, esgotamento mental, falta de tempo, ansiedade e outros fatores que podem contribuir para uma maior dificuldade de resolução ao item (ALEXANDROWICZ, MATSCHINGER, 2008).

Podemos afirmar que a posição do item em um teste pode comprometer suas propriedades psicométricas, assim como pode comprometer o desempenho dos candidato em exames concorridos (ALBANO, 2014). Tal efeito já foi comprovado no Enem 2016 por Mendes, Araújo, Airton e Trompieri (2016) no trabalho “Efeito de posição na dificuldade dos itens no Enem”, publicado na revista ABAVE.

O estudo sobre o DIF em avaliações de larga escala apresenta-se como ferramenta fundamental ao avanço da qualidade educacional no país, uma vez que revela diferenças educacionais, as quais podem ser corrigidas com o apoio do poder público e de uma efetiva gestão educacional.

## 6 ANÁLISE DE VARIÂNCIA (ANOVA)

No campo educacional, é comum definir grupos com base em determinadas características para construir inferências entre os mesmos. Ao trabalhar com avaliações de larga escala, é possível perceber que em geral há diferenças de proficiências entre grupos. Diante das diferenças observadas, é necessário se basear em ferramentas estatísticas que comprovem as hipóteses iniciais.

Os testes estatísticos para tomada de decisão são fundamentais em toda pesquisa científica (BUSSAB, 2006). Os testes de diferença entre médias, ANOVA e os Testes de significância são muito conhecidos e aplicados em pesquisas nas universidades, laboratórios, departamento de controle de qualidade e outros (MONTGOMERY, 2003). Os testes de comparação entre médias avaliam a probabilidade de se cometer o erro tipo I, porém, no teste entre médias, os erros podem ser os seguintes:

\* Erro tipo I, é o erro que se rejeita a hipótese nula ( $H_0$ ) quando ela deve ser mantida.

\* Erro tipo II, é o erro que se aceita a hipótese nula ( $H_0$ ) quando ela deve ser rejeitada.

O teste da diferença entre médias (teste t) trabalha com duas amostras extraídas de uma mesma população ou populações equivalentes, avaliando uma variável quantitativa. O teste trabalha com duas hipóteses: a hipótese nula ( $H_0: \mu_1 = \mu_2$ ), que trata de verificar se as médias ( $\mu_1 =$  média do grupo 1,  $\mu_2 =$  média do grupo 2) são estatisticamente iguais; e a hipótese alternativa ( $H_1: \mu_1 \neq \mu_2$ ) que busca perceber se as médias são diferentes estatisticamente. O teste t avalia, especificamente, a hipótese nula com probabilidade  $\theta$  de concluir, de forma incorreta, por uma diferença inexistente.

A formulação matemática para o teste t é:

$$t = \frac{\overline{X}_1 - \overline{X}_2}{S_{\overline{X}_1 - \overline{X}_2}}$$

Equação 12

$\overline{X}_i$ , corresponde a média do grupo i;

$\overline{X}_1 - \overline{X}_2$ , corresponde a diferença de médias dos grupos 1 e 2;

$S_{\overline{X}_1 - \overline{X}_2}$ , corresponde ao desvio padrão da distribuição da diferença entre médias.

O valor de  $\theta$  utilizado para a conclusão do teste é conhecido como nível de significância. O valor  $\theta$  corresponde a probabilidade em que a hipótese nula pode ser rejeitada

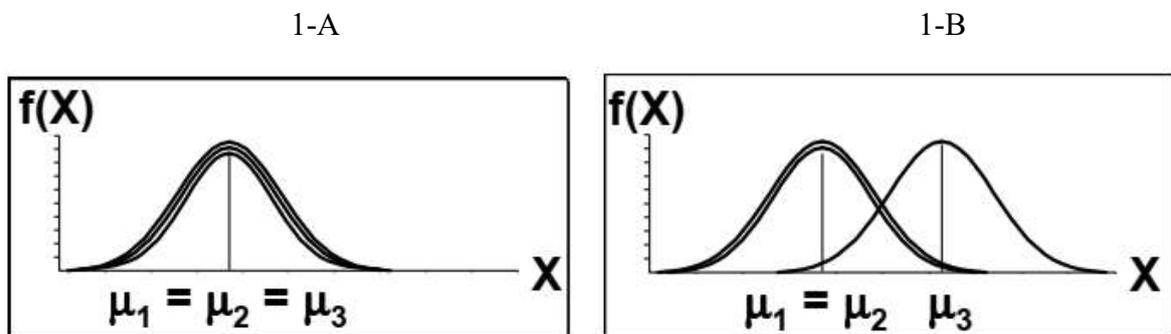
com confiança e a hipótese alternativa aceita com confiança. Se o teste gerar um valor menor que  $\theta$ , então, a hipótese nula é rejeitada, caso contrário será aceita. Quando for necessário verificar a diferença com respeito a uma dada variável para mais de dois grupos, o teste ANOVA é o mais adequado.

O objetivo do teste ANOVA (unidirecional) é verificar se as médias são iguais, ou se alguma (as) difere (m) das demais. Para se utilizar o teste ANOVA, devemos seguir alguns pressupostos: as amostras aleatórias e independentes, as populações normais e homocedasticidade. O teste ANOVA unidirecional trata a variação total dos escores dividindo-a em desvios dos escores brutos, em relação à média do próprio grupo (variação dentro do grupo), e o desvio das médias de cada grupo, em relação às médias dos demais grupos (variação entre grupos) (LEVIN, 2004).

As hipóteses para o teste ANOVA são as seguintes:

$$\begin{cases} H_0: \mu_1 = \mu_2 = \mu_3 = \dots = \mu_n & ; \\ H_1: \exists \mu_i \neq \mu_j, & i \neq j, \quad i, j \leq n \end{cases}$$

Figura 10 - Distribuição das médias para três grupos



Fonte: desconhecido

A figura mostrada anteriormente ilustra o comportamento dos dados para três diferentes grupos submetidos a um teste. Na figura 1-A, ocorre a coincidência das médias, ou seja, o resultado obtido por meio do teste F implica em aceitar a  $H_0$ . Por outro lado, a figura 1-B mostra que existe pelo menos uma das médias difere das demais, portanto podemos concluir que  $H_0$  deve ser rejeitada com base no teste F. O teste F utiliza uma soma de quadrados e graus de liberdade em sua formulação (LEVIN, 2004). O cálculo para as somas de quadrados a serem utilizadas são dadas por:

$$SQ_{total} = \sum_1^n X^2_{total} - N_{total}(\bar{X})^2_{total}$$

Equação 13

$$SQ_{dentro} = \sum_1^n X^2_{total} - \sum_1^n N_{grupo} (\bar{X})^2_{grupo}$$

Equação 14

$$SQ_{entre} = \sum_1^n N_{grupo} (\bar{X})^2_{grupo} - N_{total} (\bar{X})^2_{total}$$

Equação 15

Sendo,

$SQ_{total}$ , a soma dos quadrados total (soma dos quadrados dos desvios de cada escore em relação à média total);

$SQ_{dentro}$ , a soma dos quadrados dentro do próprio grupo (soma dos quadrados dos desvios de cada escore em relação à média do próprio grupo);

$SQ_{entre}$ , a soma dos quadrados entre os grupos (a soma do produto da quantidade de escores pelo quadrado da diferença da média de cada grupo e a média total);

$\sum_1^n X^2_{total}$ , a soma do quadrado de todos os escores;

$\bar{X}_{total}$ , é a média de todos os grupos combinados;

$N_{total}$ , é o número total de escores em todos os grupos combinados;

$\bar{X}_{grupo}$ , é a média de cada grupo;

$N_{grupo}$ , é o número de escores em um grupo.

A soma dos quadrados se torna maior quando a variação tende a aumentar. Por esse motivo, é necessário controlar o número de escores envolvidos para que a medida obtida seja satisfatória (LEVIN, 2004). A média quadrática (variância) é obtida ao se dividir  $SQ_{dentro}$  ou  $SQ_{entre}$  por um número apropriado de graus de liberdade. O cálculo para as médias quadráticas são:

$$MQ_{entre} = \frac{SQ_{entre}}{gl_{entre}}$$

Equação 16

$$MQ_{dentro} = \frac{SQ_{dentro}}{gl_{dentro}}$$

Equação 17

Onde,

$MQ_{entre}$ , média quadrática entre grupos;

$MQ_{dentro}$ , média quadrática dentro do grupo;

$SQ_{dentro}$  e  $SQ_{entre}$  representam as somas quadráticas já definidas anteriormente;

$gl_{dentro}$ , são os graus de liberdade dentro dos grupos

( $gl_{dentro} = N_{total} - k$ ,  $k = \text{quantidade de grupos}$ );

$gl_{entre}$ , são os graus de liberdade entre grupos ( $gl_{entre} = k - 1$ );

$N_{total}$ , já foi definido.

A razão F é definida pela fórmula matemática:

$$F = \frac{MQ_{entre}}{MQ_{dentro}}$$

Equação 18

Após o cálculo da razão  $F$ , é necessário verificar a tabela de distribuição  $F$  (Fischer) e a sua significância de acordo com o número de graus de liberdade adequado, caso o teste  $F$  aponte para uma diferença significativa, deve ser verificado o teste DMS de Fischer.

O teste de diferença mínima significativa (DMS) de Fischer serve para comparar pares de médias, esse método controla a taxa de erro de acordo com o nível de significância adotado, por outro lado não é capaz de controlar a taxa de erro ocorrida no experimento. O teste DMS só deve ser utilizado se a estatística  $F$  for significativa ao nível de significância escolhido. A expressão para o cálculo da  $DMS$  de Fisher é dada por:

$$DMS = t \sqrt{\frac{2 \cdot SQ_{entre}}{r}}$$

Equação 19

A quantidade de grupos é representada por  $r$ , associado ao nível de significância e aos graus de liberdade está o valor de  $t$ . A diferença será significante se for no mínimo igual à  $DMS$  (SONIA, 2016).

## 7 MATERIAL E MÉTODOS

O Enem a partir de 2009 foi construído com base na TRI, essa teoria é aplicada em avaliações de larga escala como Saeb (Sistema de avaliação da educação básica) e PISA (Programa internacional da avaliação de estudantes) e possui grande credibilidade entre especialistas e pesquisadores em avaliação educacional (Inep, 2011). Este trabalho faz um estudo sobre a equidade de dificuldade entre os cadernos de Matemática do Enem 2017 referente a primeira aplicação do exame

A pesquisa e análise construída utiliza o arquivo de Microdados do Enem 2017. O arquivo de dados (Microdados.rar) requer grande poder computacional para que seja possível ler e inferir algo sobre os dados. De posse do arquivo é necessário descompactá-lo e utilizar softwares específicos para ter acesso aos dados. Os softwares Microsoft Excel e o software estatístico R foram utilizados no presente trabalho.

### 7.1 Dos arquivos gerados

Os dados utilizados neste trabalho seguiram algumas definições que representam a grande massa do atual sistema educacional público. A seleção dos dados analisados foi feita da seguinte maneira: os dados devem ser referentes a 1ª aplicação do Enem 2017, os dados utilizados são dos alunos que estão concluindo o ensino médio, idade maior ou igual a 16 anos e menor ou igual a 19 anos, candidatos presentes e não eliminados em nenhuma das provas. Os critérios acima foram adotados com o objetivo de analisar os candidatos mais comprometidos com o exame, que estão em idade adequada para concluir o ensino médio e que estão concluindo o ensino médio, assim este trabalho construirá uma boa representação sobre os dados estudados.

A idade combinada com o fator concluindo são fatores determinantes na análise, visto que esses educandos tornam a amostra mais representativa quando nos referimos aos alunos de escolas públicas. A diretora executiva do Todos pela educação (2013) Priscila Cruz, afirma que a idade ideal para o término do ensino médio é 17 anos, porém, há uma grande quantidade de alunos que concluem aos 19 anos (TOKARNIA, 2013). Não houve inclusão de candidatos com idade maior que 19 anos pelo fato de tais alunos estarem fora da escola ou mesmo sem estudar a muito tempo e a inclusão desses poderia criar um viés em nossas análises. Após a definição dos critérios para seleção dos dados foi gerado uma série de arquivos para serem utilizados nas análises. Os gabaritos das provas foram dicotomizados, após isso foram gerados

vários outros arquivos em formato de planilha e em formato de texto para serem utilizados e modificados quando necessário.

Tabela 5 - Gabaritos por prova

CÓDIGO DA PROVA	COR DA PROVA	QUANTIDADE
403	AZUL	292.517
404	AMARELA	292.823
405	ROSA	292.049
406	CINZA	291.597

Fonte: Autor (2019)

## 7.2 Dos resultados

O quadro abaixo mostra que as provas AMARELA e CINZA possuem estatísticas descritivas semelhantes, as provas AZUL e ROSA possuem desvios semelhantes com diferença média aproximada de 4 pontos. O teste ANOVA é necessário para comprovar se há diferença significativa entre as médias das proficiências das provas.

Tabela 06 - Estatísticas Descritivas

PROVA	QUANTIDADE	MÍNIMO	MÁXIMO	MÉDIA	DESVIO PADRÃO
ROSA	292.049	350,1000	988,3000	521,3018	106,7713
AZUL	292.517	350,1000	990,2000	525,1197	106,8220
CINZA	291.597	350,1000	993,9000	527,5871	105,8579
AMARELA	292.823	350,1000	992,8000	527,6583	105,2526

Fonte: Autor (2019)

O teste Anova (unidirecional) confirmou que existe pelo menos uma das médias diferentes das demais.

Tabela 7 - O teste ANOVA

NOTA	SOMA DOS QUADRADOS	DF	MÉDIA QUADRÁTICA	F	SIGNIFICÂNCIA
ENTRE GRUPOS	7.815.915,9090	3	2.605.305,3030	231,0950	,0000
DENTRO DO GRUPO	13.178.782.017,3830	1.168,9820	11.273,7250		
Total	13.186.597.933,2920	1.168,9850			

Fonte: Autor (2019)

Os resultados apresentados na tabela abaixo comprovam que as provas AMARELA e CINZA não apresentam diferença significativa, por outro lado qualquer outra combinação apresenta tal diferença. A semelhança entre as provas AMARELA e CINZA era esperada desde a visualização das estatísticas descritivas.

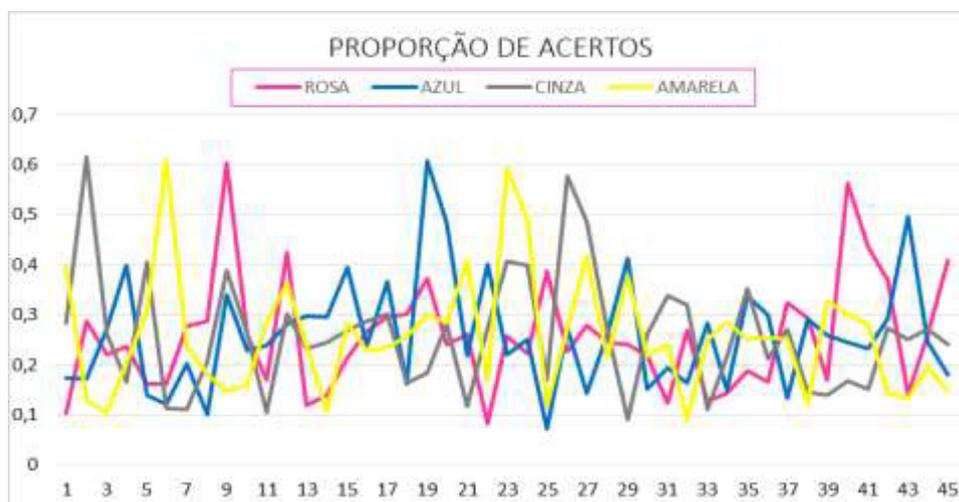
Tabela 8 - DMS DE FISCHER

PROVAS		DIFERENÇA DE MÉDIAS	DESVIO PADRÃO	SIGNIFICÂNCIA	95% DE CONFIANÇA	
					LIM.INFERIOR	LIM.SUPERIOR
ROSA	AZUL	-3,8179	0,2777	0,0000	-4,3623	-3,2735
ROSA	CINZA	-6,2853	0,2779	0,0000	-6,8301	-5,7405
ROSA	AMARELA	-6,3565	0,2776	0,0000	-6,9007	-5,8123
AZUL	CINZA	-2,4674	0,2778	0,0000	-3,0120	-1,9228
AZUL	AMARELA	-2,5385	0,2775	0,0000	-3,0826	-1,9946
CINZA	AMARELA	-0,0711	0,2777	0,7978	-0,6156	0,4733

Fonte: Autor (2019)

Os dois próximos gráficos mostrarão a proporção de acertos por prova na ordem normal dos cadernos e na ordem dos itens idênticos tomando como referência a prova amarela. Pelo gráfico 1, é possível notar que as há uma maior presença de itens com alta proporção de acerto no final dos cadernos ROSA e CINZA, tal fato não ocorre com os cadernos CINZA e AMARELO que apresenta tais itens sempre em posições anteriores ao item 30. A disposição dos itens nos cadernos parece favorecer aos candidatos que receberam as provas CINZA e AMARELA.

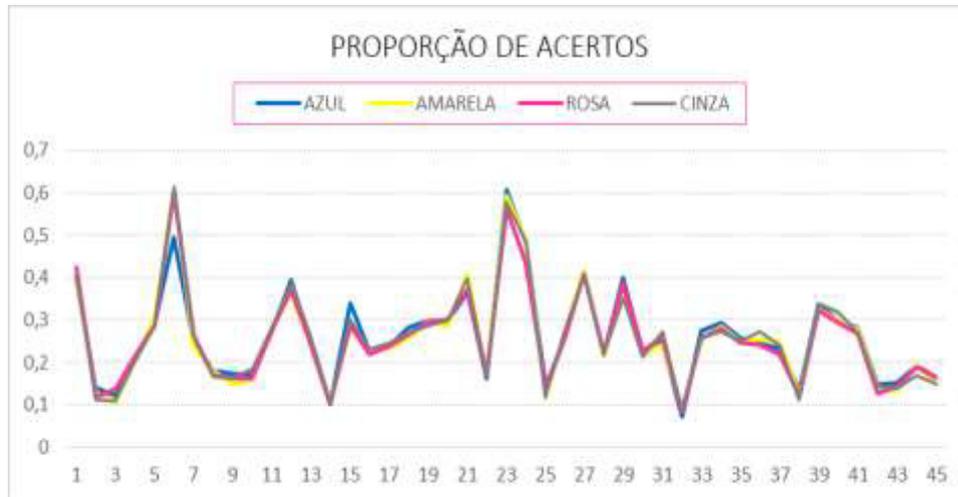
Gráfico 1- Proporção de acertos (ordem normal)



Fonte: Autor. 2019.

