

Medidas de Tendência Central: Análise da Qualidade das Questões do ENEM de 2016 a 2018

Measures Of Central Tendency: Evaluation of ENEM Questions From 2016 to 2018

Talita Emídio Andrade Soares^a; Denilson Junio Marques Soares^{ab}; Wagner dos Santos^{*ac}

^aUniversidade Federal do Espírito Santo, Programa de Pós-Graduação Stricto Sensu em Educação. ES, Brasil.

^bInstituto Federal de Minas Gerais – Campus Piumhi. MG, Brasil.

^cUniversidade Federal do Espírito Santo, Programa de Pós-Graduação Stricto Sensu em Educação Física. ES, Brasil.

*E-mail: wagnercefd@gmail.com

Resumo

O Exame Nacional do Ensino Médio (ENEM) tem se constituído como uma das principais políticas de avaliação da qualidade da educação básica brasileira. A expansão do exame e as atribuições a ele investidas tornaram imprescindíveis as discussões acerca da sua estrutura e qualidade. Com o intuito de contribuir para esta temática, este artigo tem o objetivo de analisar a qualidade de 11 itens de Matemática das provas do ENEM de 2016 a 2018, que se referem às Medidas de Tendência Central: média aritmética simples e ponderada, mediana e moda. Para isto, adotou-se uma metodologia predominantemente quantitativa e exploratória, desenvolvida a partir da aplicação de técnicas psicométricas a uma amostra representativa de participantes do exame. Os resultados apontaram para itens adequados e bem estruturados, que privilegiam o raciocínio lógico e a interpretação. A análise pedagógica indicou para a comum confusão entre os conceitos de média aritmética simples e ponderada, que representam a grande maioria dos itens analisadas. Espera-se que os resultados obtidos possam esclarecer alguns conceitos da psicometria e auxiliar no processo de ensino e aprendizagem de estatística na educação básica.

Palavras-chave: Medidas de Tendência Central. Psicometria. Análise de Itens. ENEM.

Abstract

The National High School Examination (ENEM) has become one of the most of the main Brazilian quality evaluation policies in the field of basic education. The expansion and attributions invested in the test make indispensable the discussions about its structure and quality. In order to contribute to this theme, we propose to analyze 11 Maths questions to ENEM from 2016 to 2018 and related to Measures of Central Tendency: arithmetic mean; weighted arithmetic mean, median and mode. It was used a quantitative and exploratory methodology, developed through psychometrics methods applied to a significant sample of students that made the exam. The results pointed to well-structured items, that favor logical reasoning and interpretation and that have an excellent ability discriminate groups with different levels of. Its highlight the confusion between concepts of arithmetic mean and weighted arithmetic mean, that represent the most of the questions analyzed. It is hoped that this research may help clarify some concepts of psychometry and can assist in the teaching-learning process of statistics in basic education.

Keywords: Measures of Central Tendency. Psychometry. Item Analysis. ENEM.

1 Introdução

Segundo Fernandes, Sousa & Ribeiro (2004), o ensino de Estatística na Educação Básica é fundamental para o desenvolvimento de capacidades que permitem que o aluno tome atitudes críticas face ao que veem ou ouvem, desenvolvendo competências no âmbito da argumentação e da tomada de decisões. Os Parâmetros Curriculares Nacionais (PCN's), instituídos em 1998, já alertavam para isto:

[...] a compreensão e a tomada de decisões diante de questões políticas e sociais dependem da leitura crítica e interpretação de informações complexas, muitas vezes contraditórias, que incluem dados estatísticos e índices divulgados pelos meios de comunicação. Ou seja, para exercer a cidadania é necessário saber calcular, medir, raciocinar, argumentar, tratar informações estatisticamente etc. (Brasil, 1998, p. 27).

A ideia de uniformizar alguns aspectos considerados fundamentais a cada área do conhecimento, trazida pelos

PCN's, pôde se materializar no documento básico que fez parte da implementação do Exame Nacional do Ensino Médio (ENEM), também em 1998. Desde então, o ENEM se constituiu como uma das principais políticas de avaliação da qualidade da educação básica brasileira.

Atualmente, além de compor o processo de admissão de estudantes de um considerável número de instituições públicas de ensino superior, o ENEM também pode ser utilizado para a obtenção de bolsas de estudo ou no financiamento de cursos em instituições privadas. Estas atribuições corroboraram para a expansão do exame que tem atingido números exorbitantes de inscritos, tornando-o uma das maiores avaliações em larga escala do mundo (Travitzki, 2017). A expansão do exame e as atribuições a ele investidas ao longo dos anos tornaram imprescindíveis as discussões acerca da sua estrutura e qualidade.

O ENEM é composto por uma redação e por quatro provas de múltipla escolha: Matemática e suas Tecnologias,

Linguagens Códigos e suas Tecnologias, Ciências Humanas e suas Tecnologias e Ciências da Natureza e suas Tecnologias. Cada prova é constituída por 45 itens ancorados na interdisciplinaridade e na contextualização, desenvolvidas a partir de uma Matriz de Referência.

A prova de Matemática e suas Tecnologias está organizada em sete competências e trinta habilidades a elas associadas. A Estatística está incluída na Competência 7 da Matriz de Referência da área que, por sua vez, se desdobra em 4 habilidades (H27, H28, H29 e H30), definidas como:

Competência de área 7 - Compreender o caráter aleatório e não-determinístico dos fenômenos naturais e sociais e utilizar instrumentos adequados para medidas, determinação de amostras e cálculos de probabilidade para interpretar informações de variáveis apresentadas em uma distribuição estatística.