Organizando os itens semelhantes de acordo com a prova AMARELA temos o gráfico de proporção de acertos a seguir.

Gráfico 2 - PROPORÇÃO DE ACERTOS (ORDEM DA PROVA AMARELA)

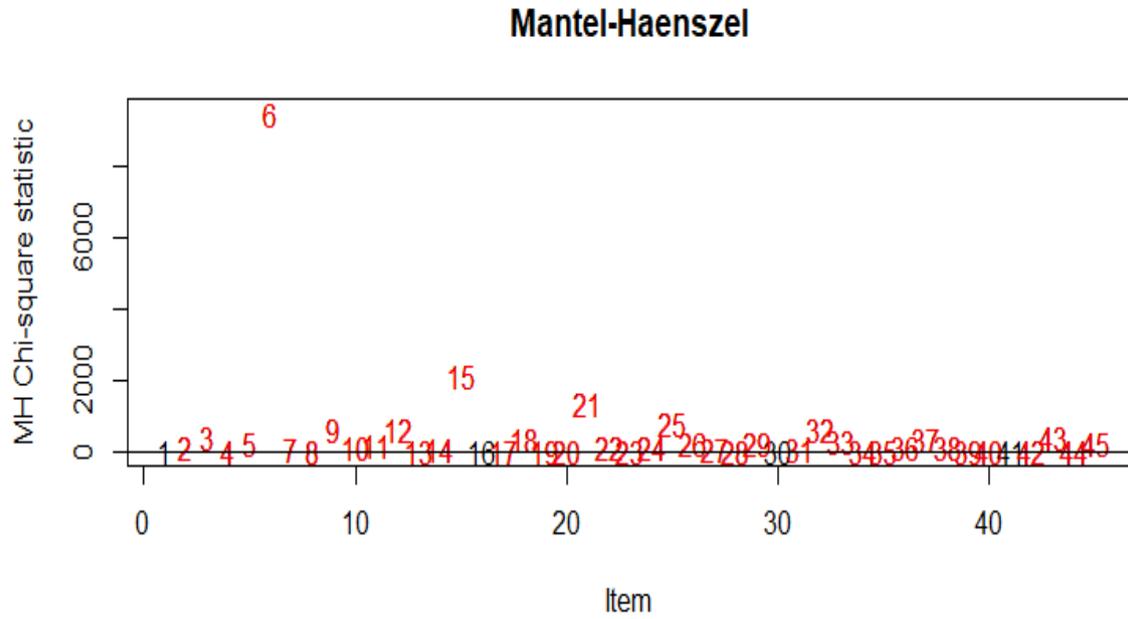


Fonte: Autor (2019)

Visualizando o gráfico acima fica fácil perceber que há vários itens que apresentam significativa diferença de proporção de acertos. Os itens 6, 15, 21, 23, 24, 27, 29, 33, 34, 36, 39, 40, 42, 43, 44 e 45 podem ser facilmente identificados no gráfico acima e devem ter causado as diferenças observadas nas estatísticas descritivas. Os gráficos AMARELO e CINZA são muito próximos, nos itens com proporção de acertos maiores que 0,3 é possível notar que o gráfico ROSA quase sempre está abaixo dos gráficos CINZA, AZUL e AMARELO.

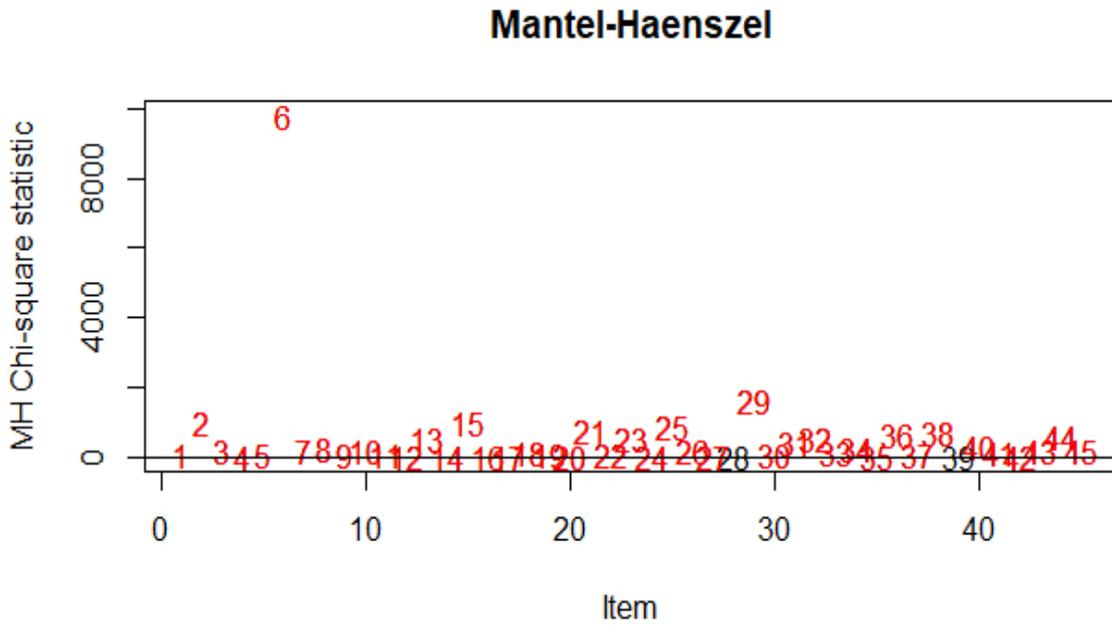
O teste DIF foi executado, utilizaremos a notação REF (grupo de maior nota média) para indicar o grupo de referência e FO (grupo de menor nota média) para indicar o grupo focal. Após a verificação de DIF por meio da estatística qui-quadrado Mantel-Haenszel a confirmação será validada pelo teste de DIF alpha Mantel-Haenszel. As tabelas com os valores das estatísticas calculadas estão no apêndice, os gráficos a seguir correspondem aos valores do teste DIF para a Estatística Chi-quadrado Mantel Haenzel.

Gráfico 3 - DIF (AMARELA (REF)-AZUL (FO))



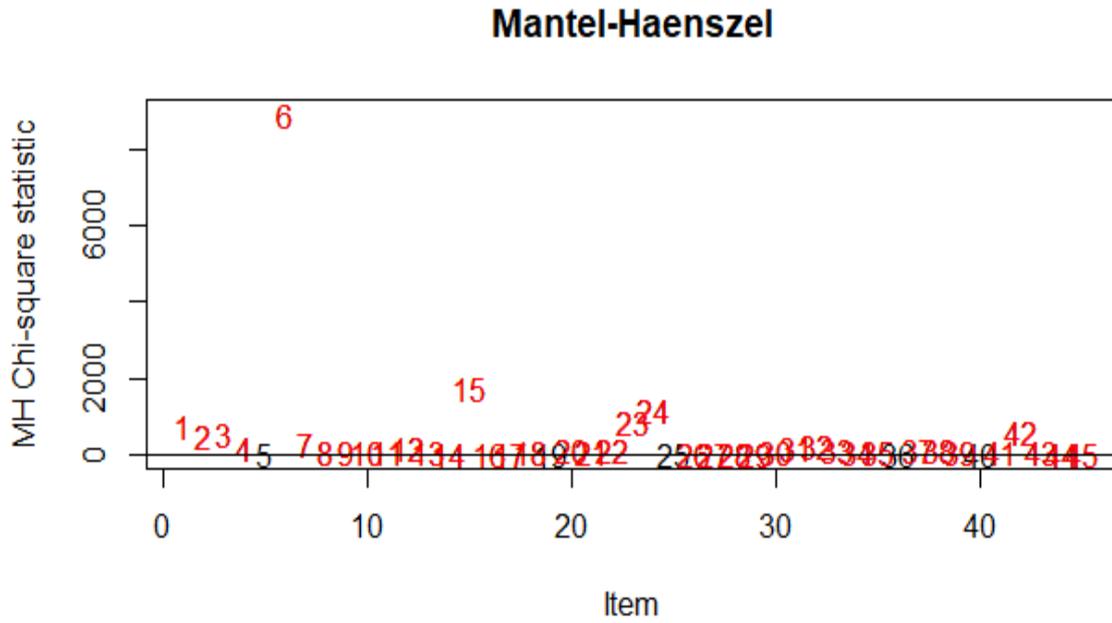
Fonte: Autor (2019)

Gráfico 4 - DIF(AZUL(REF)-CINZA(FO))



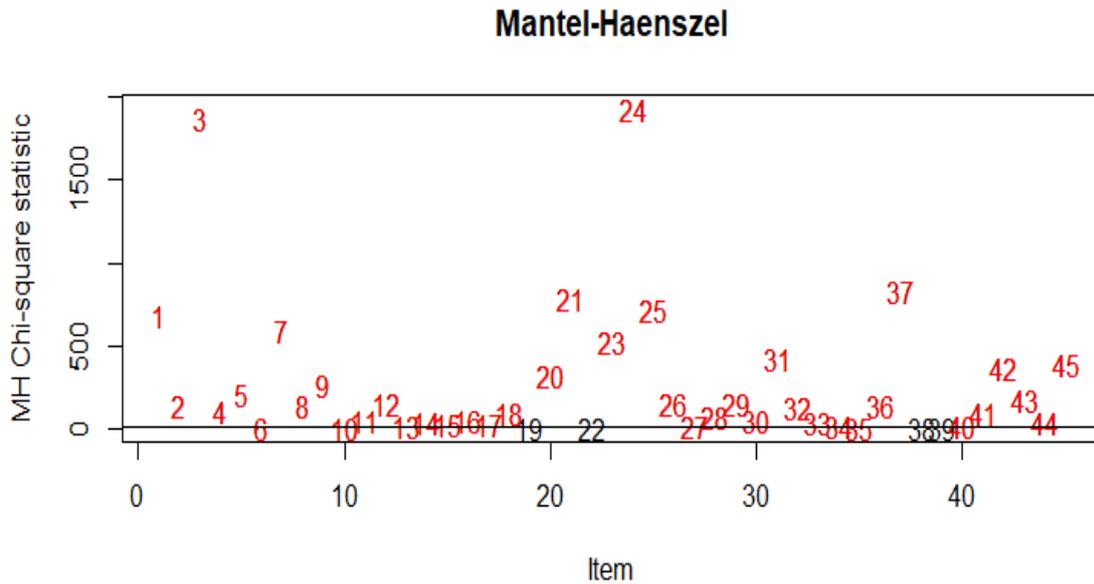
Fonte: Autor (2019)

Gráfico 5 - DIF(AZUL(REF)-ROSA(FO))



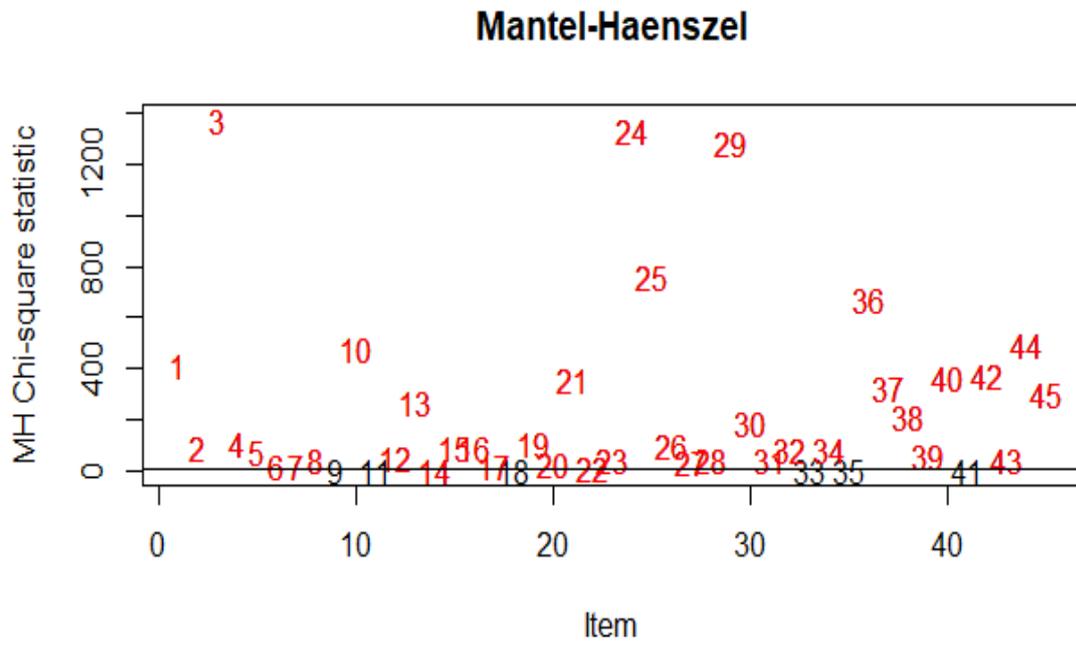
Fonte: Autor (2019)

Gráfico 6 - DIF(AMARELA(REF)-ROSA(FO))



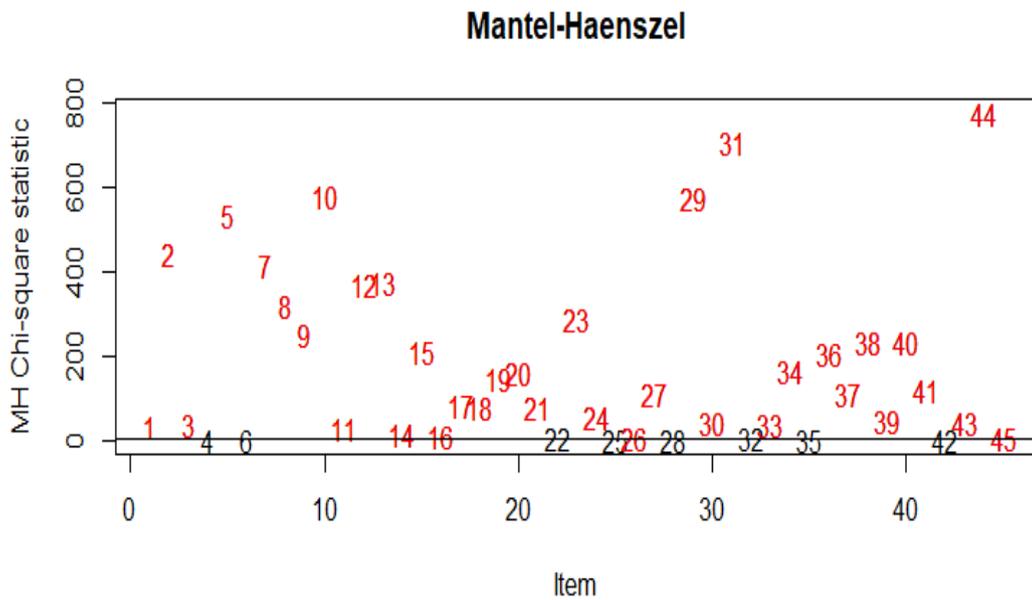
Fonte: Autor (2019)

Gráfico 7- DIF(CINZA(REF)-ROSA(FO))



Fonte: Autor (2019)

Gráfico 8 - DIF(AMARELA(REF)-CINZA (FO))

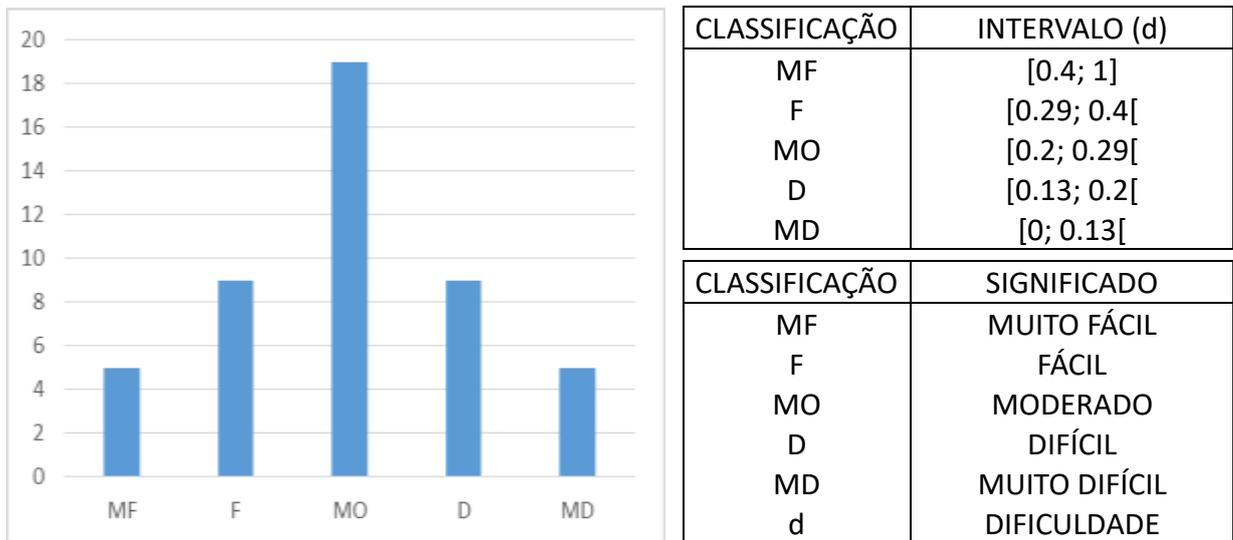


Fonte: Autor (2019)

Em todas as combinações de provas há presença de DIF de acordo com as estatísticas referenciadas no texto. Nota-se que entre a estatística  $\chi^2_{MH}$  as provas AMARELA e CINZA é sempre inferior a 800 com apenas 4 itens com valores entre 500 e 800. Entre as provas AMARELA e ROSA há 6 itens com estatística  $\chi^2_{MH}$  com valores acima de 600. A presença de DIF em um teste de rendimento é natural, por outro lado o efeito não pode causar grandes discrepâncias de rendimento entre os avaliados pois a presença de DIF significativo já indica injustiça entre as provas (ANDRIOLA, 2002).

A distribuição dos itens por nível de dificuldade (Gráfico 9) seguirá o modelo proposto por Pasquali (2003). A distribuição proposta por Cerdá (1978) não pode ser aplicada por não se adequar aos dados.

Gráfico 9 - Classificação dos itens por nível de dificuldade



Fonte: Autor (2019)

Na tabela abaixo estão dispostos os itens por prova ordenados de acordo com o mesmo item da prova AMARELA. Como exemplo temos o item 1 (questão 136) da prova AMARELA que corresponde ao item 4 (questão 139) na prova AZUL, ao item 12 (questão 147) na prova ROSA e ao item 5 (questão 140) na prova CINZA.

Tabela 9 - Itens por prova ordenados pela prova amarela

Am	Az	Rs	Cz		Am	Az	Rs	Cz		Am	Az	Rs	Cz		Am: AMARELA			
1	4	12	5		16	10	3	13		31	24	21	28		Az: AZUL			
2	5	13	6		17	11	4	14		32	25	22	29		Rs: ROSA			
3	6	14	7		18	12	16	15		33	26	23	30		Cz: CINZA			
4	7	15	8		19	13	17	16		34	38	27	42					
5	42	8	1		20	14	18	17		35	39	28	43		MUITO FÁCIL			
6	43	9	2		21	17	42	24		36	40	29	44		FÁCIL			
7	44	10	3		22	18	39	25		37	41	30	45		MODERADA			
8	45	11	4		23	19	40	26		38	37	31	33		DIFÍCIL			
9	1	5	18		24	20	41	27		39	35	37	31		MUITO DIFÍCIL			
10	2	6	19		25	27	43	21		40	36	38	32					
11	3	7	20		26	28	44	22		41	33	32	37					
12	15	19	9		27	29	45	23		42	34	33	38					
13	16	20	10		28	21	24	34		43	30	34	39					
14	8	1	11		29	22	25	35		44	31	35	40					
15	9	2	12		30	23	26	36		45	32	36	41					

Fonte: Autor (2019)

Tabela 10 - Quantidade de itens por dificuldade nos cadernos

QUANTIDADE DE ITENS POR PROVA				
CLASSIFICAÇÃO	ROSA	AZUL	CINZA	AMARELA
MF	5	5	5	5
F	9	9	9	9
MO	17	17	17	17
D	10	9	9	9
MD	4	5	5	5

Fonte: Autor (2019)

Figura 11- PROPORÇÃO DE ACERTOS POR ITEM E POR CADERNO

COR DA PROVA				COR DA PROVA				COR DA PROVA						
1	0,1635847	0,1730051	0,282751	0,3954265	16	0,1466731	0,2375525	0,2873025	0,2174882	31	0,3223123	0,1932796	0,3374739	0,1405896
2	0,2871421	0,1715596	0,6142173	0,1129567	17	0,2946246	0,3654065	0,3010531	0,133621	32	0,2680384	0,1634549	0,3189988	0,3881181
3	0,2195155	0,2734781	0,2605138	0,1032151	18	0,3018025	0,1608773	0,1639334	0,2574971	33	0,1269459	0,2636182	0,1119094	0,2515711
4	0,237321	0,3975628	0,1661787	0,203971	19	0,3714933	0,606855	0,1841277	0,2984909	34	0,1429515	0,1497097	0,1168475	0,1853451
5	0,1623732	0,2394819	0,4051812	0,3061373	20	0,1411824	0,4836126	0,2820071	0,2846556	35	0,1884933	0,336852	0,3517668	0,1498882
6	0,161059	0,1108103	0,2121599	0,6081441	21	0,2620951	0,1184203	0,1179469	0,4071556	36	0,1654743	0,2991905	0,212868	0,2540384
7	0,2764913	0,2027249	0,2103111	0,2382885	22	0,1805386	0,3995892	0,1393971	0,1693515	37	0,3237777	0,1341309	0,2898932	0,2503464
8	0,2880458	0,1014171	0,205315	0,1821446	23	0,256418	0,2111029	0,4073825	0,5823256	38	0,2933068	0,2918229	0,1452371	0,2138179
9	0,8027904	0,3398318	0,3873631	0,1479141	24	0,1111521	0,2902554	0,3985349	0,4888736	39	0,1698249	0,257852	0,1399615	0,3270489
10	0,2617909	0,2281719	0,2620794	0,159493	25	0,3871185	0,0722838	0,1687909	0,1173438	40	0,5623246	0,1437319	0,1681053	0,2990396
11	0,1698831	0,2390708	0,1044928	0,2858503	26	0,1263325	0,2796563	0,5773218	0,2738249	41	0,4340779	0,2303388	0,1506397	0,2804705
12	0,4147025	0,2820961	0,3021431	0,361342	27	0,179617	0,1442045	0,4848311	0,4163668	42	0,3709627	0,2919527	0,2724495	0,1437905
13	0,1195367	0,2978408	0,1325376	0,2389681	28	0,1452071	0,2998297	0,2714524	0,2135574	43	0,1406274	0,4868515	0,2502281	0,1331708
14	0,1386117	0,2951171	0,2453175	0,1071309	29	0,1400119	0,4132874	0,0896434	0,3760819	44	0,2537187	0,2465581	0,2716531	0,1950694
15	0,2133902	0,395263	0,1692141	0,2834642	30	0,1880367	0,1505298	0,1801225	0,2187222	45	0,4086319	0,1805814	0,2400177	0,1476103
	MUITO FÁCIL					MODERADO					MUITO DIFÍCIL			
	FÁCIL					DIFÍCIL								

Fonte: Autor (2019)

No primeiro terço da prova os itens fáceis ou muito fáceis se apresentam da seguinte forma: 4 itens na prova ROSA, 5 itens na prova AZUL, 5 itens na prova CINZA e 5 na prova AMARELA. No segundo terço da prova os itens fáceis ou muito fáceis tem a distribuição: 4 itens na prova ROSA, 5 na prova AZUL, 6 na prova CINZA e 7 itens na prova AMARELA. Para o último terço da prova nota-se que há 6 itens fáceis ou muito fáceis na prova ROSA, 4 itens na prova AZUL, 3 itens na prova CINZA e apenas 2 itens na prova AMARELA. Ao chegar aos últimos itens no fim dos cadernos é possível notar a exaustão em muitos candidatos, vários sinais de esgotamento físico são perceptíveis como esticar o braço ou pernas e se retorcer nas carteiras, esse esgotamento físico tende a fazer com que o candidato tente ler e responder ao item de forma rápida não observando os detalhes que podem fazer com que o erro ocorra. A perda de itens fáceis determina uma grande perda de proficiência pois a TRI utiliza o modelo matemático de avalia o padrão de resposta informado pelo candidato.

A TRI utiliza a coerência pedagógica no cálculo das notas. Por exemplo, se em um teste de 5 itens (com 2 difíceis e 3 fáceis) há dois candidatos com os seguintes resultados:

Candidato 1- acertou 2 itens difíceis e 1 item fácil;

Candidato 2-acertou 3 itens fáceis.

O candidato 1 terá menor proficiência que o candidato 2 pois a TRI avalia a sequência

(padrão de resposta) de itens apontados como certo pelo candidato, neste caso o acerto aos 2 itens difíceis pode ser visto como acerto ao acaso uma vez que o mesmo errou um item fácil. A ideia parte de que se o candidato possui habilidade suficiente para responder corretamente a itens difíceis, então ele também deveria responder corretamente a itens de menor dificuldade, e se isso não se confirmar, o acerto a um item difícil será visto como acerto ao acaso.

Verificar a quantidade de itens em que uma prova possui maior proporção de acertos com relação a outra pode não se configurar como fator decisório para uma maior nota. A tabela abaixo compara a quantidade de itens de acordo com maior proporção de acertos por item. Por exemplo, houve 21 itens na prova AMARELA com proporção de acertos maior que na prova ROSA que teve 24, por outro lado a nota média da prova AMARELA é superior a nota média da prova ROSA em mais de seis pontos. Com relação as provas AMARELA e CINZA apesar da taxa de acerto ser maior em 25 itens da prova AMARELA contra 20 itens da prova CINZA não houve diferença significativa entre médias confirmando que nem sempre maior quantidade de acertos resulta em maior nota.

Tabela 11- Quantidade de itens com maior proporção de acerto na prova 1 contra prova 2

PROVA 1	PROVA 2	QUANTIDADE
AMARELA	ROSA	21
AMARELA	AZUL	28
AMARELA	CINZA	25
ROSA	AZUL	16
ROSA	CINZA	20
AZUL	CINZA	24

Fonte: Autor (2019)

A distribuição dos itens nos cadernos por nível de dificuldade é determinante na nota final do candidato. Os itens abaixo (ver tabela 10) serão analisados pedagogicamente assim como outros que julgemos ser necessário. A justificativa se dá pelo fato dos itens estarem em posições que alteram sua dificuldade devido a fatores internos e externos ao teste.

Tabela 12 - Posição do item e proporção de acerto por prova

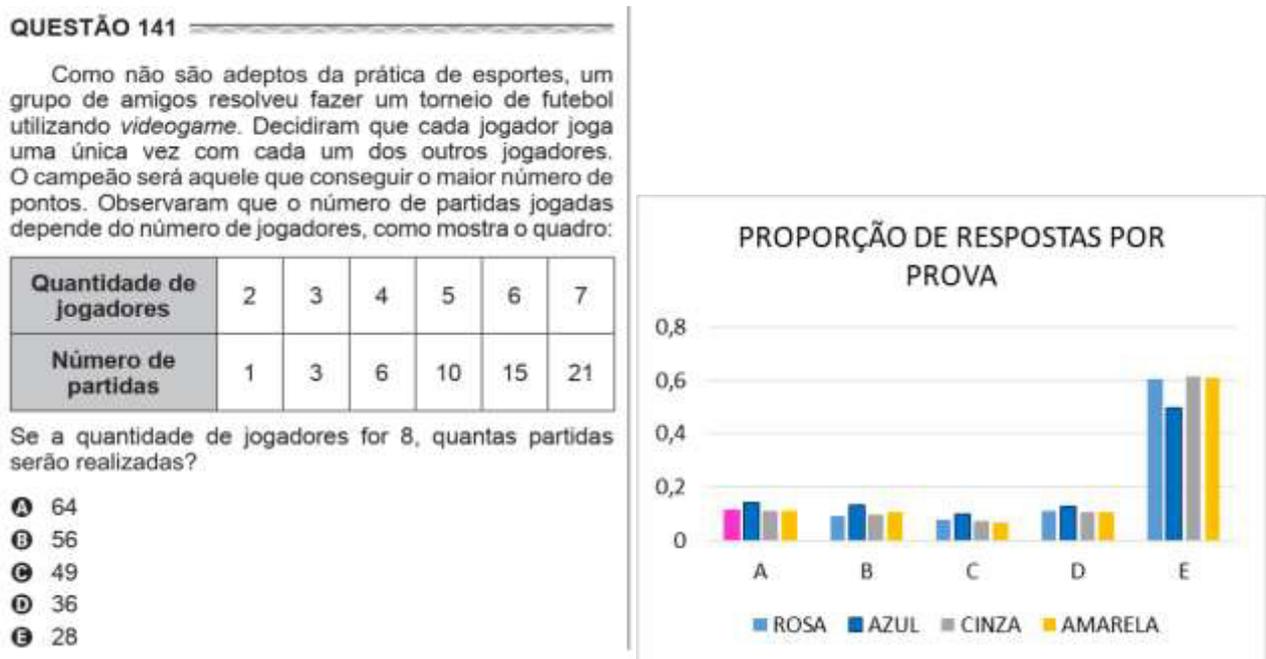
PROVA								
ITEM	ROSA		AZUL		CINZA		AMARELA	
	POSIÇÃO	PROPORÇÃO	POSIÇÃO	PROPORÇÃO	POSIÇÃO	PROPORÇÃO	POSIÇÃO	PROPORÇÃO
1	9	0,60279	43	0,495652	2	0,614217	6	0,608144
2	2	0,287142	9	0,339832	12	0,302143	15	0,283464
3	42	0,370963	17	0,365406	24	0,398535	21	0,407156
4	40	0,562325	19	0,606555	26	0,577322	23	0,592326
5	41	0,434078	20	0,483613	27	0,484831	24	0,489874
6	45	0,408632	29	0,413087	23	0,407383	27	0,416369
7	37	0,323778	35	0,336852	31	0,337479	39	0,327049
8	38	0,293307	36	0,29919	32	0,318999	40	0,299034
9	33	0,126946	34	0,14971	38	0,145237	42	0,143791
10	34	0,142952	30	0,15053	39	0,139961	43	0,133171
11	35	0,188435	31	0,19328	40	0,168105	44	0,195069
12	36	0,165474	32	0,163495	41	0,15064	45	0,14761

Fonte: Autor (2019)

## 8 ANÁLISE PEDAGÓGICA

Os três itens detectados com DIF de maior relevância e presentes no maior número de cadernos serão analisados. As sequências de itens fáceis ou difíceis que surgirem próximos ao final da prova (últimos itens) serão estudados para agregar ainda mais validade aos resultados já obtidos. Os itens serão apresentados conforme a questão na prova amarela.

Figura 12 - Item 1



Fonte: Autor (2019)

Os problemas de contagem são comuns em exames de proficiência e provas de concurso público, em geral não são fáceis. O item exige que o candidato consiga identificar padrões numéricos ou princípios de contagem para poder solucionar o problema. A questão pode ser solucionada através de combinação simples de  $n$  elementos tomados  $p$  a  $p$ , por aritmética básica ou mesmo pela percepção lógica em que a soma dos elementos de cada coluna geram o elemento da próxima coluna localizado na segunda linha. É preciso que o candidato esteja atento a cada detalhe informado pelo item fato esse que pode levar a soluções mais simples. Através do gráfico PROPORÇÃO DE RESPOSTAS percebe-se que houve uma alta taxa de acerto ao item (item E). O item A poderia ser um distrator (falsa resposta mas plausível), os candidatos poderiam ter calculado  $8^2=64$  (por serem 8 jogadores combinados de dois em dois), porém isso não ocorreu (pelo menos não pode ser justificado) mostrando que grande

parte dos alunos conhecem o problema das combinações simples.

**Solução:**

I- Combinação de oito elementos tomados dois a dois.  $C_{8,2} = \frac{8!}{2!(8-2)!} = 28$ .

II- Poderíamos ter utilizado aritmética para resolver verificando os pares:

\*(2,1), (3,3) = (2+1. 1+2)

\*\* (3,3), (4,6) = (3+1. 3+(2+1))

\*\*\* (4,6), (5,10) = (4+1. 6+(2+1+1))

\*\*\*\* (5,10), (6,15) = (5+1. 10+(2+1+1+1))

... continuando de forma análoga chegamos ao par (8, 28). (Resposta. D)

(Resposta. D)

Figura 13 - Uma solução para o item 06

Quantidade de jogadores	2	3	4	5	6	7	8
Número de partidas	1	3	6	10	15	21	28

*(Note: Red arrows in the original image point from the top row to the bottom row, showing the cumulative sum: 2 to 1, 3 to 3, 4 to 6, 5 to 10, 6 to 15, 7 to 21, 8 to 28.)*

Fonte: Inep (2017)

A questão de aparência difícil por envolver uma análise combinatória, pode se tornar fácil por intermédio de uma verificação lógica da resposta. Uma simples soma dos elementos em cada coluna irá gerar o valor disposto na segunda linha da próxima coluna. Neste item era de se esperar que as maiores proporções de acerto quando o item fosse localizado na primeira metade dos cadernos. No caderno AZUL o item está localizado próximo ao fim do caderno, tal fato justifica a menor proporção de acertos ao item pois se trata de um item fácil. Para este item quanto mais próximo do início do caderno maior foi a proporção de acertos e isso era esperado (ver tabela 12, pág. 73).

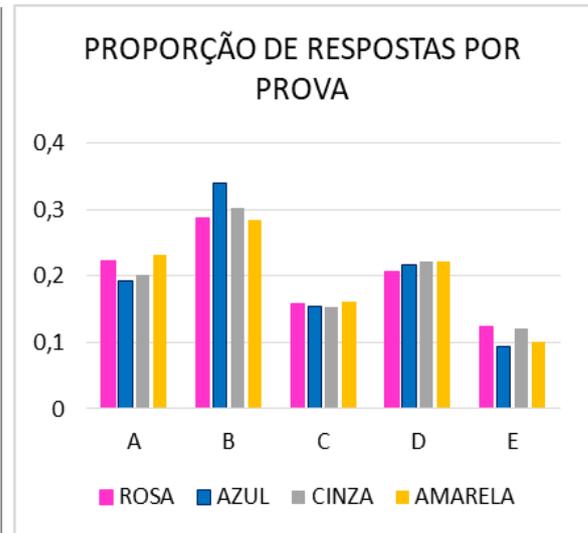
Figura 14 - Item 2

**QUESTÃO 150**

Uma empresa especializada em conservação de piscinas utiliza um produto para tratamento da água cujas especificações técnicas sugerem que seja adicionado 1,5 mL desse produto para cada 1 000 L de água da piscina. Essa empresa foi contratada para cuidar de uma piscina de base retangular, de profundidade constante igual a 1,7 m, com largura e comprimento iguais a 3 m e 5 m, respectivamente. O nível da lâmina d'água dessa piscina é mantido a 50 cm da borda da piscina.

A quantidade desse produto, em mililitro, que deve ser adicionada a essa piscina de modo a atender às suas especificações técnicas é:

- A 11,25.
- B 27,00.
- C 28,80.
- D 32,25.
- E 49,50.



MT - 2º dia | Caderno 5 - AMARELO - Página 20

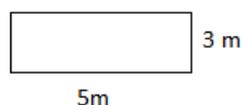
Fonte: Autor (2019)

Há várias informações que induzem ao candidato a marcar um distrator. O item busca avaliar o conhecimento geométrico do candidato com respeito a noções de área, volume e suas variações tratadas por meio de proporções. Após uma cuidadosa leitura o candidato deve identificar os objetivos do problema. A segunda etapa é verificar quais os elementos informados no texto são úteis para solucionar o problema. Esboçar um desenho que reflita o problema pode ajudar bastante. Deve-se compreender que a piscina não está completamente cheia. O volume da piscina é simples de ser calculado, por outro lado os problemas com multiplicação de números decimais podem ter afetado a solução de muitos candidatos. O volume de produto a ser colocado é calculado por proporção simples e direta. O distrator A induz o candidato ao erro quando o mesmo trocar a altura do nível da água por 0,5 m (0,5 m=distância do bordo da piscina até a lâmina d'água). O item E poderia ter se mostrado um distrator, basta que a altura da lâmina d'água seja trocada por 2,2 m e os cálculos são os mesmos para se chegar ao resultado correto.