H27 - Calcular medidas de tendência central ou de dispersão de um conjunto de dados expressos em uma tabela de frequências de dados agrupados (não em classes) ou em gráficos;

H28 - Resolver situação-problema que envolva conhecimentos de estatística e probabilidade;

H29 - Utilizar conhecimentos de estatística e probabilidade como recurso para a construção de argumentação;

H30 - Avaliar propostas de intervenção na realidade utilizando conhecimentos de estatística e probabilidade (Brasil, 2009, p.7).

Com questões que privilegiam a análise e interpretação de dados, geralmente dispostos em quadros, tabelas e gráficos, a Estatística esteve presente em todas as edições do exame. Apesar dessa periodicidade, ainda é pequeno o número de estudos voltados à avaliar a qualidade dessas questões, sobretudo pautados em técnicas psicométricas.

De fato, não há muitos documentos oficiais que tratam dos procedimentos metodológicos que garantem o bom funcionamento do exame e da qualidade dos seus itens, ficando a cargo dos pesquisadores interessados em avaliar a avaliação. Felizmente, o INEP disponibiliza informações detalhadas sobre o exame, potencializando inúmeros estudos estatísticos, pedagógicos, políticos, entre outros.

Com o intuito de contribuir para o preenchimento desta lacuna, este trabalho tem o objetivo de avaliar a qualidade técnica dos itens de Matemática das provas do ENEM de 2016 a 2018, relacionadas às Medidas de Tendência Central: média aritmética simples e ponderada, mediana e moda. Essa análise irá se basear na aplicação de técnicas psicométricas e no conteúdo dos itens.

2 Análise Psicométrica de Itens

Segundo Pasquali (2009), a Psicometria se fundamenta na Teoria da Medida para explicar o sentido das respostas dadas por indivíduos a um conjunto de itens, que podem, no caso de avaliações de conhecimento, ser interpretados como questões. A Psicometria Moderna está estruturada em duas vertentes: Teoria Clássica dos Testes (TCT) e Teoria de Resposta ao Item (TRI).

De acordo com Muñiz (1990), a principal diferença entre estas teorias está no sentido em que considera estas respostas. Enquanto que a TCT se preocupa em explicar o conjunto de respostas, a TRI está voltada para a análise individual das respostas de cada item.

As técnicas de análise via TCT são predominantemente descritivas e podem ser executadas facilmente por professores e pesquisadores que se interessam pelo tema. Para definir o índice de dificuldade de uma questão, por exemplo, ela considera o seu percentual de acertos. Condé (2001) apresenta uma classificação para este índice, representada pelo Quadro 1:

Quadro 1 – Classificação para o índice de dificuldade pela TCT

Classificação para o Índice de Dificuldade	Percentual de Acerto da Questão
Baixa	Superior a 70%
Moderado	Superior a 30% e inferior ou igual a 70%
Alta	Inferior ou igual a 30%

Fonte: Adaptado de Condé (2001).

Para analisar a capacidade que um item possui em discriminar níveis distintos de habilidades, na TCT faz-se uso do Coeficiente de Correlação Ponto-Bisserial. Trata-se de uma generalização do Coeficiente de Correlação de Pearson para as variáveis desempenho na questão (dicotomizada como certa ou errada) e total de acertos na prova. Segundo Tórres (2015) é esperado que o valor retornado para esta estatística seja superior a para que ela seja considerada de boa discriminação.

Entretanto, apesar de sua praticidade e fácil aplicabilidade em diferentes contextos, a TCT possui limitações que corroboraram para a expansão do uso da TRI como metodologia de análise psicométrica de itens em avaliações standardizadas. Entre essas limitações, destacam-se as dificuldades para se comparar o desempenho de respondentes submetidos a provas distintas, o que pode comprometer a sua isonomia, ou a edições distintas de um mesmo exame, dificultando estudos longitudinais que podem ser úteis na avaliação da eficiência de uma política pública adotada, por exemplo (Pasquali, 2009, Rabelo, 2013, Moreira *et al.*, 2020, Costa, Lima & Soares, 2020).

De acordo com Andrade, Tavares e Valle (2000, p.17), a TRI pode ser compreendida como “um conjunto de modelos matemáticos que procuram representar a probabilidade de um indivíduo dar uma certa resposta a um item como função dos parâmetros do item e da habilidade (ou habilidades) do respondente”. Para tanto, comumente são utilizados os modelos logísticos, admitindo-se que a probabilidade de acerto de um item está condicionada ao grau de conhecimento do seu respondente.

O ENEM utiliza o modelo logístico de três parâmetros, que considera os índices de discriminação, dificuldade e a probabilidade de acerto ao acaso do item. Ainda segundo Andrade, Tavares & Valle (2000), a expressão Matemática

desse modelo é dada por:

$$P(U_{ij} = 1 | \theta_j, a_i, b_i, c_i) = c_i + (1 - c_i) \frac{e^{a_i(\theta_j - b_i)}}{1 + e^{a_i(\theta_j - b_i)}}$$

em que c_i é a probabilidade do indivíduo com habilidade acertar o item i , θ_j é o parâmetro de discriminação do item j , b_i é o parâmetro de dificuldade do item i medido na mesma escala da habilidade e a_i é o parâmetro do modelo que representa a probabilidade de acerto ao acaso do item i .

Para analisar a qualidade dos itens, Alnabhan & Harwell (2001), ao avaliarem um exame utilizado para auxiliar as universidades nas decisões de admissão na Jordânia, baseando-se nos estudos realizados por Muthén, Kao & Burstein (1991), consideraram adequados itens com um bom ajuste pela TRI, com índices de dificuldade com valores entre -3 e 3, índice de discriminação acima de 0,5 e cujo percentual de acerto fosse inferior à 90%, considerando que itens fora desses intervalos trazem pouca (ou nenhuma) informação relevante acerca de seus respondentes.