**Solução:**

O nível da lâmina d'água será igual a  $h=1,7\text{m}-0,5\text{m}=1,2\text{ m}$ .

A área da base (A) será...



$$A=3\text{m} \cdot 5\text{m}=15\text{ m}^2$$

O volume será então...  $V=15 \text{ m}^2 * 1.2 \text{ m}=18 \text{ m}^3$  de água. É necessário saber que  $1 \text{ m}^3= 1000 \text{ l}$  (litros) de água. daí temos que  $V=18.1000=18.000 \text{ l}$  de água na piscina. Finalmente chegamos na proporção: A relação de proporção é dada por:

$$\frac{1.5 \text{ ml}}{x} = \frac{1000 \text{ l}}{18000 \text{ l}}$$

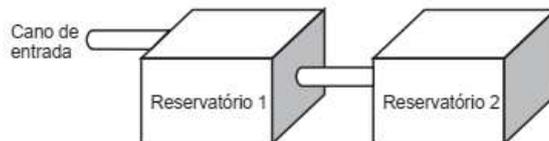
Daí concluímos que  $x=1.5 \text{ ml} * 18=27 \text{ ml}$  de produto (Resposta. B)

Pela tabela 12 (pág. 73), é possível notar que a proporção de acertos entre os candidatos que responderam a prova ROSA e a prova AMARELA é bem próxima. O possível problema ocorrido na prova ROSA se deve ao fato do item estar localizado após um item muito difícil (item 1), vir muito no início da prova no momento em que o cérebro está começando a trabalhar de forma mais intensa e necessita de um tempo para conseguir buscar na memória ideias que venham a contribuir com a resolução dos problemas da prova.

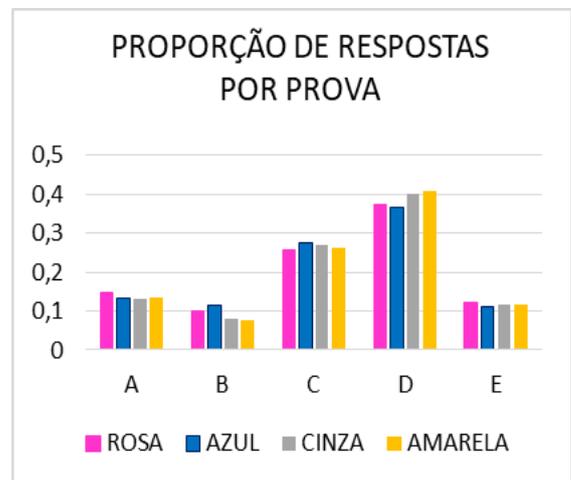
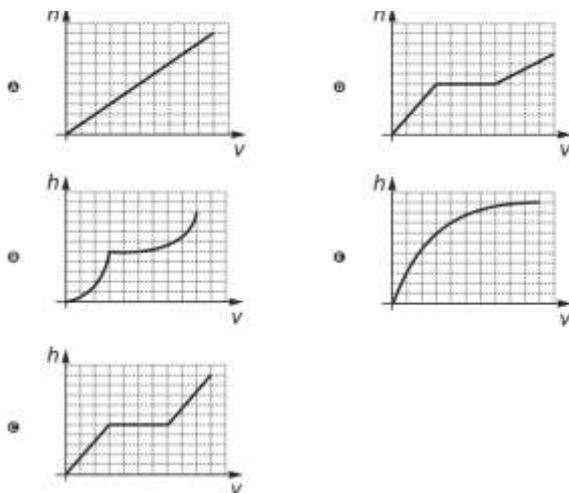
Figura 15 - Item 3

**QUESTÃO 156**

A água para o abastecimento de um prédio é armazenada em um sistema formado por dois reservatórios idênticos, em formato de bloco retangular, ligados entre si por um cano igual ao cano de entrada, conforme ilustra a figura.



A água entra no sistema pelo cano de entrada no Reservatório 1 a uma vazão constante e, ao atingir o nível do cano de ligação, passa a abastecer o Reservatório 2. Suponha que, inicialmente, os dois reservatórios estejam vazios. Qual dos gráficos melhor descreverá a altura  $h$  do nível da água no Reservatório 1, em função do volume  $V$  de água no sistema?



Fonte: Autor (2019)

O item é muito interessante, exige do candidato conhecimentos básicos sobre mecânica dos fluidos que é conteúdo abordado na disciplina de Física. Um bom raciocínio sobre construção de gráficos e aumento de volume nos reservatórios é o objetivo a ser medido pelo item. A análise feita de forma clara e com calma ajuda a não marcar o distrator que há na questão, não há cálculos a serem feitos. O item C é o distrator do item pelo fato do volume d'água crescer de forma constante no reservatório 1 antes e após o nível de água do reservatório 2 alcançar a altura de água do reservatório 1, porém os candidatos que marcaram C não perceberam que nesse instante o crescimento do volume de água no reservatório 1 seria menor (a água que cairia pelo cano de entrada iria encher os dois reservatórios ao mesmo tempo).

### **Solução:**

Até que o líquido consiga chegar a altura do cano que transfere água de um reservatório para outro teremos o reservatório 1 enchendo de forma constante (gráfico será uma reta). Após a altura da água superar a altura do cano de transferência de água teremos que o volume de água no reservatório 1 permanecerá inalterado (constante) até que a água no reservatório 2 tenha o mesmo volume de água que o reservatório 1, nesse instante os dois reservatórios serão enchidos ao mesmo tempo, daí teremos uma reta com inclinação menor que a inicial representando o aumento de volume no reservatório 1, portanto temos que o item D é o mais adequado como resposta para o item.

(Resposta. D)

Quanto mais próximo do início dos cadernos maior é a proporção de acertos, exceto para o caderno AZUL onde o item veio entre um item moderado e um item muito difícil (nesta ordem) que provavelmente tenha causado menor proporção de acertos quando comparado aos demais cadernos.

### **8.1 A Sequência de itens por nível de dificuldade**

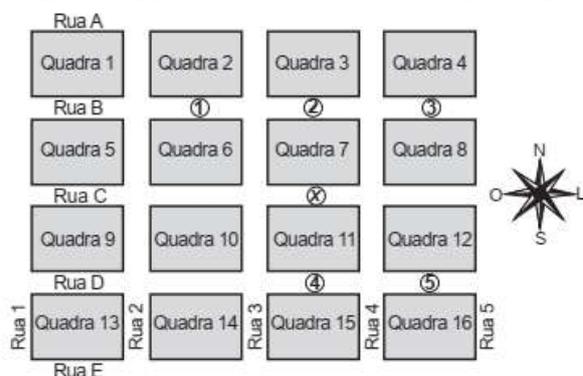
Em um todo teste de rendimento a fadiga e o esgotamento mental crescem com o tempo. No Enem estas características são facilmente observadas (candidatos se retorcendo nas carteiras escolares, com sono ou somente olhando para os itens sem conseguir resolver) nos candidatos pelo fato da prova ter 90 itens (45 de Ciências da Natureza e 45 de Matemática) com duração de no máximo 4 horas. Uma sequência de itens fáceis ou difíceis pode ser determinante na nota do candidato. Se uma sequência de itens fáceis está localizada no início

ou até na metade da prova então se espera que a taxa de acerto por item seja grande pois a fadiga ainda não afeta tanto o rendimento, porém se a localização da sequência estiver no final da prova então a taxa de certo por item esperada é menor que nas sequências com localização anteriormente comentadas. Uma sequência de itens difíceis não afeta tanto quando surge no fim da prova pois espera-se que o candidato tivera boas condições e maior tempo para resolver os itens fáceis que surgiram anteriormente beneficiando-se do critério de correção da TRI. A combinação de sequência de itens fáceis combinada com sequência de itens difíceis no fim dos cadernos do Enem e no modelo atual de correção beneficia candidatos que receberam cadernos com sequência de itens fáceis no início da prova e com o nível de dificuldade aumentando quando se aproxima dos últimos itens do exame. Uma sequência de quatro itens fáceis foi identificada no final da prova ROSA e uma sequência de itens difíceis no final da prova AMARELA. Como os cadernos do Enem possuem os mesmos itens então as duas sequências estão presentes nos quatro cadernos do exame, a localização da sequência dos itens nos cadernos e a taxa de acerto em cada item serão objetos de estudo deste tópico. A primeira sequência de itens a ser analisada é composta pelos itens (fáceis ou muito fáceis): 156. 158. 159. 162. 174 e 175 ordenados de acordo com a prova AMARELA. O item 156 já foi analisado.

Figura 16 - Item 4

#### QUESTÃO 158

Um menino acaba de se mudar para um novo bairro e deseja ir à padaria. Pediu ajuda a um amigo que lhe forneceu um mapa com pontos numerados, que representam cinco locais de interesse, entre os quais está a padaria. Além disso, o amigo passou as seguintes instruções: a partir do ponto em que você se encontra, representado pela letra X, ande para oeste, vire à direita na primeira rua que encontrar, siga em frente e vire à esquerda na próxima rua. A padaria estará logo a seguir.



A padaria está representada pelo ponto numerado com

- A 1.
- B 2.
- C 3.
- D 4.
- E 5.



Fonte: Autor (2019)

Verifica se os candidatos sabem ler e interpretar informações fornecidas além de verificar como se posicionar em um sistema de eixos coordenados. O item é muito simples, não exige cálculos mas exige atenção com relação a cada comando (a ordem de virar ou seguir em frente) que foi dado. O ponto mais importante para o acerto do item era se posicionar com relação a estrela cardeal, não há distrator.

### Solução:

Posicionando-se de acordo com a estrela cardeal o menino deve caminhar em direção as quadras 6 e 10, virar à direita no fim da quadra 7 (ele estará entre a quadra 6 e 7) caminhando em direção às quadras 2 e 3, porém deve virar à esquerda ao fim da quadra 7 e chegará ao destino.

(Resposta. A)

Nos cadernos em que o item estava mais próximo do início, houve uma maior proporção de acertos e isso é o esperado pois não há distrator e o item vem seguido de outro item fácil para todos os cadernos (ver tabela 12, pág. 73).

Figura 17 - Item 5

#### QUESTÃO 159

Três alunos, X, Y e Z, estão matriculados em um curso de inglês. Para avaliar esses alunos, o professor optou por fazer cinco provas. Para que seja aprovado nesse curso, o aluno deverá ter a média aritmética das notas das cinco provas maior ou igual a 6. Na tabela, estão dispostas as notas que cada aluno tirou em cada prova.

Aluno	1ª Prova	2ª Prova	3ª Prova	4ª Prova	5ª Prova
X	5	5	5	10	6
Y	4	9	3	9	5
Z	5	5	8	5	6

Com base nos dados da tabela e nas informações dadas, ficará(ão) reprovado(s)

- A** apenas o aluno Y.
- B** apenas o aluno Z.
- C** apenas os alunos X e Y.
- D** apenas os alunos X e Z.
- E** os alunos X, Y e Z.



Busca verificar se o candidato conhece e sabe aplicar o conceito de média aritmética simples que é um tema muito aplicado no cotidiano seja no comércio, indústria ou mesmo no campo educacional. Se o candidato tivesse problema com somas e divisões bastava que fosse distribuídas as unidades da nota que excedem o valor 6 para as notas que não alcançaram o valor 6, dessa forma o aluno verificaria que o único a não atingir média 6 seria o aluno Z. Um segundo modo de resolver é verificar se o somatório é maior ou igual a 30, se isso ocorrer o aluno será aprovado e reprovado caso contrário. O método clássico utiliza a equação que representa a média aritmética simples (soma dos elementos dividido pela quantidade de elementos). Não há distratores.

### Solução:

Seja X=conjunto de n notas de um candidato, a média aritmética simples ( $\bar{X}$ ) do conjunto X é dada por:

$$\bar{X} = \sum_{i=1}^n \frac{x_i}{n}, \text{ onde } x_i \text{ é o valor da nota } i \text{ do candidato.}$$

$$\bar{X}_X = (5+5+5+10+6)/5 = 6.2 \text{ (X foi aprovado);}$$

$$\bar{X}_Y = (4+9+3+9+5)/5 = 6.0 \text{ (Y foi aprovado);}$$

$$\bar{X}_Z = (5+5+8+5+6)/5 = 5.8 \text{ (Z não-aprovado);}$$

(Resposta. B)

O item C poderia ser pelo fato dos candidatos não perceberem que o item pedia os candidatos reprovados. Novamente temos que nos cadernos em que o item estava mais próximo do início houve uma maior taxa de acerto ao item.

Figura 18 - Item 6.

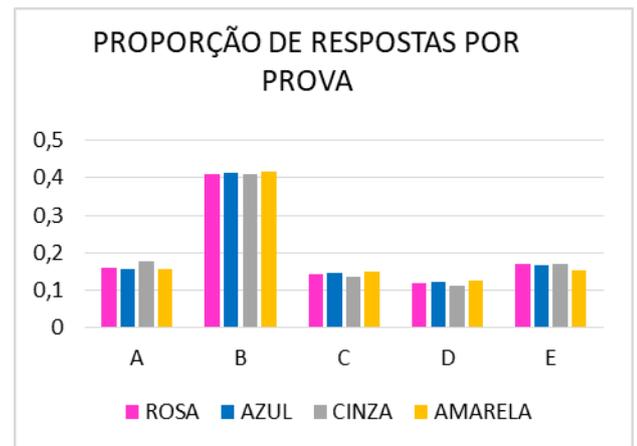
**QUESTÃO 162**

A mensagem digitada no celular, enquanto você dirige, tira a sua atenção e, por isso, deve ser evitada. Pesquisas mostram que um motorista que dirige um carro a uma velocidade constante percorre “às cegas” (isto é, sem ter visão da pista) uma distância proporcional ao tempo gasto ao olhar para o celular durante a digitação da mensagem. Considere que isso de fato aconteça. Suponha que dois motoristas (X e Y) dirigem com a mesma velocidade constante e digitam a mesma mensagem em seus celulares. Suponha, ainda, que o tempo gasto pelo motorista X olhando para seu celular enquanto digita a mensagem corresponde a 25% do tempo gasto pelo motorista Y para executar a mesma tarefa.

Disponível em: <http://g1.globo.com>. Acesso em: 21 jul. 2012 (adaptado).

A razão entre as distâncias percorridas às cegas por X e Y, nessa ordem, é igual a

- A  $\frac{4}{3}$
- B  $\frac{4}{1}$
- C  $\frac{3}{4}$



Fonte: Autor (2019)

Item simples que exige do candidato conhecimento básico sobre velocidade média, razões e proporções entre distância percorrida e tempo. Uma leitura desatenta pode levar o candidato a marcar qualquer um dos itens.

I) O candidato que não entendeu o objetivo do item pode erroneamente chegar ao item A da seguinte forma: O candidato soma 25% (de tempo) a 100% (de tempo) e chega a 125% (de tempo) percorrido. Sem pensar muito ele simplifica a razão  $125/100=5/4$ .

II) O item B é o item correto e será apresentado adiante na solução.

III) Para chegar ao item C o candidato fará a simplificação de forma semelhante ao que pode ter ocorrido com os candidatos que marcaram o item A. porém teriam diminuído  $100\%-25\%=75\%$  e fariam a simplificação da razão  $75/100=3/4$ .

IV) Para o candidato chegar ao item D ele deveria ter simplificado a fração  $100/25$ . para chegar ao item E o candidato deve ter pensado de forma semelhante aos que marcaram o item C invertendo a fração em questão apresentada.

**Solução:**

O item já afirmou que um motorista que dirige com velocidade constante percorre “às cegas” uma distância proporcional ao tempo. Podemos afirmar então que a distância percorrida pelo motorista 1 é proporcional ao tempo que o mesmo leva para digitar a mensagem e o mesmo ocorre com o motorista 2. Assim qualquer variação que ocorrer na distância percorrida terá verificada a mesma variação no tempo para qualquer dos motoristas, como o motorista 1 gasta 25% do tempo do motorista 2 para digitar a mensagem temos a equação a seguir.

$$T_1 = (25/100)T_2.$$

$T_1$  = tempo gasto para o motorista 1 digitar a mensagem;

$T_2$  = tempo gasto para o motorista 2 digitar a mensagem.

$$\text{Logo, } \frac{T_1}{T_2} = \frac{25}{100} = \frac{1}{4}.$$

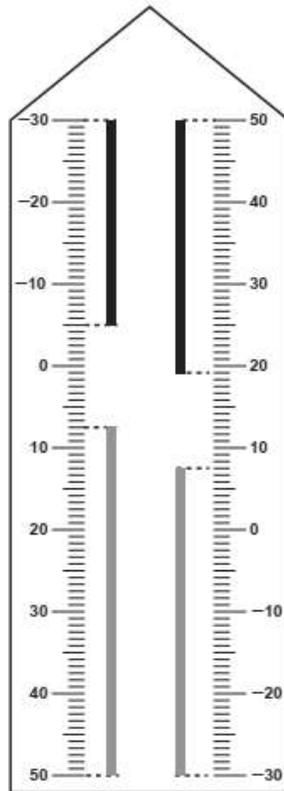
(Resposta. B)

A prova CINZA deveria apresentar uma taxa de acerto maior que a prova AMARELA para o item (posição anterior ao item na AMARELA), isso pode não ter ocorrido pelo fato do item ser antecedido por uma sequência de 3 itens muito difíceis e um moderado. Para as provas AMARELA, AZUL e ROSA ocorreu o previsto (ver tabela 12, pág. 73).

Figura 19 - Item 7

**QUESTÃO 174**

Neste modelo de termômetro, os filetes na cor preta registram as temperaturas mínima e máxima do dia anterior e os filetes na cor cinza registram a temperatura ambiente atual, ou seja, no momento da leitura do termômetro.



Por isso ele tem duas colunas. Na da esquerda, os números estão em ordem crescente, de cima para baixo, de  $-30\text{ }^{\circ}\text{C}$  até  $50\text{ }^{\circ}\text{C}$ . Na coluna da direita, os números estão ordenados de forma crescente, de baixo para cima, de  $-30\text{ }^{\circ}\text{C}$  até  $50\text{ }^{\circ}\text{C}$ .

A leitura é feita da seguinte maneira:

- a temperatura mínima é indicada pelo nível inferior do filete preto na coluna da esquerda;
- a temperatura máxima é indicada pelo nível inferior do filete preto na coluna da direita;
- a temperatura atual é indicada pelo nível superior dos filetes cinza nas duas colunas.