Além desses critérios, sob a ótica da TRI, Baker (2001) apresenta uma classificação para o índice de discriminação e Rabelo (2013) para o índice de dificuldade que estão descritas nas Quadros 2 e 3, respectivamente.

Quadro 2 – Classificação para o índice de discriminação da TRI

Classificação	Índice de discriminação
Muito Baixa	(0; 0,34]
Baixa	(0,34; 0,64]
Moderada	(0,64; 1,34]
Alta	(1,34; 1,69]
Muito alta	(1,69; +

Fonte: Adaptado de Baker (2001).

Quadro 3 – Classificação para o índice de dificuldade da TRI

Classificação	Índice de dificuldade
Muito Baixa	(-)
Baixa	(;]
Moderada	(]
Alta	(0,52; 1,28]
Muito alta	(1,28

Fonte: Adaptado de Rabelo (2013).

Entretanto, antes de proceder à estimação dos parâmetros dos itens e dos respondentes por meio desse modelo é necessário verificar duas pressuposições básicas da TRI: unidimensionalidade e independência local. O pressuposto da unidimensionalidade se refere à presença de uma habilidade dominante, responsável pelo conjunto de itens. Já o pressuposto da independência local assume que, considerando esta habilidade, as respostas aos diferentes itens do teste são independentes. Em outras palavras, o desempenho em um item não interfere no desempenho em outro (Travitzki, 2017, Pasquali, 2018, Soares, 2018).

Pasquali & Primi (2003) afirmam que estes pressupostos são axiomas, não havendo uma demonstração lógica ou empírica para eles. Hambleton, Swaminathan & Rogers (1991) postulam que a unidimensionalidade implica em

independência local, sendo sua verificação suficiente para a aplicação das técnicas da TRI.

Para tanto, a literatura da área indica diversos métodos, pautados na análise fatorial, como a análise do gráfico *scree plot*, análise confirmatória e análise paralela modificada (Chernyshenko *et al.*, 2001; Hair Jr. *et al.*, 2005). Essa última, envolve o cálculo da matriz de correlação tetracórica dos itens dicotômicos, derivando seu segundo autovalor que é comparado com a média dos segundos autovalores obtidos usando simulações de Monte Carlo (Dragow & Lissak, 1983).

3 Material e Métodos

As análises iniciais foram realizadas considerando as provas de Matemática do ENEM de 2016 a 2018. Este recorte temporal se justifica considerando o quantitativo de itens a serem analisados (135) e por serem, até o momento das análises, as edições cujos microdados disponibilizados no portal eletrônico do INEP, eram mais recentes.

As informações referentes aos itens, presentes nos microdados do ENEM, permitiram eleger os que eram referentes à competência 7 da Matriz de Referência da área. Estes itens foram resolvidos e classificados conforme os conceitos matemáticos exigidos para as suas resoluções. Em seguida, foram selecionados os itens relacionados às Medidas de Tendência Central, para comporem as análises deste trabalho. A escolha por este tópico se deu por se tratar de um assunto recorrente em avaliações standardizadas e que possibilita ricas discussões acerca das metodologias que envolvem o seu ensino.

Para a leitura dos microdados e para a análise psicométrica desses itens, foi utilizado o *software* estatístico R (R Core Team, 2020), versão 3.6.2. Trata-se de um *software* livre e de código aberto de fácil manuseio que oferece uma grande diversidade de funcionalidades. Os pacotes utilizados foram o *ltm* (Rizopoulos, 2006), *mirt* (Chalmers, 2012) e *psych* (Revelle & Revelle, 2015). Para os testes de hipóteses realizados, adotou-se 5% como nível de significância.

Inicialmente, para cada edição analisada, foram selecionados, aleatoriamente, 1 milhão de respondentes e destes considerou-se apenas os que cursavam e concluiriam o Ensino Médio no ano de realização do ENEM. Excluíram-se os participantes cuja planilha de dados não estava completa (apresentavam dados ausentes – NA's) e, considerando a divergência dos gabaritos dos diferentes cadernos de prova de cada exame, adotou-se o caderno azul (códigos 459, 403 e 303, respectivamente) como o referencial, excluindo os respondentes de outro tipo de caderno. Assim, obteve-se amostras de tamanho 49.970, 67.330 e 68.438 para as edições de 2016, 2017 e 2018, respectivamente

Em seguida, para essas amostras, realizou-se análises descritivas que contemplam a TCT. Foram consideradas as frequências de resposta a cada uma das cinco alternativas dos itens analisados, representando-as através de gráficos

de barras, e foram calculados os coeficientes de correlação ponto-bisserial de cada item.

Para a análise via TRI, primeiramente foi verificado o pressuposto da unidimensionalidade, por meio da Análise Paralela Modificada (APM), que compõe a Análise Fatorial (AF), conforme proposto por Drasgow & Lissak (1983). Ressalta-se que a realização dessa análise foi adequada, considerando que obteve-se estatística Kaiser-Meyer-Olkin (KMO) igual a 0,82 e teste de esfericidade de Bartlett significativo.

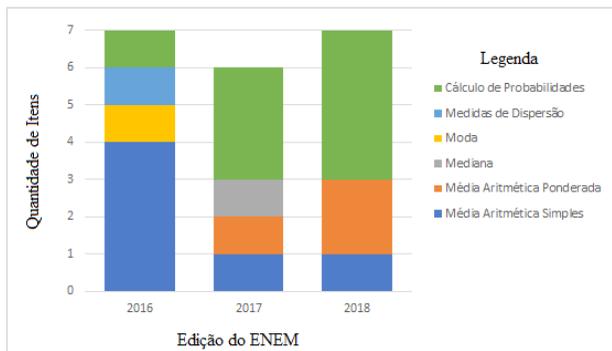
Em seguida, estimou-se os parâmetros dos itens e dos respondentes por meio de um modelo logístico de 3 parâmetros, como o utilizado no ENEM. Para isto, foram considerados 61 pontos de quadratura, 1000 iterações quase-Newton e 400 iterações *Estimation-Maximization*. De acordo com Travitzki (2017), essas configurações normalmente são suficientes para uma boa calibração dos itens.

Para verificar a qualidade do ajuste do modelo procedeu-se ao teste Qui-Quadrado de Bock (1972), cuja estatística é obtida comparando as respostas do modelo real com as respostas do modelo esperado. A hipótese de nulidade do teste é a de que o item possui um bom ajuste. Por fim, é apresentada uma classificação para os parâmetros psicométricos estimados para cada item e discute-se acerca do seu conteúdo.

4 Resultados e Discussão

Nas provas de Matemática das edições de 2016, 2017 e 2018 do ENEM, foram identificados 20 itens referentes a competência 7, dos quais 11 exigiam o cálculo das medidas de tendência central: média aritmética simples (6 itens) e ponderada (3 itens), mediana (1 item) e moda (1 item), no seu desenvolvimento. Os demais eram referentes às medidas de dispersão (1 item sobre desvio-padrão) e ao cálculo de probabilidades (8 itens). A Figura 1 mostra a distribuição dos itens.

Figura 1 – Distribuição dos 20 itens referentes a competência 7 da área de Matemática do ENEM, identificados nas provas de 2016, 2017 e 2018 do exame



Fonte: Baseado nos microdados do ENEM (Brasil, 2016, 2017, 2018).

A seguir, são expostos os resultados da análise psicométrica realizada para cada um desses 11 itens, separados por ano. Considerando a TCT, apresenta-se gráficos de barras que descrevem o percentual de respondentes que assinalou cada

uma das alternativas de resposta do item e o coeficiente de correlação Ponto-Bisserial (pBis). Para fins de classificação, considerou-se, para avaliar a dificuldade do item, os intervalos propostos por Condé (2001), conforme Tabela 1. Quanto a discriminação, assumiu-se um valor mínimo de 0,30 como satisfatório, de acordo com o estudo de Tôres (2015).

Sob a ótica da TRI, apresenta-se os parâmetros de discriminação (a), dificuldade (b) e probabilidade de acerto ao acaso (c), bem como seus respectivos erros-padrão (ep). Ademais, apresenta-se o p-valor obtido para o teste Qui-Quadrado (Chisq), para avaliar a qualidade do ajuste ao modelo. Quanto a classificação, assumiu-se os intervalos descritos por Baker (2001), para a discriminação e por Rabelo (2013) para a dificuldade, conforme Quadro 2 e 3, respectivamente.

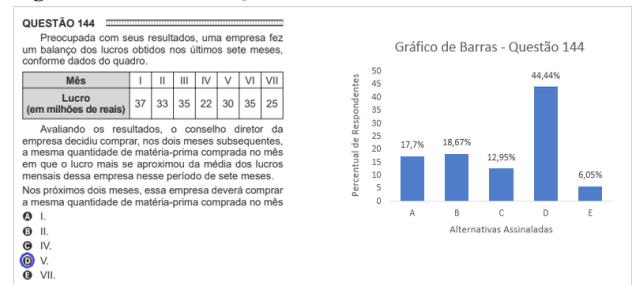
Também é descrita a Habilidade (Hab) pela qual a questão está relacionada. Essa informação foi extraída da planilha de itens dos microdados do ENEM. O Gabarito encontra-se assinalado nos itens.

4.1 Análise das questões de Medidas de Tendência Central do ENEM 2016

Na prova de Matemática e suas Tecnologias do ENEM 2016 foram identificados 5 itens relacionados as Medidas de Tendência Central, dos quais 4 exigiam o cálculo da média aritmética simples e 1 exigia o cálculo da moda.

4.1.1 Questão 144

Figura 2 – Análise da Questão 144 do ENEM 2016



Fonte: Baseado nos microdados do ENEM (Brasil, 2016, 2017, 2018).

Quadro 4 – Parâmetros Psicométricos da Questão 144 do ENEM 2016

Hab	pBis	a	ep(a)	b	ep(b)	c	ep(c)	Chisq (p-valor)
29	0,4484	1,743	0,05	0,452	0,01	0,18	< 0,01	< 0,05

Fonte: Baseado nos microdados do ENEM (Brasil, 2016, 2017, 2018).

Assunto: Média Aritmética Simples.

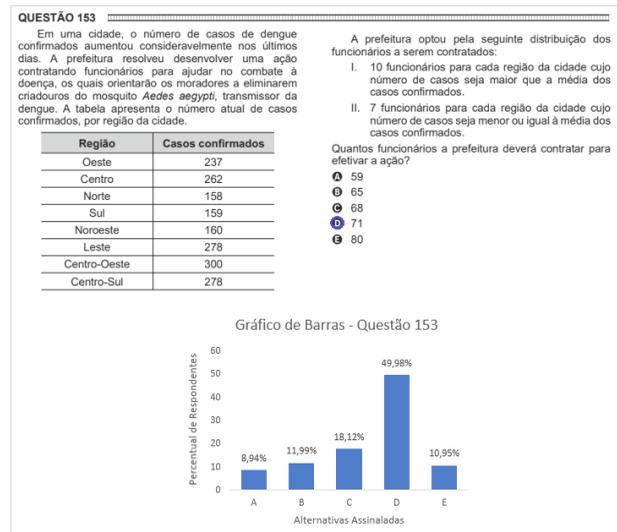
Classificação via TCT: Discriminação satisfatória e dificuldade moderada.