Disponível em: [www.if.ufrgs.br](http://www.if.ufrgs.br). Acesso em: 28 ago. 2014 (adaptado).

Qual é a temperatura máxima mais aproximada registrada nesse termômetro?

- A**  $5\text{ }^{\circ}\text{C}$
- B**  $7\text{ }^{\circ}\text{C}$
- C**  $13\text{ }^{\circ}\text{C}$
- D**  $15\text{ }^{\circ}\text{C}$
- E**  $19\text{ }^{\circ}\text{C}$

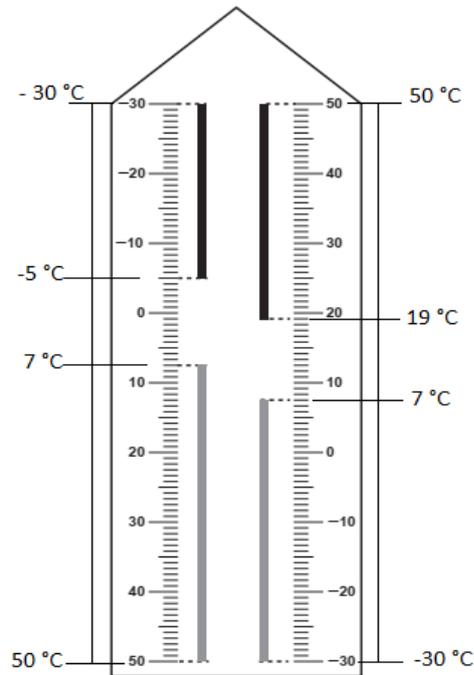


Fonte: Autor (2019)

O Item requer muita atenção e compreensão dos dados informados, a habilidade medida pelo item consiste em interpretar informações contidas em gráficos. O texto é extenso com muitas informações que são úteis para a solução do problema. O modo mais fácil e preciso de resolver o problema é indicar os valores das leituras direto na figura. As alternativas B e C são os distratores, para chegar a opção B o candidato interpretou que a temperatura máxima do dia anterior estava também no filete cinza. Para chegar ao item C a leitura no filete cinza foi feita na ordem crescente quando na verdade deveria ser feita na ordem decrescente.

Sol:

Figura 20 - TERMÔMETRO COM LEITURAS



Fonte: Enem (2019)

A temperatura mínima do dia anterior é  $-5\text{ }^{\circ}\text{C}$  (filete esquerdo, preto e inferior), a temperatura máxima do dia anterior é  $19\text{ }^{\circ}\text{C}$  (filete direito, preto e inferior) e a temperatura atual é  $7\text{ }^{\circ}\text{C}$  (filetes cinza superior esquerdo ou direito), assim a temperatura máxima mais próxima é  $19\text{ }^{\circ}\text{C}$ .

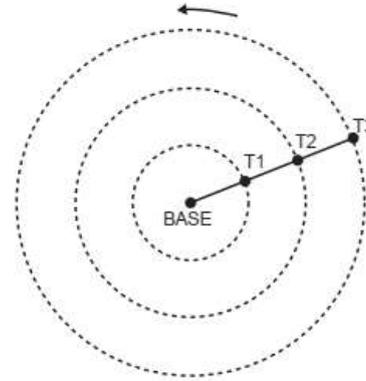
(Resposta. E)

O tempo já gasto com os itens anteriores, as dificuldades com itens difíceis, a proximidade com itens difíceis e com o fim da prova causou grande dificuldade no item que não é difícil. Apesar das proporções entre ROSA e AMARELA serem muito próximas houve uma sequência de 4 itens muito difíceis antes do item em questão na prova ROSA, na prova AMARELA os quatro últimos itens eram moderados (4) e difícil (1). A prova AMARELA obteve uma maior proporção de acertos entre seus candidatos, isso se deve ao fato de o item (na prova AMARELA) não ser antecedido de itens muito difíceis como ocorreu com as demais provas.

Figura 21 - Item 8

**QUESTÃO 175**

Pivô central é um sistema de irrigação muito usado na agricultura, em que uma área circular é projetada para receber uma estrutura suspensa. No centro dessa área, há uma tubulação vertical que transmite água através de um cano horizontal longo, apoiado em torres de sustentação, as quais giram, sobre rodas, em torno do centro do pivô, também chamado de base, conforme mostram as figuras. Cada torre move-se com velocidade constante.

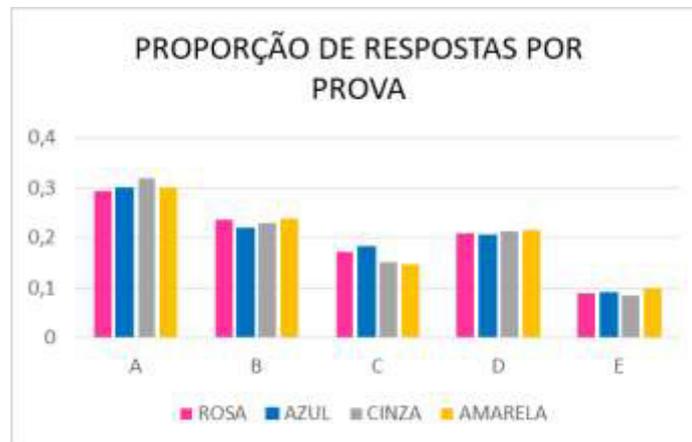


Um pivô de três torres ( $T_1$ ,  $T_2$  e  $T_3$ ) será instalado em uma fazenda, sendo que as distâncias entre torres consecutivas bem como da base à torre  $T_1$  são iguais a 50 m. O fazendeiro pretende ajustar as velocidades das torres, de tal forma que o pivô efetue uma volta completa em 25 horas. Use 3 como aproximação para  $\pi$ .

Para atingir seu objetivo, as velocidades das torres  $T_1$ ,  $T_2$  e  $T_3$  devem ser, em metro por hora, de

- A 12, 24 e 36.
- B 6, 12 e 18.
- C 2, 4 e 6.
- D 300, 1 200 e 2 700.
- E 600, 2 400 e 5 400.

MT - 2º dia | Caderno 5 - AMARELO - Página 29

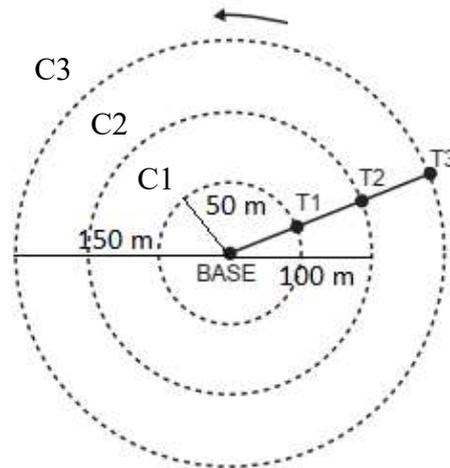


Fonte: Autor (2019)

O item traz um longo texto com muitas informações necessárias à solução do problema. Neste item o candidato deve ter conhecimento sobre comprimento do círculo e sobre velocidade média. A percepção que se espera do candidato é que ele perceba que o raio para os três círculos corresponde ao espaço da Base do pivô para cada um dos pontos  $T_1$ ,  $T_2$  e  $T_3$  respectivamente. Não há distrator.

**Sol:**

Figura 22 - Caminho do pivô



Fonte: Enem adaptada (2019)

Se o candidato marcar o comprimento dos raios direto na figura então o problema pode se tornar mais simples.

O raio da Base a T1 (R1) tem valor igual a 50 m;

O raio da Base a T2 (R2) tem valor igual a 100 m;

O raio da Base a T3 (R3) tem valor igual a 150 m;

O candidato deve reconhecer que o comprimento de cada círculo (espaço a ser percorrido pelos pontos T1, T2 e T3) é calculado através da equação  $C=2\pi R$ ;

$C$ =comprimento do círculo;  $\pi=3$  (neste problema);  $R$ = raio de cada círculo;

O comprimento de cada círculo deve ser calculado:

$$C1=2\pi \cdot 50\text{m} = 2 \cdot 3 \cdot 50\text{ m} = 300\text{ m};$$

$$C2=2\pi \cdot 100\text{m} = 2 \cdot 3 \cdot 100\text{ m} = 600\text{ m};$$

$$C3=2\pi \cdot 150\text{m} = 2 \cdot 3 \cdot 150\text{ m} = 900\text{ m};$$

A velocidade média linear em T1, T2 e T3 é o objetivo do item. A equação da velocidade média pode ter sido um problema para muitos candidatos uma vez que tal conceito é visto principalmente na disciplina de Física. A equação da velocidade média corresponde a razão entre a variação de espaço e a variação de tempo nesta ordem.

$$V = \frac{\Delta E}{\Delta t};$$

$V$ = velocidade média;

$\Delta E$ = variação de espaço;

$\Delta t$  = variação de tempo;

O problema informou que o tempo para que o pivô efetue uma volta completa é de 25 h (25 horas).

$V_1 = C_1 / 25 \text{ h} = 300 \text{ m} / 25 \text{ h} = 12 \text{ m/h}$  (12 metros/ hora);

$V_2 = C_2 / 25 \text{ h} = 600 \text{ m} / 25 \text{ h} = 24 \text{ m/h}$  (24 metros/ hora);

$V_3 = C_3 / 25 \text{ h} = 900 \text{ m} / 25 \text{ h} = 36 \text{ m/h}$  (36 metros/ hora);

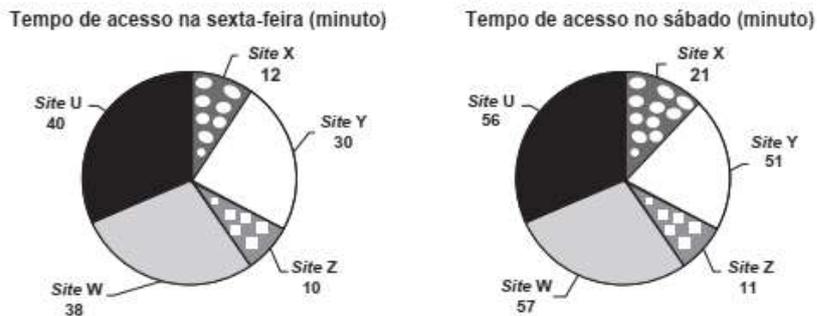
Resposta A.

O item pode se tornar simples se o candidato conhecer o conceito de velocidade média, comprimento do círculo e saber calcular o valor de cada objetivo do item. A prova CINZA apresentou maior proporção de acerto ao item que as demais sendo que nessa prova a localização do item está mais próxima da metade da prova. por outro lado nas demais provas o item estava localizado próximo aos últimos itens dos cadernos. A prova AMARELA obteve proporção de acerto (ver tabela 12, pág.73) quase idêntica as proporções das provas ROSA e AZUL apesar de estar localizada mais próxima ao fim da prova. O item 7 vem antecedido de 3 itens muito difíceis e um fácil na prova ROSA, na prova AZUL vem antecedido de 2 muito difíceis, um moderado e um fácil, na prova AMARELA vem depois de 2 itens moderados, um difícil e um fácil.

Figura 23 - Item 9

**QUESTÃO 177**

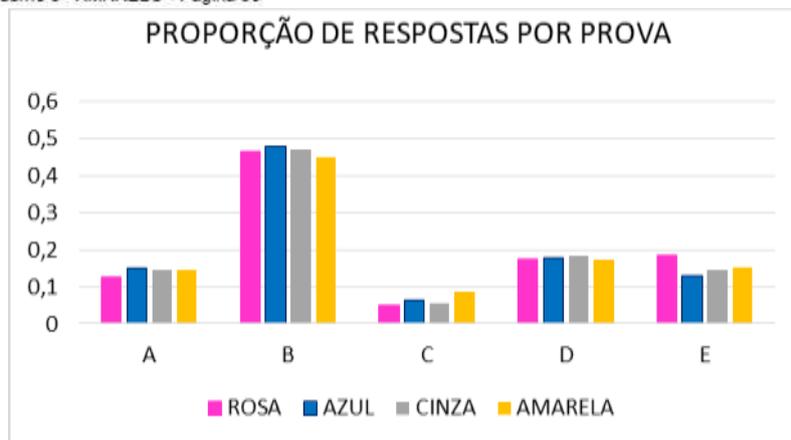
Quanto tempo você fica conectado à internet? Para responder a essa pergunta foi criado um miniaplicativo de computador que roda na área de trabalho, para gerar automaticamente um gráfico de setores, mapeando o tempo que uma pessoa acessa cinco sites visitados. Em um computador, foi observado que houve um aumento significativo do tempo de acesso da sexta-feira para o sábado, nos cinco sites mais acessados. A seguir, temos os dados do miniaplicativo para esses dias.



Analisando os gráficos do computador, a maior taxa de aumento no tempo de acesso, da sexta-feira para o sábado, foi no site

- A** X.
- B** Y.
- C** Z.
- D** W.
- E** U.

MT - 2º dia | Caderno 5 - AMARELO - Página 30



Fonte: Autor (2019)

O item verifica o conhecimento do candidato sobre o conceito de taxa de variação. Apesar do cálculo envolvido ser simples houve uma proximidade com itens difíceis fazendo com que a real dificuldade do item fosse modificada, além disso o esgotamento físico e mental contribuiu para uma taxa de acerto tão baixa. Através do gráfico de proporção de acertos acima verifica-se que poucos candidatos marcaram o item C. Isso se deve ao fato dos candidatos perceberem que o aumento foi de apenas 1 unidade na área correspondente ao site Z. esse valor é muito inferior aos aumentos ocorridos nas demais áreas. Os candidatos poderiam ter calculado o dobro do valor de cada área inicial e verificar quantas unidades faltam (de acordo com o gráfico de pizza mais à direita) no segundo gráfico para se chegar ao

dobro. ao que faltasse a menor quantidade de unidades seria o de maior taxa de aumento. A opção B é o distrator do item pois nela o aumento é significativo unido ao tamanho da figura dá impressão de um aumento maior que nas outras áreas. O aumento ocorrido na área do site X visualmente está distante do dobro da área inicial reforçando ainda mais a opção pelo item B.

**Solução:**

A dificuldade era lembrar do conceito de taxa de variação, como efetuar o cálculo e como interpretar o resultado.

$$X=21/12=1,75 ;$$

$$Y=51/30=1,70 ;$$

$$Z=11/10=1,10;$$

$$W=58/37=1,567... ;$$

$$U=56/40= 1,40.$$

O site X teve maior taxa de aumento de acordo com as áreas informadas (aumento de 75%).

(Resposta. A)

Note que a posição não interfere tanto na proporção de acertos quando o item é muito difícil. Pensando dessa forma podemos inferir que com relação a este item as provas CINZA e AMARELA foram beneficiadas pois o item está mais próximo do fim dos cadernos.

Figura 24 - Item 10

## QUESTÃO 178

O resultado de uma pesquisa eleitoral, sobre a preferência dos eleitores em relação a dois candidatos, foi representado por meio do Gráfico 1.

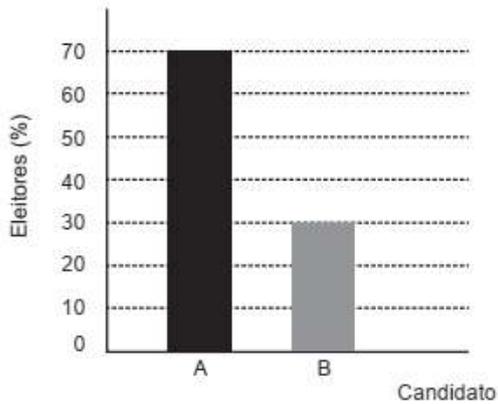


Gráfico 1

Ao ser divulgado esse resultado em jornal, o Gráfico 1 foi cortado durante a diagramação, como mostra o Gráfico 2.

A diferença entre as razões da altura da coluna B pela coluna A nos gráficos 1 e 2 é

- A 0
- B  $\frac{1}{2}$
- C  $\frac{1}{5}$
- D  $\frac{2}{15}$
- E  $\frac{8}{35}$

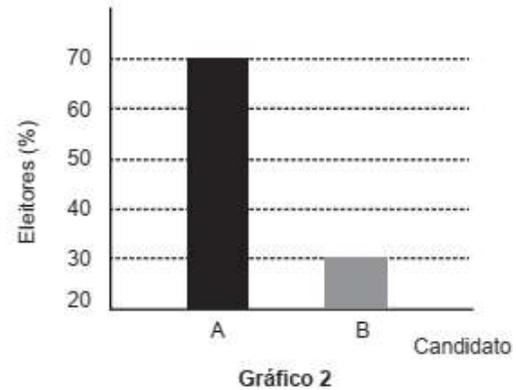
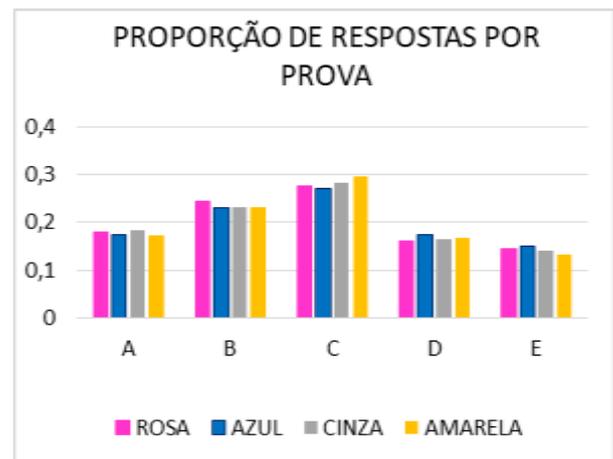


Gráfico 2

Apesar de os valores apresentados estarem corretos e a largura das colunas ser a mesma, muitos leitores criticaram o formato do Gráfico 2 impresso no jornal, alegando que houve prejuízo visual para o candidato B.



Fonte: Autor (2019)

O item avalia a capacidade do candidato em interpretar gráficos e realizar operações que envolvem razões e proporções. O tempo de prova, ansiedade, fadiga e cansaço mental podem ter prejudicado o desempenho de milhares de candidatos pois o item está localizado após o 29º item em todos os cadernos. O problema seria interpretar o gráfico e compreender o objetivo do item, poderiam ter explicado melhor a pergunta (verifique a razão entre a coluna B e a coluna A em cada gráfico e depois calcule a diferença). Uma das formas para se chegar a opção A (incorreto) pode ter sido observando a diferença de altura (em porcentagem) entre as colunas nos gráficos 1 e 2. No gráfico 1 se observarmos a diferença é de 40% da altura da coluna A contra a coluna B. ocorre o mesmo valor de diferença no gráfico 2, logo o candidato (despreparado para responder ao item no momento) deve ter calculado a diferença (40%-

40%=0) e optado pela opção A. Os candidatos que marcaram B podem ter pensado da seguinte forma: Se  $50\%=1/2$ , então retirando 20% do gráfico preto e justapondo ao cinza chegaremos a marca de 50% nos dois gráficos. Os candidatos que optaram pelo item B como resposta verificaram que houve uma redução de 20% (pensamento dos candidatos), daí lembraram que  $20\%=20/100=1/5$ . Os candidatos que optaram pela opção C pensaram que em cada coluna houve redução de 20% (altura da coluna), daí  $20\%=20/100=1/5$ .