Classificação via TRI: Discriminação muito alta e dificuldade moderada.

Comentário: Exige conhecimentos básicos de aritmética para o cálculo da média aritmética simples dos lucros mensais de uma empresa em um período de sete meses.

4.1.2 Questão 153

Figura 3 – Análise da Questão 153 do ENEM 2016



Fonte: Baseado nos microdados do ENEM (Brasil, 2016).

Quadro 5 – Parâmetros Psicométricos da Questão 153 do ENEM 2016

Hab	pBis	a	ep(a)	b	ep(b)	c	ep(c)	Chisq (p-valor)
28	0,3822	1,524	0,06	0,509	0,01	0,15	< 0,01	< 0,05

Fonte: Baseado nos microdados do ENEM (Brasil, 2016).

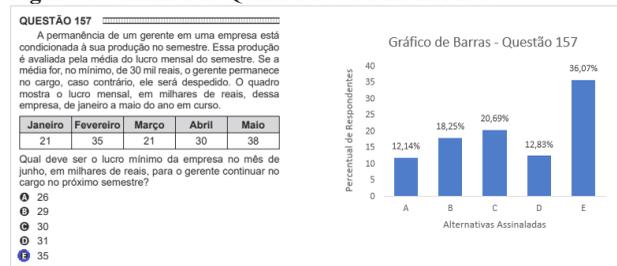
Assunto: Média Aritmética Simples.

Classificação via TCT: Discriminação satisfatória e dificuldade moderada.

Classificação via TRI: Discriminação alta e dificuldade moderada.
Comentário: Exigia conhecimentos básicos de aritmética para o cálculo da média aritmética simples do número de casos de dengue confirmados e, em seguida, do número de funcionários a serem contratados pela prefeitura.

4.1.3 Questão 157

Figura 4 – Análise da Questão 157 do ENEM 2016



Fonte: Baseado nos microdados do ENEM (Brasil, 2016).

Quadro 6 – Parâmetros Psicométricos da Questão 157 do ENEM 2016

Hab	pBis	a	ep(a)	b	ep(b)	c	ep(c)	Chisq (p-valor)
27	0,4682	3,274	0,08	0,987	0,01	0,17	< 0,01	< 0,05

Fonte: Baseado nos microdados do ENEM (Brasil, 2016).

Assunto: Média Aritmética Simples.

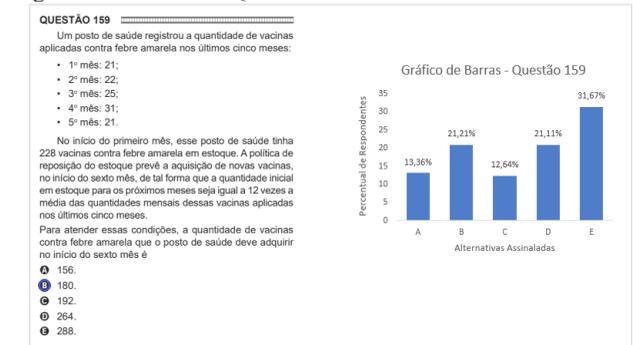
Classificação via TCT: Discriminação satisfatória e dificuldade moderada.

Classificação via TRI: Discriminação muito alta e dificuldade alta.
Comentário: Exige conhecimentos básicos de aritmética para o cálculo da média aritmética simples dos lucros de uma empresa

em um período de seis meses, a fim de determinar o lucro mínimo do sexto mês dessa empresa, considerando uma média de 30 mil reais mensais.

4.1.4 Questão 159

Figura 5 – Análise da Questão 159 do ENEM 2016



Fonte: Baseado nos microdados do ENEM (Brasil, 2016).

Quadro 7 – Parâmetros Psicométricos da Questão 159 do ENEM 2016

Hab	pBis	a	ep(a)	b	ep(b)	c	ep(c)	Chisq (p-valor)
30	0,3957	3,866	0,09	2,110	0,03	0,18	< 0,01	< 0,05

Fonte: Baseado nos microdados do ENEM (Brasil, 2016).

Assunto: Média Aritmética Simples.

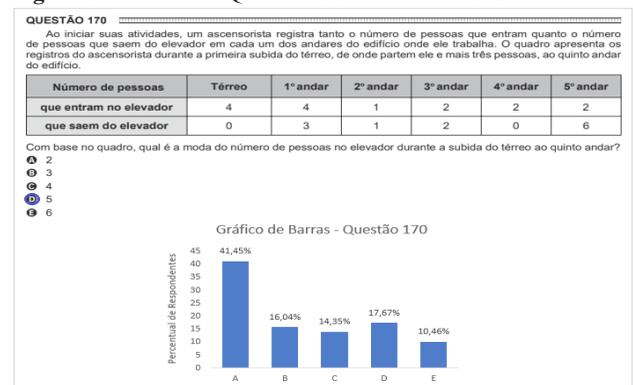
Classificação via TCT: Discriminação satisfatória e dificuldade alta.

Classificação via TRI: Discriminação muito alta e dificuldade muito alta.

Comentário: Exige conhecimentos básicos de aritmética para o cálculo da média aritmética simples das quantidades mensais de vacinas aplicadas em cinco meses. Entretanto, requer atenção na interpretação do enunciado em que é pedido a quantidade de vacinas que o posto de saúde deve adquirir, sendo necessário a subtração da quantidade já existente de vacinas. A maioria dos respondentes desta questão assinalaram a alternativa “E”, possivelmente por não ter se atentado a este fato.

4.1.5 Questão 170

Figura 6 - Análise da Questão 170 do ENEM 2016



Fonte: Baseado nos microdados do ENEM (Brasil, 2016).

Quadro 8 – Parâmetros Psicométricos da Questão 170 do ENEM 2016

Hab	pBis	a	ep(a)	b	ep(b)	c	ep(c)	Chisq (p-valor)
27	0,4227	3,345	0,07	1,608	0,01	0,19	< 0,01	< 0,05

Fonte: Baseado nos microdados do ENEM (Brasil, 2016).

Assunto: Moda.

Classificação via TCT: Discriminação satisfatória e dificuldade alta.

Classificação via TRI: Discriminação muito alta e dificuldade muito alta.

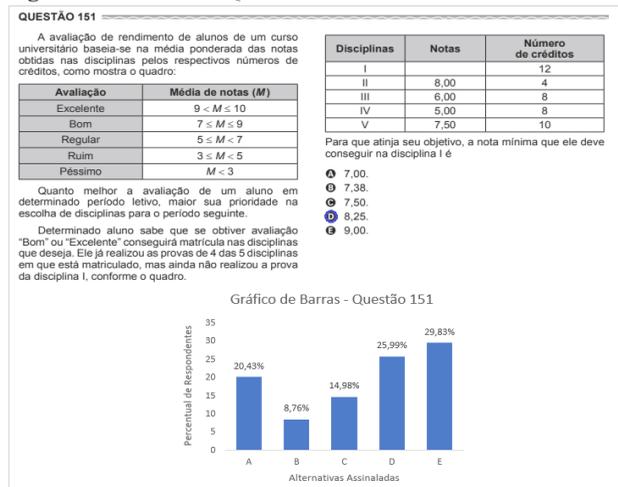
Comentário: Exige conhecimentos básicos de aritmética para o cálculo da moda do número de pessoas no elevador em cada andar do edifício. O estudante deve considerar, para cada andar, a quantidade de pessoas que entram e que saem do elevador. Observou-se que a maioria dos respondentes assinalaram a alternativa “A”, possivelmente pelo fato de que 2 é a quantidade mais frequente de pessoas que entram no elevador do edifício.

4.2 Análise das questões de Medidas de Tendência Central do ENEM 2017

Na prova de Matemática e suas Tecnologias do ENEM 2017 foram identificados 3 itens relacionados as Medidas de Tendência Central, dos quais 1 exigia o cálculo da média aritmética simples e 1 exigia o cálculo da média aritmética ponderada e 1 exigia o cálculo da mediana.

4.2.1 Questão 151

Figura 7 – Análise da Questão 151 do ENEM 2017



Fonte: Baseado nos microdados do ENEM (Brasil, 2017).

Quadro 9 - Parâmetros Psicométricos da Questão 151 do ENEM 2017

Hab	pBis	a	ep(a)	b	ep(b)	c	ep(c)	Chisq (p-valor)
30	0,3069	3,577	0,08	1,738	0,03	0,12	< 0,01	< 0,05

Fonte: Baseado nos microdados do ENEM (Brasil, 2017).

Assunto: Média Aritmética Ponderada.

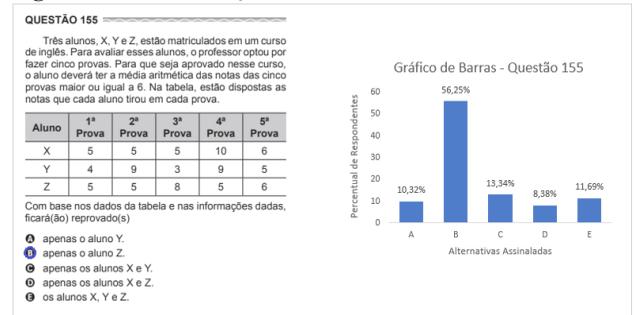
Classificação via TCT: Discriminação satisfatória e dificuldade alta.

Classificação via TRI: Discriminação muito alta e dificuldade muito alta.

Comentário: Exige conhecimentos básicos de aritmética para o cálculo da média aritmética ponderada das notas, cujos pesos são os números de crédito das disciplinas cursadas. Observou-se que a maioria dos estudantes assinalaram a alternativa “E”, possivelmente por apresentar o único valor, dentre os das alternativas, que faz com que a média aritmética simples das notas seja igual ou superior à 7, o que faz com que o aluno obtenha avaliação “Bom”.

4.2.2 Questão 155

Figura 8 - Análise da Questão 155 do ENEM 2017



Fonte: Baseado nos microdados do ENEM (Brasil, 2017).

Quadro 10 - Parâmetros Psicométricos da Questão 155 do ENEM 2017

Hab	pBis	a	ep(a)	b	ep(b)	c	ep(c)	Chisq (p-valor)
29	0,3463	0,943	0,03	-0,309	0,06	0,04	0,01	< 0,05

Fonte: Baseado nos microdados do ENEM (Brasil, 2017).

Assunto: Média Aritmética Simples

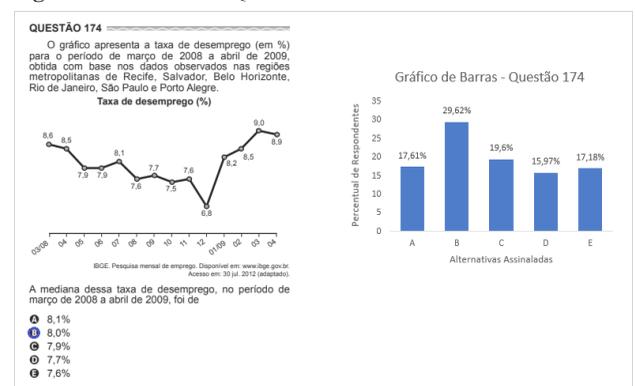
Classificação via TCT: Discriminação satisfatória e dificuldade moderada.