### Solução:

No gráfico 1 temos que a altura na coluna B corresponde a uma variação de 30% e na coluna A de 70%, logo a razão será 30/70. No gráfico 2 a altura da coluna B corresponde a uma variação de 10% e na coluna A de 50%, logo teremos 10/50. Neste momento deve ser calculado a diferença que é dada por  $(30/70)-(10/50)=(3/7)-(1/5)=8/35$

(Resposta. E)

Item de interpretação complicada, note que a diferença na proporção de acertos entre as provas CINZA e AMARELA foi muito pequena apesar da distância para o fim dos cadernos (item mais próximos do fim da prova tem menor taxa de acerto).

Figura 25 - Item 11

#### QUESTÃO 179

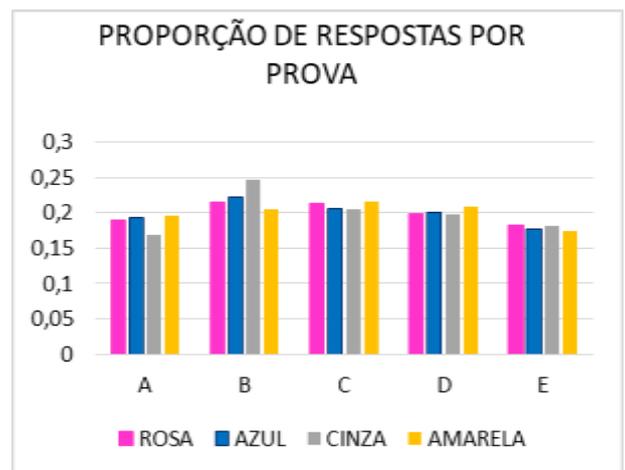
Um cientista, em seus estudos para modelar a pressão arterial de uma pessoa, utiliza uma função do tipo  $P(t) = A + B\cos(kt)$  em que  $A$ ,  $B$  e  $K$  são constantes reais positivas e  $t$  representa a variável tempo, medida em segundo. Considere que um batimento cardíaco representa o intervalo de tempo entre duas sucessivas pressões máximas.

Ao analisar um caso específico, o cientista obteve os dados:

Pressão mínima	78
Pressão máxima	120
Número de batimentos cardíacos por minuto	90

A função  $P(t)$  obtida, por este cientista, ao analisar o caso específico foi

- A  $P(t) = 99 + 21\cos(3\pi t)$
- B  $P(t) = 78 + 42\cos(3\pi t)$
- C  $P(t) = 99 + 21\cos(2\pi t)$
- D  $P(t) = 99 + 21\cos(t)$
- E  $P(t) = 78 + 42\cos(t)$

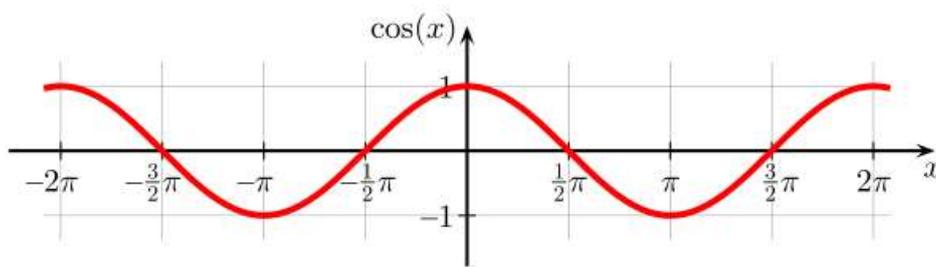


O item é complicado, questões envolvendo funções trigonométricas costumam ser “uma pedra no sapato” dos candidatos. O candidato deve saber valores limites das funções seno e cosseno, período da função cosseno, resolver sistemas com 2 equações e 2 incógnitas e relacionar quantidade de batimentos com tempo. Este item é muito difícil independente da posição em que está localizado nos cadernos, por outro lado se o item está localizado próximo ao fim da prova ele beneficia os candidatos que receberam tal caderno (CINZA e ROSA).

### Solução:

O candidato deve conhecer os valores máximo, mínimo e período da função cosseno.

Figura 26 - GRÁFICO DA FUNÇÃO COSSENO



Fonte: desconhecido (Internet)

Pelo gráfico acima podemos perceber que os valores limite para a função cosseno são 1 e -1 e o período da função é  $2\pi$  radianos.

O valor máximo da função será  $P_{\text{máx}}(t)$  e ocorre quando  $\cos(kt)$  for igual a 1. De forma análoga o valor mínimo  $P_{\text{mín}}(t)$  ocorre quando  $\cos(kt)$  for igual a -1. A tabela informa que a pressão máxima vale 120 e a mínima 78, substituindo os valores na equação teremos que o maior valor do  $\cos(kt)=1$  e o menor é  $\cos(kt)=-1$ ;

$$P_{\text{máx}}(t)=A + B.\cos(kt)$$

$$P_{\text{mín}}(t)=A + B.\cos(kt)$$

$$120= A+ B.(1)$$

$$78=A +B.(-1)$$

$$\text{Logo: } \begin{cases} 120 = A + B \\ 78 = A - B \end{cases}$$

Se somarmos as duas equações teremos  $198=2A$ . daí  $A=198/2$ , portanto  $A=99$ . Com a informação já poderiam ser excluídos os itens C e E como possíveis respostas, assim ampliando as chances de sucesso do candidato no item. Em qualquer uma das equações pode ser obtido o valor de B.

$120=A+ B$ . como o valor de  $A=99$  teremos  $120=99+B$ . subtraindo 99 em ambos os lados da

igualdade chegaremos ao valor de B ( $B=120-99= 21$ ). A equação então é  $P(t)= 99+ 21\cos(kt)$ , o problema agora consiste em determinar o valor de k. O texto informou que um batimento cardíaco equivale ao tempo de batimento entre duas pressões máximas. Sabemos que há 90 batimentos em 1 minuto e montamos a proporção;

batimentos                      tempo

90                                      1 minuto=60 segundos

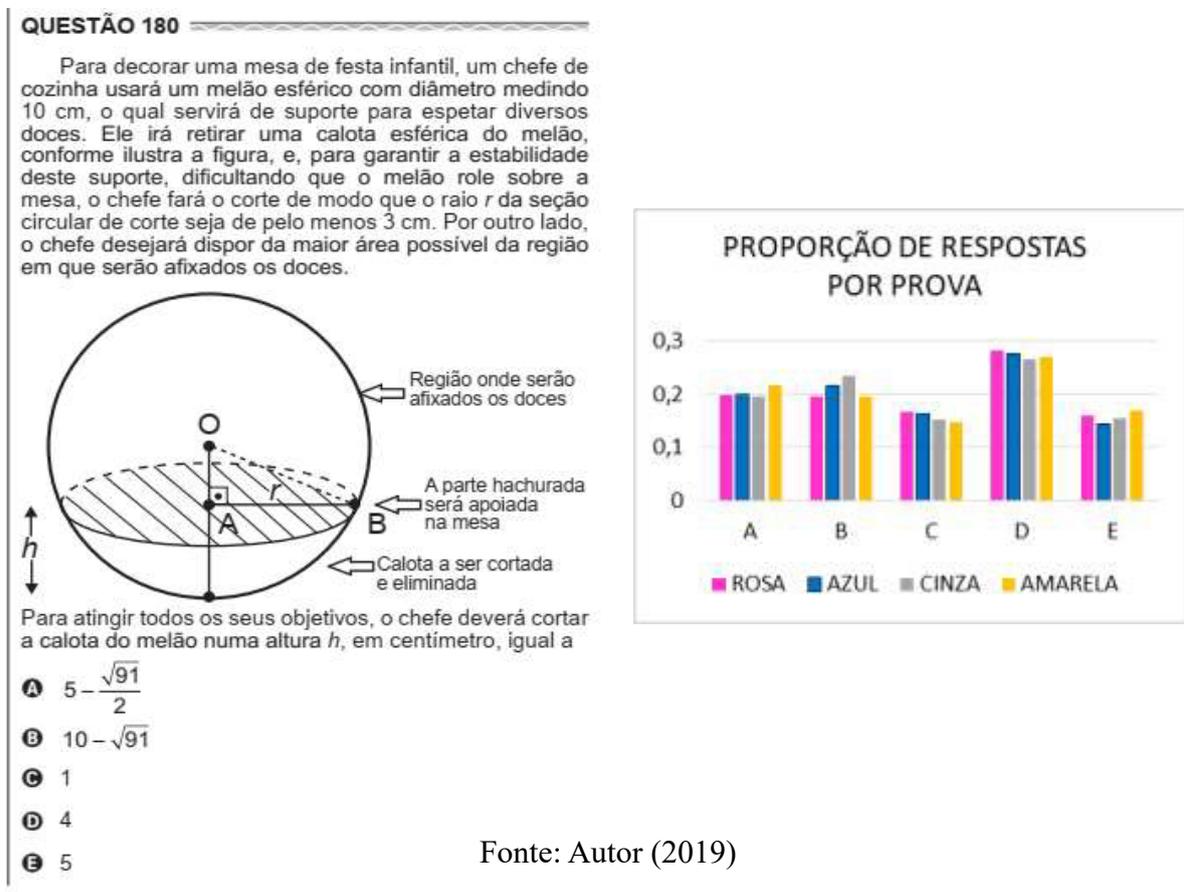
1                                        x (segundos). As grandezas são diretamente proporcionais.

$$(90/1)=(60 \text{ s}/t). t=(2/3) \text{ s.}$$

A última parte do problema trata de descobrir o valor de k. sabendo que o argumento deve valer  $2\pi$ , teremos  $kt= 2\pi$ . daí  $k.2/3=2\pi$ , simplificando por 2 teremos  $k/3=\pi$ , multiplicando ambos os lados por 3 chegamos ao valor de k ( $k=3\pi$ ), finalmente a equação toma a forma  $P(t)=99+21\cos(3\pi t)$ . (Resposta. A)

Os itens difíceis não sofrem tanta influência da posição quando se trata de proporção de acertos, por outro lado a localização no fim dos cadernos corresponde a um maior tempo para resolver os itens mais fáceis que fazem grande diferença na correção pela TRI.

Figura 27 - Item 12

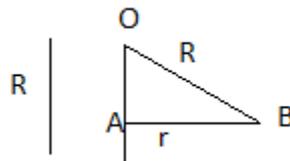


O item verifica o conhecimento espacial do candidato sobre esferas, círculos, áreas e Teorema de Pitágoras. O item possui o viés da figura que quando associado a localização próxima do fim dos cadernos afeta o desempenho do candidato. Outros fatores como a fadiga, o cansaço mental e tempo são fatores que dificultam a análise do item. O item D é o distrator do item pelo fato da figura dar a impressão de que o tamanho  $h$  seja muito maior que o tamanho da semirreta  $OA$ . Os candidatos podem ter calculado o valor de  $OA$  e já o tomaram como resposta. O fator fadiga pode ser evidenciado pela alta taxa de indicação de resposta ao item E que equivale exatamente ao tamanho do raio da esfera. A falta de atenção com relação ao objetivo do item combinado com o viés da imagem levam a conclusões incorretas.

**Solução:**

O tamanho  $OB$  é exatamente o raio da esfera ( $R=5$  cm) e se  $r=3$  cm, temos então um terno Pitagórico no triângulo  $AOB$ , sendo retângulo em  $A$ .

Figura 28: TRIÂNGULO RETÂNGULO



Fonte: Autor (2019)

Pelo Teorema de Pitágoras temos que  $R^2 = r^2 + OA^2$ . substituindo seus valores teremos:

$5^2 = 3^2 + OA^2$ . logo  $AO=4$  cm. Fazendo a subtração  $R-OA = (5-4)\text{cm}=1$  cm.

(Resposta. C)

Note que a posição do item nas provas ROSA e AZUL (ver tabela 12, pág. 73) não gerou grandes diferenças pelo fato do item não ser fácil.

## 9 CONCLUSÃO

Um bom desempenho no Enem é o grande sonho de muitos alunos do ensino médio. A importância de se ter cadernos diferentes é fundamental para que se evite minimizar possíveis irregularidades (cola) no processo avaliativo, porém os diferentes cadernos devem promover a equidade do exame fazendo com que a diferença de cor e ordem dos itens nos cadernos não se transforme em uma diferença significativa de notas prejudicando a validade do exame.

Conforme a Análise das notas há diferença de proficiência média entre os candidatos que responderam aos cadernos ROSA, AZUL, CINZA e AMARELO. O teste ANOVA confirmou a diferença de proficiência entre candidatos que responderam cadernos diferentes. O teste DMS de Fischer mostrou que as diferenças somente não se confirmam entre candidatos que responderam as provas CINZA e AMARELA (as distribuições dos itens são mais próximas). Entre as provas a maior diferença de pontos observada ocorreu entre as provas AMARELA e ROSA com mais de 6 pontos a favor da prova AMARELA.

O efeito da diferença foi confirmado pelo teste DIF que apontou o efeito diferencial em todas as provas indicando falta de equidade (ANDRIOLA, 2002). O erro de itens fáceis nos cadernos foram determinantes para uma queda significativa na proficiência visto que a correção utiliza a TRI.

Apesar de uma maior quantidade de itens respondidos de forma correta nas provas ROSA quando comparada com a prova AMARELA, houve registro de maior proficiência média dos candidatos que responderam a prova AMARELA, o mesmo fato ocorreu entre as provas AZUL e CINZA nessa ordem.

A sequência de itens fáceis surgida no final do caderno ROSA contribuiu para uma menor proficiência quando comparada aos demais cadernos. Os itens 37, 38, 40, 41, 42 e 45 no caderno ROSA são idênticos aos itens 39, 40, 23, 24, 21 e 27 nesta ordem, são todos itens fáceis ou muito fáceis que apresentaram taxa de acerto favorável ao caderno AMARELO em 5 dos 6 itens quando comparado ao caderno ROSA. A sequência de itens 42, 43, 44, 45 no caderno AMARELO corresponde a sequência dos itens 33, 34, 35 e 36 no caderno ROSA nesta ordem, onde foi verificado maior taxa de acerto favorável ao caderno ROSA em 2 itens de um total de 4 itens, isso reforça que a posição de itens difíceis não definem melhores resultados de proficiência em função da posição nos cadernos corroborando com o tipo de correção utilizado pela TRI. A perda de itens difíceis não traz redução tão significativa na nota quanto a perda de itens fáceis, isso ficou comprovado pela proporção de acertos por item que

foi maior para o caderno ROSA quando comparado ao caderno AMARELO. “Os efeitos de localização de itens distorcem a estimativa das dificuldades do item, eles invalidam as conclusões sobre as propriedades psicométricas de um item ou o nível de desempenho de um respondente (MENDES et al, 2016). Os candidatos que receberam o caderno AMARELO tem em média 6 pontos de vantagem contra candidatos que receberam o caderno ROSA.

A análise pedagógica mostrou que itens fáceis em alguns cadernos foram classificados como moderados ou difíceis em outros cadernos por estarem localizados próximos ao fim ou mesmo por serem muito próximos a sequências de itens com maior dificuldade. Os longos textos confundem e levam o candidato à exaustão. Os distratores são essenciais em todo teste mas o tempo, falta de estratégia de solução de itens, a ansiedade, o esgotamento físico e o esgotamento mental podem distorcer a verdadeira proficiência do candidato.

O presente estudo é de extrema relevância para que as autoridades do campo educacional, gestão de escolas, população e para os candidatos que participaram e participarão de novas edições do exame percebam que há dificuldades distintas entre os cadernos. Tal problema deve ser estudado e verificado em edições posteriores para buscar soluções que minimizem ou acabem com o problema.

Uma possível solução para o problema seria organizar o exame em cadernos onde os itens seriam organizados em 3 sequências de itens pré-testadas de acordo com seu nível de dificuldade (fácil, moderado e difícil). As três sequências podem ser arranjadas de 6 modos distintos onde a quantidade de itens poderia ser de 12 itens fáceis, 21 itens moderados e 12 difíceis, dessa forma o candidato ao identificar a dificuldade da sequência de itens fáceis poderia resolvê-los primeiro, depois resolver os itens da sequência moderada e finalmente a sequência de itens difíceis, assim o candidato ganharia tempo, aumentaria sua proficiência no teste e o exame ganharia em equidade.

Outra forma de tornar o exame mais justo é corrigindo os gabaritos de acordo com a cor do caderno utilizando um processo de normalização, assim todos os cadernos teriam as mesmas médias e os mesmos desvios padrões.

## REFERÊNCIAS

AFONSO, Almerindo Janela. Avaliar a escola e a gestão escolar: elementos para uma reflexão crítica. *In*: ESTEBAN. Maria Teresa (org.). **Escola, currículo e avaliação**. São Paulo: Cortez, 2005. (Série Cultura, Memória e Currículo ; v. 5)

ALBANO, A. D. Multilevel modeling of item position e effects. **Journal of Educational Measurement**, v. 50, n. 4, 2014.

ALEXANDROWICZ, R.; MATSCHINGER, H. Estimation of item location e effects by means of the generalized logistic regression model: a simulation study and an application. **Psychology Science Quarterly**, v. 50, 2008.

ALMEIDA, A. C. O efeito do contexto e posição da pergunta no questionário sobre o resultado da medição. **Opin. Publica**, v.8, n.2, p.328-339, 2002.

ANDRADE, D. F. de ;TAVARES, H. R. ;VALLE.R. da C. **Teoria de resposta ao item: conceitos e aplicações**. São Paulo: ABE–Associação Brasileira de Estatística, 2000.

ANDRIOLA, W. B. Avaliação da aprendizagem: uma análise descritiva segundo a teoria de resposta ao item (TRI). **Revista Educação em Debate**, v. 20, nº 36, p. 93-102,1998.