Classificação via TRI: Discriminação moderada e dificuldade baixa.

Comentário: Exige conhecimentos básicos de aritmética para o cálculo das médias aritméticas simples das notas dos três alunos, permitindo identificar o aluno que ficará reprovado.

4.2.3 Questão 174

Figura 9 – Análise da Questão 174 do ENEM 2017



Fonte: Baseado nos microdados do ENEM (Brasil, 2017).

Quadro 11 - Parâmetros Psicométricos da Questão 174 do ENEM 2017

Hab	pBis	a	ep(a)	b	ep(b)	c	ep(c)	Chisq (p-valor)
27	0,3524	1,742	0,08	1,501	0,04	0,14	< 0,01	< 0,05

Fonte: Baseado nos microdados do ENEM (Brasil, 2017).

Assunto: Mediana.

Classificação via TCT: Discriminação satisfatória e dificuldade alta.

Classificação via TRI: Discriminação muito alta e dificuldade muito alta.

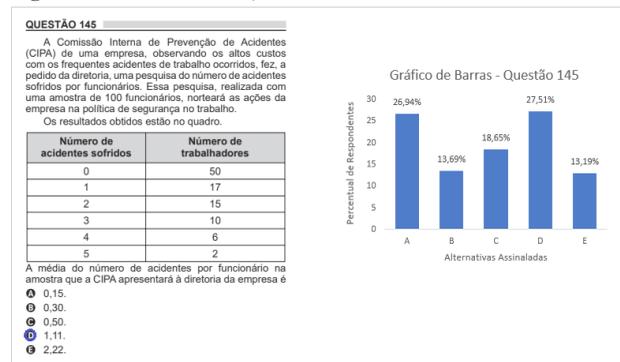
Comentário: Exige conhecimentos básicos de ordenação de valores e de média aritmética simples, considerando que há um número par de valores, implicando em dois termos centrais.

4.3 Análise das questões de Medidas de Tendência Central do ENEM 2018

Na prova de Matemática e suas Tecnologias do ENEM 2018 foram identificados 3 itens relacionados as Medidas de Tendência Central, dos quais 1 exigia o cálculo da média aritmética simples e 2 exigiam o cálculo da média aritmética ponderada.

4.3.1 Questão 145

Figura 10 – Análise da Questão 145 do ENEM 2018



Fonte: Baseado nos microdados do ENEM (Brasil, 2018).

Quadro 12 - Parâmetros Psicométricos da Questão 145 do ENEM 2018

Hab	pBis	a	ep(a)	b	ep(b)	c	ep(c)	Chisq (p-valor)
27	0,3475	2,245	0,08	1,595	0,03	0,16	< 0,01	< 0,05

Fonte: Baseado nos microdados do ENEM (Brasil, 2018).

Assunto: Média Aritmética Ponderada

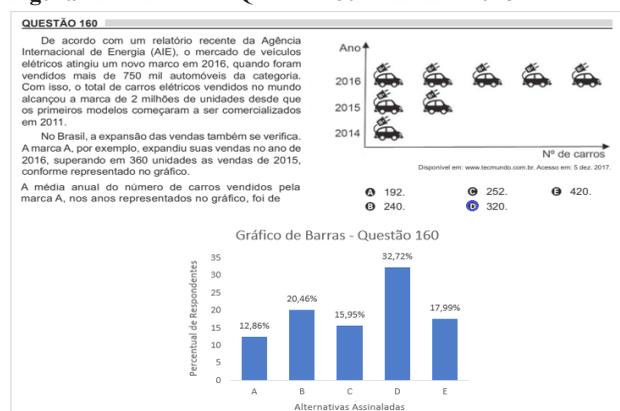
Classificação via TCT: Discriminação satisfatória e dificuldade alta.

Classificação via TRI: Discriminação muito alta e dificuldade muito alta.

Comentário: Exige conhecimentos básicos de aritmética para o cálculo da média aritmética ponderada de acidentes sofridos, considerando como pesos os números de trabalhadores. Observe a grande incidência de respondentes que assinalaram a alternativa “A”, possivelmente por calcularem a média aritmética simples, desconsiderando os pesos.

4.3.2 Questão 160

Figura 11 – Análise da Questão 160 do ENEM 2018



Fonte: Baseado nos microdados do ENEM (Brasil, 2018).

Quadro 13 - Parâmetros Psicométricos da Questão 160 do ENEM 2018

Hab	pBis	a	ep(a)	b	ep(b)	c	ep(c)	Chisq (p-valor)
27	0,4027	3,281	0,07	1,260	< 0,01	0,13	< 0,01	< 0,05

Fonte: Baseado nos microdados do ENEM (Brasil, 2018).

Assunto: Média Aritmética Simples.

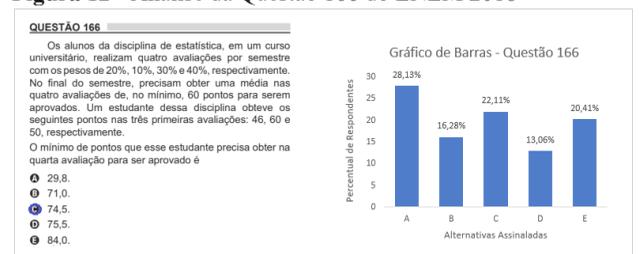
Classificação via TCT: Discriminação satisfatória e dificuldade moderada.

Classificação via TRI: Discriminação muito alta e dificuldade alta.

Comentário: Exige conhecimentos básicos de aritmética para o cálculo do valor representado por um carro no gráfico, para, em seguida, calcular a média anual do número de carros vendidos pela marca.