\_\_\_\_\_. Descrição dos principais métodos para detectar o funcionamento diferencial dos itens DIF. **Psicologia: Reflexão e Crítica**, v. 14, n. 3, p. 643-652, 2001.

\_\_\_\_\_. **Detección del funcionamiento diferencial del ítem(DIF) em tests de rendimento**. Tese (doutorado em Educação), Facultad de Educación, Universidad Complutense de Madrid, Madrid, 2002.

\_\_\_\_\_. Estudos sobre o viés do item em testes de rendimento: uma restrospectiva . **Estudos em Avaliação Educacional**, v. 17, n. 35, set./dez., 2006.

\_\_\_\_\_. **Funcionamento diferencial do item (DIF)**: indicador de justiça das avaliações em larga escala. (Tese professor titular)- Faculdade de Educação, Departamento de Fundamentos da Educação, Universidade Federal do Ceará, Fortaleza , 2018.

ANASTASI, A. **Testes psicológicos**. São Paulo, EPU, 1977.

ANGOFF, W, H . Perspectivas sobre a metodologia de funcionamento de itens diferenciais. *In*: HOLLAND, P. W. ; WAINER, H.(eds.). **Funcionamento diferencial de itens**. Hilldsdale: Lawrence Erlbaum, 1993. p. 3-23.

BAKER, F. B. **The basics of item response theory**. College Park, MD,:ERIC Cleanringhouse on Assessment and Evaluation, 2001.Disponível em <https://eric.ed.gov/?id=ED458219>. Acesso em: 13out. 2019.

BECKER. Fernanda da Rosa. **Avaliação educacional em larga escala**: a experiência brasileira. 2010. Disponível em: <https://rieoei.org/RIE/article/view/1751>. Acesso em: 08 set. 2019.

BERTAGA, R . H . ; MELLO , L . R . ; POLATO , A . **Política e avaliação educacional**. Disponível em : <http://dx.doi.org/10.14244/19827199904>. Acesso em: 03 jun.2019.

BOCK, A. M. B. Psicologia da educação: cumplicidade ideológica. *In*: MEIRA, M. E. M.; ANTUNES, M. A. M. (orgs.). **Psicologia escolar**: teorias críticas São Paulo: Casa do Psicólogo, 2003. p.79-103.

BOECK, P. Random item IRT models. **Psychometrika**, 73, p.533-559, 2008. Disponível em: <https://doi.org/10.1007/s11336-008-9092-x>. Acesso em: 08 set. 2009.

BRASIL. **Constituição da República Federativa do Brasil**. Artigos 205 e 210. Das providências e diretrizes à Educação, Cultura e Desporto.

BRASIL. **LDB**: Lei de diretrizes e bases da educação nacional (Lei n. 9394/96). Disponível em: [http://www.planalto.gov.br/ccivil\\_03/leis/19394.htm](http://www.planalto.gov.br/ccivil_03/leis/19394.htm). Acesso em: 08 set. 1999.

BUSSAB, W. O.; MORETTIN, P. A. **Estatística básica**. São Paulo: Saraiva, 2006.

CAMILLI, G. ; SHEPARD, L. A. **Methods for identifying biased test items**. Thousand Oaks, Califórnia, : Sage Publications, 1994.

CARLOS, A. A. **O efeito do contexto e posição da pergunta no questionário sobre o resultado da medição**; Universidade Federal Fluminense. OPINIÃO PÚBLICA, Campinas. Vol. VIII. nº2, p.328-339, ano 2004.

CENTRO DE INFORMAÇÃO DAS NAÇÕES UNIDAS PARA O BRASIL (UNIC). **ENEM 2010** : nota do coordenador residente do sistema das nações Unidas no Brasil. Brasília, 2010. Disponível em: <https://unicrio.org.br/onubrasil/coordenador-residente-da-onu-no-brasil/>. Acesso em : 08 set. 2019.

CERDÁ, E. **Psicometria general**. 2. ed. [S.l.]: Editorial Herder, 1978. 190 p.

COLE, N. S. History and development of DIF. *In* : HOLLAND, P. W.; WAINER, H.(e d s ) . **Differential item functioning** . New Jersey : Lawrence Erlbaum Associates, 1993. p. 25- 29.

COLE, N. S.; MOSS, P. A. Bias in test use. *In*: LINN, R. L. (ed.). **Educational Measurement**. New York: McMillan, 1989. p. 201-219.

COUTO, G.; PRIMI, R. **Teoria de resposta ao item (TRI)**: conceitos elementares dos modelos para itens dicotômicos. *Boletim de Psicologia*, São Paulo, v. 61, n. 134: p.1-15, jun. 2011.

CRONBACH, L. J. **Designing evaluation of educational and social programs**. San Francisco : Jossey Bass, 1982.

\_\_\_\_\_. Course improvement through evaluation. **Teachers College Record**, New York, v. 64n. 8. p. 672-683, May 1963.

CUNHA, O. E. **Ações afirmativas**: o princípio constitucional da igualdade e das cotas raciais.

Trabalho de Conclusão de Curso (Graduação em Direito)- Universidade Federal de Uberlândia, Uberlândia, 2017.

DIFERENÇA entre nota de corte com e sem cota chega a 180 pontos no Sisu.G1, Rio Grande do Sul, 2017. Disponível em em:<<http://g1.globo.com/rs/rio-grande-dosul/noticia/2014/03/diferencaentre-nota-de-corte-com-e-sem-cota-chega-180-pontosno-sisu.html>>. Acesso em: 12 dez.2019.

EBEL, R. L. **Producers for the analysis of classroom tests**. [S. l.] :State University of Iowa, 1954.

ERTHAL, T. C. **Manual de psicometria**. 6. ed. Rio de Janeiro: Jorge Zahar, 2001.

EVARISTO. O. C. **Ações afirmativas: O princípio constitucional da igualdade e das cotas raciais**. TCC- Universidade Federal de Uberlândia, 2017.

FIDALGO, A. M. Mantel-Haenszel methods. *In* : EVERITT, B. S. ; HOWELL, D. C. (eds.) **Encyclopedia of Statistics in Behavioral Science**. Chichester, UK: John Wiley & Sons, 2005.p.1120-1126. v. 3.

FIDALGO, A. M. ; FERRERES, D. Supuestos y consideraciones en los estudios empíricos sobre el funcionamiento diferencial de los items. **Psicothema**, v.14, p. 491-496, 2002.

FIDALGO, A. M. ; SCALON, J. D. Uso dos métodos Mantel-Haenszel para detecção do funcionamento diferencial dos itens e Software relacionado. **Psicologia: Reflexão e Crítica**, Porto Alegre,v. 25, n. 1, 2012.

GATTI, B. A. **A construção da pesquisa em educação no Brasil**. [S. l.] :Editora Plano. 2002.

GOMES. P. A educação é a arma mais poderosa que você pode usar para mudar o mundo – Nelson Mandela. **Revista, Prosa, Verso e Arte**. Rio de Janeiro, 2019. Disponível em : <https://www.revistaprosaversoarte.com/a-educacao-e-a-arma-maispoderosa-que-voce-pode-usarpara-mudar-o-mundo-nelson-mandela>. Acesso em: 08 dez. 2019.

HAMBLETON, R. K. Perspectivas futuras y aplicaciones. *In*: MUÑIZ, J. **Introducción a la teoría de respuesta a los items**. Madrid: Ediciones Psicología Pirámide, 1997.

IOETA, P. A. Enem 2019: 'Começar pelo exercício mais fácil' e outros erros comuns de alunos em provas, segundo a ciência. **BBC News Brasil**, São Paulo, 26 out. 2019. Disponível em: <https://www.bbc.com/portuguese/geral-49411981>. Acesso em: 26 out.2019.

INSTITUTO NACIONAL DE ESTUDOS E PESQUISAS EDUCACIONAIS ANÍSIO TEIXEIRA. **Documento Básico do Enem**. Brasília, 2002.

\_\_\_\_\_. **Exame nacional do ensino médio(ENEM)**: fundamentação teórico-metodológica. Brasília : MEC/INEP, 2005.

\_\_\_\_\_. **Exame nacional do ensino médio**: interpretação pedagógica das escalas de proficiência no exame nacional do ensino médio-Enem. Brasília: MEC/INEP, 2011.

\_\_\_\_\_. **Nota técnica**. Brasília, MEC/INEP, 2011.

\_\_\_\_\_. **Enem**: documento básico. Brasília - DF: [s.n.], 2018. Disponível em:<http://portal.inep.gov.br/documents/186968/484421/Exame+Nacional+do+Ensino+Médio++ENEM++documento+básico/e2cf61a8-fd80-45b8-a36f-af6940e56113?> Acesso em: 10 set. 2018.

\_\_\_\_\_. **Enem 2017 tem 7,6 milhões de inscritos**. Disponível em: [http://portal.inep.gov.br/artigo/-/asset\\_publisher/B4AQV9zFY7Bv/content/enem-2017-tem-7-6-milhoes-de-inscritos/21206](http://portal.inep.gov.br/artigo/-/asset_publisher/B4AQV9zFY7Bv/content/enem-2017-tem-7-6-milhoes-de-inscritos/21206). Acesso em 2017.

IOETA, P. A. Enem 2019: 'Começar pelo exercício mais fácil' e outros erros comuns de alunos em provas, segundo a ciência. **BBC News Brasil**, São Paulo, 26 out. 2019. Disponível em: <https://www.bbc.com/portuguese/geral-49411981>. Acesso em: 26 out. 2019.

KENNETH, D. R. The Impact of item sequence order on local item dependence: an item response theory perspective. **Survey Practice**, v. 9, n. 5, 2016.

KLEIN, R.; FONTANIVE, N. S. Avaliação em larga escala: uma proposta inovadora. **Em Aberto**, Brasília, v. 15, n. 66, abr./jun. 1995.

LEVIN, J.; FOX. J. A. **Estatística para ciências humanas**. São Paulo: Prentice Hall, 2004.

LORD, F. M. **Applications of item response theory to practical testing problems**. New Jersey: Lawrence Erlbaum Associates, 1980.

LORD, F. M. **A theory of test scores**. New York : Psychometric Society, c1952.  
(Psychometric Monograph ; n. 7)

LUCKESI, C. C. **Avaliação da aprendizagem**: componente do ato pedagógico. São Paulo: Cortez, 2011.

\_\_\_\_\_. **Avaliação da aprendizagem escolar**: estudos e proposições. 19. ed. São Paulo: Cortez. 2008.

MACDONALD, B. C. **Esboços em avaliação educacional**. Fortaleza: Editora da UFC, 2003.

MARÔCO J. **Avaliação das qualidades psicométricas de uma escala**. Lisboa: Manuscrito; 2009.

MELO, João. **No Sisu diferença da nota de corte de cotistas e não cotistas é no máximo de 4%**. GGN, Caderno Educação, 20 jan, 2016. Disponível em <https://jornalggn.com.br/educacao/no-sisudiferenca-da-nota-de-corte-de-cotistas-e-naocotistas-e-no-maximo-de-4/>. Acesso em: 03 jul. 2019.

MENDES, L. F.; ARAUJO, L. S.; AIRTON. J. F. P. J.; TROMPIERI, N.F. **Efeito de posição na dificuldade dos itens do Enem**. ABAVE. 2016.

MICRODADOS ENEM 2017. Disponível em: <<http://portal.inep.gov.br/microdados>> Acesso em: 03 jul. 2019.

MONTGOMERY, Douglas C.; RUNGER, George C. **Estatística aplicada e probabilidade para engenheiros**. Rio de Janeiro : LTC, 2003.

MUÑIZ, J. **Teoría de respuesta a los ítems**: un nuevo enfoque en la evolución psicológica y educativa. Madri: Ediciones Pirámide, 1990.

OBINO, F. C. W. **Políticas de avaliação de larga escala na educação básica: do controle de resultados à intervenção no processos de operacionalização do ensino**. Ensaio: aval. pol. públ. Educ., Rio de Janeiro, v. 19, n. 73, p. 769-792, out./dez. 2011.

ONU. Enem 2010- **Nota do sistema do coordenador residente das nações unidas do Brasil**. Brasília, 2010.

PASQUALI, L. **Psicometria: teoria dos testes na psicologia e na educação**. Petrópolis: Vozes, 2003.

\_\_\_\_\_. **Psicometria: teoria dos testes na psicologia e na educação**. [S.l.]: Vozes. 2017. 392 p.

PAYNE. D. A. Toward a characterization of curriculum evaluation. *In*: Payne. D.A. (ed.). **Curriculum evaluation: commentaries on purpose, process, product**. Lexington, Mass.: D.C. Heath and Company. 1974.

RIBEIRO, Vanda Mendes. Justiça como equidade na escola, igualdade de base, currículo e avaliação externa. **Cenpec**, v. 3, n.1, 2013.

SILVA, Assis Leão; GOMES, Alfredo Macedo. Avaliação educacional: concepções e embates teóricos. **Estudos em Avaliação Educacional**, v. 29, n. 71, 2018.

SOARES, T. M.; GENOVEZ, S. F. M. ; GALVÃO, A. F. Análise do comportamento diferencial dos itens de geografia : estudo da 4ª série avaliada no Proeb/Simave 2001. **Estudos em Avaliação Educacional**, v. 16. n. 32, p. 81-102, jul./dez. 2005.

SPAGNA, D. J. **Enem**: veja estratégias que fazem a diferença na hora da prova. Disponível em: <https://guiadoestudante.abril.com.br/enem/enem-veja-estrategias-quefazem-a-diferenca-na-hora-daprova/> Acesso em 08 set. 2019.

THURSTONE, L. L. Attitudes can be measured. **American Journal of Sociology**, v. 33, n. 4, p. 529-554, 1928.

TOKARNIA, M. **Apenas metade dos estudantes conclui ensino médio na idade certa**. Disponível em: <https://www.terra.com.br/noticias/educacao/apenas-metade-dos-estudantes-com-cluiensino-medio-na-idade-certa,0813c3d1f7a3d310VgnCLD2000000ec6eb0aRCRD.html>. Acesso em: 08 set. 2019.

UNIVERSIDADE FEDERAL DO CEARÁ. Biblioteca Universitária. **Guia de normalização de trabalhos acadêmicos da Universidade Federal do Ceará**. Fortaleza, 2013.

VENABLES, W.N.; SMITH, D.M; R Development Core Team. **An Introduction to R**. [S.l.] : R. Evolution Computing, 2009.

VIANNA, H. M. Avaliação do desempenho em matemática e ciências: uma experiência em São Paulo e em Fortaleza. **Estudos em Avaliação Educacional**, São Paulo, n. 5, p. 9-14 . 1999.

\_\_\_\_\_. Avaliação educacional: algumas ideias percursoras. **Revista Educação e Seleção** , v. 6, p.61-70, 1982 .

\_\_\_\_\_. **Avaliações nacionais em larga escala**: análises e propostas. São Paulo: Fundação Carlos Chagas / DPE, 2003. 41 p. (Textos FCC, 23).

\_\_\_\_\_. **Pesquisa em educação**: a observação. Brasília: Plano, 2000.

VIEIRA, S. **Teste da DMS de Fischer para comparação de médias**. Disponível em: <http://soniavieira.blogspot.com/2016/09/teste-da-dms-de-fisher-para-comparacao.html>

Acesso em : 08 set. 2019.

WERLE, F. O. C. Políticas de avaliação de larga escala na educação básica: do controle de resultados à intervenção no processos de operacionalização do ensino. **Ensaio: Avaliação e Políticas Públicas em Educação**, Rio de Janeiro, v. 19, n. 73, p. 769-792, 2011.

## APÊNDICE A – TABELA DE RESULTADOS DIF QUI-QUADRADO

Mantel-Haenszel Chi-square statistic:								
AMARELA-AZUL			AMARELA-ROSA			AMARELA-CINZA		
Stat.	P-value		Stat.	P-value		Stat.	P-value	
V1	1.9758	0.1598	V1	685.5735	0.0000 ***	V1	31.8120	0.0000 ***
V2	104.2936	0.0000 ***	V2	134.2688	0.0000 ***	V2	442.2073	0.0000 ***
V3	417.9545	0.0000 ***	V3	1875.6073	0.0000 ***	V3	37.6420	0.0000 ***
V4	11.1474	0.0008 ***	V4	109.2984	0.0000 ***	V4	0.0001	0.9924
V5	220.0379	0.0000 ***	V5	205.1996	0.0000 ***	V5	532.1115	0.0000 ***
V6	9486.1328	0.0000 ***	V6	5.0457	0.0247 *	V6	3.0116	0.0827 .
V7	50.1035	0.0000 ***	V7	596.0007	0.0000 ***	V7	414.7725	0.0000 ***
V8	8.6723	0.0032 **	V8	142.7749	0.0000 ***	V8	321.0945	0.0000 ***
V9	609.4196	0.0000 ***	V9	267.7154	0.0000 ***	V9	253.3127	0.0000 ***
V10	117.4714	0.0000 ***	V10	5.4809	0.0192 *	V10	578.6651	0.0000 ***
V11	178.6509	0.0000 ***	V11	48.7632	0.0000 ***	V11	29.4829	0.0000 ***
V12	618.5046	0.0000 ***	V12	147.6431	0.0000 ***	V12	368.9589	0.0000 ***
V13	12.0462	0.0005 ***	V13	10.0597	0.0015 **	V13	375.0554	0.0000 ***
V14	67.5842	0.0000 ***	V14	34.7317	0.0000 ***	V14	15.3656	0.0001 ***
V15	2151.4301	0.0000 ***	V15	24.6825	0.0000 ***	V15	208.7151	0.0000 ***
V16	1.0961	0.2951	V16	42.8101	0.0000 ***	V16	8.5632	0.0034 **
V17	8.5482	0.0035 **	V17	25.8341	0.0000 ***	V17	84.5062	0.0000 ***
V18	364.4331	0.0000 ***	V18	92.8088	0.0000 ***	V18	77.3900	0.0000 ***
V19	9.6913	0.0019 **	V19	3.0317	0.0817 .	V19	145.4479	0.0000 ***
V20	40.9113	0.0000 ***	V20	323.9161	0.0000 ***	V20	158.3918	0.0000 ***
V21	1322.2591	0.0000 ***	V21	781.9764	0.0000 ***	V21	77.6408	0.0000 ***
V22	110.2028	0.0000 ***	V22	1.7472	0.1862	V22	3.7743	0.0520 .
V23	39.2810	0.0000 ***	V23	520.7694	0.0000 ***	V23	288.1009	0.0000 ***
V24	101.8816	0.0000 ***	V24	1933.3529	0.0000 ***	V24	54.7956	0.0000 ***
V25	809.1597	0.0000 ***	V25	716.0171	0.0000 ***	V25	0.6956	0.4043
V26	220.8659	0.0000 ***	V26	148.0858	0.0000 ***	V26	4.9859	0.0256 *
V27	45.3242	0.0000 ***	V27	17.5495	0.0000 ***	V27	108.8781	0.0000 ***
V28	6.3986	0.0114 *	V28	75.2464	0.0000 ***	V28	3.2154	0.0729 .
V29	263.2026	0.0000 ***	V29	148.3698	0.0000 ***	V29	575.4916	0.0000 ***
V30	0.3572	0.5501	V30	49.1285	0.0000 ***	V30	43.1023	0.0000 ***
V31	47.7948	0.0000 ***	V31	419.6256	0.0000 ***	V31	709.3042	0.0000 ***
V32	627.4716	0.0000 ***	V32	124.0281	0.0000 ***	V32	3.7086	0.0541 .
V33	303.4292	0.0000 ***	V33	31.5493	0.0000 ***	V33	35.7343	0.0000 ***
V34	9.7240	0.0018 **	V34	13.9265	0.0002 ***	V34	164.5329	0.0000 ***
V35	26.1034	0.0000 ***	V35	6.9035	0.0086 **	V35	1.7843	0.1816
V36	133.1703	0.0000 ***	V36	133.4761	0.0000 ***	V36	205.8998	0.0000 ***
V37	333.7228	0.0000 ***	V37	830.0958	0.0000 ***	V37	111.9639	0.0000 ***
V38	123.9064	0.0000 ***	V38	0.6623	0.4157	V38	231.3615	0.0000 ***
V39	31.3244	0.0000 ***	V39	0.8893	0.3457	V39	46.4469	0.0000 ***
V40	4.9863	0.0255 *	V40	14.3165	0.0002 ***	V40	233.3760	0.0000 ***

v41	0.0493	0.8243	v41	88.6378	0.0000 ***	v41	120.2035	0.0000 ***
v42	31.8466	0.0000 ***	v42	359.8100	0.0000 ***	v42	0.2616	0.6090
v43	386.9462	0.0000 ***	v43	168.8073	0.0000 ***	v43	43.2280	0.0000 ***
v44	10.5276	0.0012 **	v44	35.9897	0.0000 ***	v44	776.7668	0.0000 ***
v45	239.1375	0.0000 ***	v45	383.3351	0.0000 ***	v45	4.6214	0.0316 *

Detection threshold: 3.8415 (significance level: 0.05)

## APÊNDICE B – TABELA DE RESULTADOS DIF MANTEL HAENSZEL

DIF MH								
AMARELA-AZUL			AMARELA-ROSA			AMARELA-CINZA		
alphaMH deltaMH			alphaMH deltaMH			alphaMH deltaMH		
V1	0.9920	0.0189 A	V1	1.1597	-0.3482 A	V1	1.0328	-0.0759 A
V2	1.0821	-0.1854 A	V2	0.9117	0.2173 A	V2	0.8434	0.4002 A
V3	1.1991	-0.4267 A	V3	1.4491	-0.8718 A	V3	1.0576	-0.1315 A
V4	0.9778	0.0527 A	V4	1.0717	-0.1627 A	V4	1.0001	-0.0002 A
V5	0.9156	0.2073 A	V5	0.9164	0.2053 A	V5	0.8682	0.3322 A
V6	0.5747	1.3015 B	V6	0.9871	0.0305 A	V6	1.0101	-0.0237 A
V7	1.0455	-0.1046 A	V7	1.1647	-0.3583 A	V7	1.1356	-0.2989 A
V8	0.9798	0.0480 A	V8	0.9189	0.1988 A	V8	0.8801	0.3001 A
V9	1.1953	-0.4193 A	V9	1.1273	-0.2816 A	V9	1.1234	-0.2735 A
V10	1.0799	-0.1807 A	V10	1.0170	-0.0396 A	V10	1.1834	-0.3958 A
V11	0.9205	0.1946 A	V11	0.9577	0.1016 A	V11	0.9671	0.0785 A
V12	1.1619	-0.3527 A	V12	1.0769	-0.1740 A	V12	1.1237	-0.2740 A
V13	0.9781	0.0520 A	V13	1.0204	-0.0475 A	V13	1.1290	-0.2851 A
V14	0.9319	0.1657 A	V14	0.9508	0.1186 A	V14	0.9671	0.0786 A
V15	1.3172	-0.6475 A	V15	1.0310	-0.0718 A	V15	1.0915	-0.2057 A
V16	0.9933	0.0158 A	V16	0.9586	0.0993 A	V16	1.0189	-0.0439 A
V17	1.0190	-0.0443 A	V17	1.0334	-0.0772 A	V17	1.0606	-0.1383 A
V18	1.1245	-0.2758 A	V18	1.0617	-0.1408 A	V18	1.0561	-0.1282 A
V19	0.9814	0.0441 A	V19	0.9895	0.0248 A	V19	0.9293	0.1723 A
V20	1.0406	-0.0936 A	V20	1.1180	-0.2621 A	V20	1.0812	-0.1836 A
V21	0.8182	0.4717 A	V21	0.8571	0.3625 A	V21	0.9529	0.1133 A
V22	0.9265	0.1795 A	V22	1.0095	-0.0222 A	V22	0.9861	0.0329 A
V23	1.0378	-0.0873 A	V23	0.8749	0.3141 A	V23	0.9041	0.2370 A
V24	0.9443	0.1348 A	V24	0.7770	0.5928 A	V24	0.9587	0.0992 A
V25	1.2493	-0.5230 A	V25	1.2344	-0.4949 A	V25	0.9932	0.0161 A
V26	0.9135	0.2127 A	V26	0.9287	0.1739 A	V26	0.9866	0.0318 A
V27	0.9621	0.0907 A	V27	0.9765	0.0559 A	V27	0.9416	0.1414 A
V28	1.0165	-0.0384 A	V28	1.0573	-0.1309 A	V28	1.0117	-0.0273 A
V29	1.1001	-0.2243 A	V29	1.0744	-0.1687 A	V29	0.8665	0.3368 A
V30	1.0038	-0.0090 A	V30	1.0456	-0.1049 A	V30	0.9586	0.0994 A
V31	1.0456	-0.1048 A	V31	1.1395	-0.3069 A	V31	1.1835	-0.3959 A
V32	0.7847	0.5699 A	V32	0.9006	0.2460 A	V32	0.9823	0.0419 A
V33	1.1128	-0.2512 A	V33	1.0355	-0.0821 A	V33	1.0377	-0.0870 A
V34	1.0186	-0.0432 A	V34	0.9780	0.0523 A	V34	0.9263	0.1798 A
V35	1.0327	-0.0756 A	V35	0.9834	0.0394 A	V35	0.9916	0.0199 A
V36	0.9318	0.1661 A	V36	0.9314	0.1671 A	V36	1.0898	-0.2021 A
V37	0.8930	0.2659 A	V37	0.8348	0.4243 A	V37	0.9369	0.1531 A
V38	1.0925	-0.2079 A	V38	0.9934	0.0156 A	V38	0.8813	0.2969 A
V39	1.0334	-0.0771 A	V39	0.9944	0.0131 A	V39	1.0408	-0.0940 A
V40	0.9868	0.0312 A	V40	0.9777	0.0530 A	V40	1.0940	-0.2112 A

v41	0.9986	0.0032	A	v41	0.9444	0.1344	A	v41	0.9357	0.1561	A
v42	1.0442	-0.1016	A	v42	0.8601	0.3543	A	v42	1.0040	-0.0093	A
v43	1.1751	-0.3793	A	v43	1.1133	-0.2523	A	v43	1.0562	-0.1284	A
v44	0.9786	0.0508	A	v44	0.9607	0.0943	A	v44	0.8262	0.4486	A
v45	1.1201	-0.2666	A	v45	1.1538	-0.3361	A	v45	1.0162	-0.0377	A

Effect size code:

'A': negligible effect

'B': moderate effect

'C': large effect

### APÊNDICE C – TABELA DE RESULTADOS DIF QUI-QUADRADO

Mantel-Haenszel Chi-square statistic:											
AZUL-CINZA				AZUL-ROSA				ROSA-CINZA			
V1	49.4969	0.0000	***	V1	738.3090	0.0000	***	V1	411.6799	0.0000	***
V2	971.0304	0.0000	***	V2	462.4409	0.0000	***	V2	93.5374	0.0000	***
V3	205.2729	0.0000	***	V3	523.2967	0.0000	***	V3	1375.5319	0.0000	***
V4	11.1093	0.0009	***	V4	187.1226	0.0000	***	V4	106.7326	0.0000	***
V5	59.4045	0.0000	***	V5	0.1543	0.6945		V5	75.9290	0.0000	***
V6	9813.4403	0.0000	***	V6	8940.1800	0.0000	***	V6	16.6805	0.0000	***
V7	175.0926	0.0000	***	V7	292.2872	0.0000	***	V7	14.9029	0.0001	***
V8	217.2210	0.0000	***	V8	73.9236	0.0000	***	V8	39.8694	0.0000	***
V9	76.1406	0.0000	***	V9	71.1234	0.0000	***	V9	0.0892	0.7651	
V10	173.9556	0.0000	***	V10	74.3596	0.0000	***	V10	474.0017	0.0000	***
V11	62.0922	0.0000	***	V11	42.9557	0.0000	***	V11	2.1969	0.1383	
V12	32.8858	0.0000	***	V12	166.0579	0.0000	***	V12	50.9166	0.0000	***
V13	517.2648	0.0000	***	V13	42.2922	0.0000	***	V13	265.7725	0.0000	***
V14	18.8128	0.0000	***	V14	4.8931	0.0270	*	V14	3.9964	0.0456	*
V15	1019.3371	0.0000	***	V15	1737.3834	0.0000	***	V15	91.6887	0.0000	***
V16	16.3238	0.0001	***	V16	28.8609	0.0000	***	V16	87.2024	0.0000	***
V17	38.3971	0.0000	***	V17	4.4640	0.0346	*	V17	17.7092	0.0000	***
V18	106.1541	0.0000	***	V18	91.0788	0.0000	***	V18	0.5651	0.4522	
V19	77.9860	0.0000	***	V19	1.8592	0.1727		V19	107.5137	0.0000	***
V20	37.6291	0.0000	***	V20	130.1739	0.0000	***	V20	27.2923	0.0000	***
V21	766.3488	0.0000	***	V21	78.2340	0.0000	***	V21	359.0643	0.0000	***
V22	72.7464	0.0000	***	V22	143.3380	0.0000	***	V22	10.5117	0.0012	**
V23	532.2039	0.0000	***	V23	871.7320	0.0000	***	V23	40.5132	0.0000	***
V24	7.4250	0.0064	**	V24	1144.3429	0.0000	***	V24	1334.3653	0.0000	***
V25	855.0834	0.0000	***	V25	3.2560	0.0712	.	V25	756.4506	0.0000	***
V26	160.7037	0.0000	***	V26	6.8895	0.0087	**	V26	99.7264	0.0000	***
V27	12.5568	0.0004	***	V27	4.3895	0.0362	*	V27	34.9340	0.0000	***
V28	0.5523	0.4574		V28	36.4361	0.0000	***	V28	45.9552	0.0000	***
V29	1611.4065	0.0000	***	V29	17.4147	0.0000	***	V29	1282.0808	0.0000	***
V30	50.5922	0.0000	***	V30	39.7619	0.0000	***	V30	184.2560	0.0000	***
V31	392.2766	0.0000	***	V31	187.3331	0.0000	***	V31	39.4033	0.0000	***
V32	531.6421	0.0000	***	V32	214.0690	0.0000	***	V32	78.8779	0.0000	***
V33	131.1429	0.0000	***	V33	139.7153	0.0000	***	V33	0.1293	0.7191	
V34	255.4068	0.0000	***	V34	45.9922	0.0000	***	V34	84.2948	0.0000	***
V35	42.0417	0.0000	***	V35	58.7346	0.0000	***	V35	1.5271	0.2165	
V36	649.9313	0.0000	***	V36	0.0197	0.8883		V36	674.8114	0.0000	***
V37	59.7482	0.0000	***	V37	104.8051	0.0000	***	V37	324.9751	0.0000	***
V38	691.3058	0.0000	***	V38	135.7905	0.0000	***	V38	215.6918	0.0000	***
V39	1.7185	0.1899		V39	39.5303	0.0000	***	V39	57.6095	0.0000	***
V40	303.4172	0.0000	***	V40	2.4084	0.1207		V40	364.8463	0.0000	***
V41	115.2843	0.0000	***	V41	77.7645	0.0000	***	V41	3.5775	0.0586	.

v42	27.3966	0.0000	***	v42	602.4510	0.0000	***	v42	376.2960	0.0000	***
v43	173.0036	0.0000	***	v43	43.9917	0.0000	***	v43	41.0233	0.0000	***
v44	603.0187	0.0000	***	v44	5.7387	0.0166	*	v44	493.2347	0.0000	***
v45	176.7338	0.0000	***	v45	16.8555	0.0000	***	v45	304.0622	0.0000	***

Detection threshold: 3.8415 (significance level: 0.05)

## APÊNDICE D – TABELA DE RESULTADOS DIF MANTEL HAENSZEL

DIF MH								
AZUL-CINZA			AZUL-ROSA			ROSA-CINZA		
V1	1.0409	-0.0943 A	V1	1.1657	-0.3604 A	V1	0.8919	0.2690 A
V2	0.7799	0.5843 A	V2	0.8443	0.3976 A	V2	0.9233	0.1875 A
V3	0.8818	0.2956 A	V3	1.2076	-0.4434 A	V3	0.7305	0.7381 A
V4	1.0226	-0.0525 A	V4	1.0949	-0.2131 A	V4	0.9340	0.1605 A
V5	0.9546	0.1092 A	V5	1.0024	-0.0056 A	V5	0.9475	0.1267 A
V6	1.7562	-1.3234 B	V6	1.7098	-1.2605 B	V6	1.0239	-0.0554 A
V7	1.0851	-0.1920 A	V7	1.1113	-0.2480 A	V7	0.9766	0.0556 A
V8	0.9008	0.2455 A	V8	0.9412	0.1425 A	V8	0.9552	0.1077 A
V9	0.9402	0.1450 A	V9	0.9420	0.1403 A	V9	0.9978	0.0051 A
V10	1.0952	-0.2137 A	V10	0.9408	0.1435 A	V10	1.1643	-0.3575 A
V11	1.0502	-0.1151 A	V11	1.0418	-0.0963 A	V11	1.0093	-0.0217 A
V12	0.9663	0.0805 A	V12	0.9256	0.1816 A	V12	1.0440	-0.1012 A
V13	1.1531	-0.3347 A	V13	1.0422	-0.0972 A	V13	1.1072	-0.2393 A
V14	1.0382	-0.0880 A	V14	1.0194	-0.0451 A	V14	1.0174	-0.0406 A
V15	0.8289	0.4410 A	V15	0.7802	0.5833 A	V15	1.0599	-0.1367 A
V16	1.0262	-0.0607 A	V16	0.9658	0.0817 A	V16	1.0621	-0.1416 A
V17	1.0404	-0.0930 A	V17	1.0137	-0.0320 A	V17	1.0273	-0.0633 A
V18	0.9390	0.1478 A	V18	0.9433	0.1373 A	V18	0.9954	0.0109 A
V19	0.9477	0.1261 A	V19	1.0083	-0.0194 A	V19	0.9387	0.1486 A
V20	1.0384	-0.0886 A	V20	1.0725	-0.1646 A	V20	0.9685	0.0752 A
V21	1.1659	-0.3607 A	V21	1.0507	-0.1163 A	V21	1.1108	-0.2469 A
V22	1.0641	-0.1460 A	V22	1.0904	-0.2034 A	V22	0.9771	0.0546 A
V23	0.8729	0.3195 A	V23	0.8421	0.4039 A	V23	1.0378	-0.0873 A
V24	1.0156	-0.0364 A	V24	0.8241	0.4546 A	V24	1.2331	-0.4925 A
V25	0.7954	0.5380 A	V25	0.9865	0.0320 A	V25	0.8052	0.5090 A
V26	1.0803	-0.1816 A	V26	1.0162	-0.0379 A	V26	1.0627	-0.1429 A
V27	0.9798	0.0480 A	V27	1.0120	-0.0280 A	V27	0.9668	0.0793 A
V28	0.9952	0.0113 A	V28	1.0394	-0.0908 A	V28	0.9575	0.1021 A
V29	0.7887	0.5578 A	V29	0.9759	0.0572 A	V29	0.8088	0.4987 A
V30	0.9553	0.1074 A	V30	1.0409	-0.0942 A	V30	0.9168	0.2042 A
V31	1.1330	-0.2935 A	V31	1.0909	-0.2046 A	V31	1.0399	-0.0920 A
V32	1.2512	-0.5267 A	V32	1.1563	-0.3414 A	V32	1.0876	-0.1972 A
V33	0.9325	0.1643 A	V33	0.9302	0.1701 A	V33	1.0022	-0.0053 A
V34	0.9094	0.2231 A	V34	0.9606	0.0945 A	V34	0.9464	0.1294 A
V35	0.9601	0.0958 A	V35	0.9526	0.1142 A	V35	1.0079	-0.0185 A
V36	1.1664	-0.3618 A	V36	1.0009	-0.0021 A	V36	1.1703	-0.3697 A
V37	1.0495	-0.1135 A	V37	0.9369	0.1531 A	V37	1.1208	-0.2679 A
V38	0.8070	0.5040 A	V38	0.9112	0.2186 A	V38	0.8846	0.2881 A
V39	1.0077	-0.0181 A	V39	0.9636	0.0871 A	V39	1.0458	-0.1052 A
V40	1.1080	-0.2411 A	V40	0.9908	0.0218 A	V40	1.1194	-0.2651 A
V41	0.9369	0.1532 A	V41	0.9477	0.1263 A	V41	0.9884	0.0274 A
V42	0.9607	0.0943 A	V42	0.8233	0.4568 A	V42	1.1669	-0.3628 A

v43	0.8984	0.2517	A	v43	0.9477	0.1263	A	v43	0.9488	0.1235	A
v44	0.8445	0.3973	A	v44	0.9840	0.0379	A	v44	0.8577	0.3608	A
v45	0.9073	0.2285	A	v45	1.0299	-0.0692	A	v45	0.8807	0.2987	A

Effect size code:

'A': negligible effect

'B': moderate effect

'C': large effect