4.3.3 Questão 166

Figura 12 - Análise da Questão 166 do ENEM 2018



Fonte: Baseado nos microdados do ENEM (Brasil, 2018).

Quadro 14 - Parâmetros Psicométricos da Questão 166 do ENEM 2018

Hab	pBis	a	ep(a)	b	ep(b)	c	ep(c)	Chisq (p-valor)
28	0,3279	3,205	0,08	1,753	0,02	0,20	0,01	< 0,05

Fonte: Baseado nos microdados do ENEM (Brasil, 2018).

Assunto: Média Aritmética Ponderada.

Classificação via TCT: Discriminação satisfatória e dificuldade alta.

Classificação via TRI: Discriminação muito alta e dificuldade muito alta.

Comentário: Exige conhecimentos básicos de aritmética para o cálculo do número mínimo de pontos que o estudante precisa obter na quarta avaliação para ser aprovado, considerando a média aritmética ponderada de pontos e os pesos de , , e nas avaliações. A maioria dos estudantes assinalaram a alternativa “A”, possivelmente por considerarem os pesos para as três primeiras avaliações () e desconsiderarem o peso da última (). É digno de nota a expressiva quantidade de estudantes que assinalaram a alternativa “E”, que seria a alternativa correta caso não houvessem os pesos, ou seja, se o exercício fosse de média aritmética simples, ao invés de ponderada.

4.4 Discussão

Os itens analisados compreenderam as 4 Habilidades da Competência 7 da Matriz de Referência de Matemática e suas Tecnologias do ENEM. Destaca-se que a Habilidade 27, foi a mais frequente nas provas, com 5 itens a ela relacionados. As Habilidades 28, 29 e 30 estiveram presentes em 2 questões cada. Quanto a estrutura dos itens, percebeu-se um padrão em que os dados são frequentemente apresentados em quadros, tabelas ou gráficos. Para Curcio, estes suportes:

[...] são meios de comunicar e classificar dados, permitindo

a comparação e exibindo relações. Embora a leitura literal dos dados apresentados seja uma habilidade importante na compreensão desse tipo de representação, o seu potencial máximo será atingido quando o leitor for capaz de interpretar e generalizar os dados (Curcio, 1989, p.1).

A maioria desses itens tratavam da média aritmética simples (ou ponderada). Considerando os conceitos de mediana e de moda, foram identificados apenas 1 item para cada, referente aos anos de 2016 e 2017, respectivamente.

Quanto as respostas assinaladas pelos estudantes nos itens analisados, pôde-se perceber, através de uma análise de caráter pedagógico, que há um conflito entre os conceitos de média aritmética simples e ponderada. Dentre os estudantes que erraram as questões que envolviam a média aritmética ponderada, a maioria assinalou a alternativa que correspondia ao resultado dos cálculos que envolviam a média aritmética simples.

Este tipo de análise, embora não seja o foco deste trabalho, pode auxiliar os professores de e que ensinam Matemática a avaliarem as possíveis lacunas existentes na aprendizagem de estatística no Ensino Médio e constitui um campo vasto para a pesquisa em Educação Estatística. Além disso, permite aos docentes avaliarem se as metodologias de ensino por eles adotadas tem sido eficazes (Arteaga *et al.*, 2015).

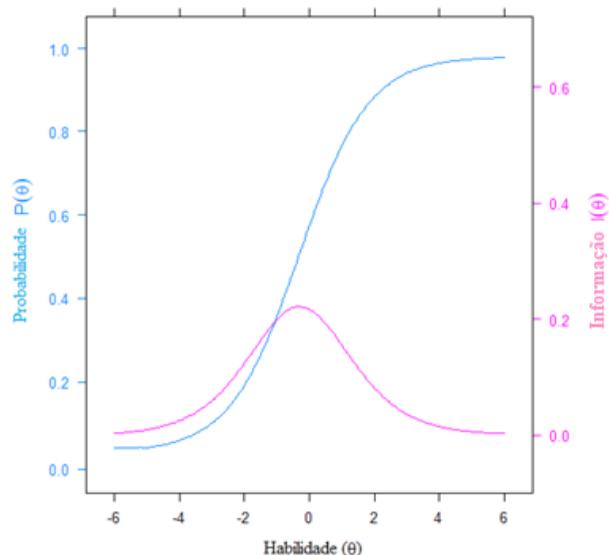
Quanto a análise da dificuldade dos itens via TCT, baseada na classificação proposta por Condé (2001), não houve itens classificados como fáceis. Dos 11 analisados, 5 foram classificados como moderados e 6 como difíceis. Sob a ótica da TRI, baseando-se na classificação proposta por Rabelo (2013), 1 item foi classificado como fácil, 2 itens como moderados, 2 itens como difíceis e 6 itens como muito difíceis. Ademais, observou-se que os índices de dificuldade estimados para todos os itens estavam dentro do intervalo -3 e 3, como esperado.

Quanto ao índice de discriminação, todos os itens possuem índice satisfatória, considerando a análise via TCT. Pela TRI, observou-se que a grande maioria das questões apresentaram discriminação muito alta. Também houve questões com discriminação classificada como alta () e moderada (). Esta análise indica um bom potencial para discriminar grupos com diferentes níveis de habilidades, tornando as questões eficientes na seleção de estudantes com maiores graus de conhecimento, o que constitui um dos objetivos do ENEM, considerando que ele tem se consolidado como a principal porta de acesso ao Ensino Superior no Brasil. A probabilidade de acerto ao acaso dos itens variou no intervalo compreendido entre 0,15 e 0,19, em 2016; 0,04 e 0,14, em 2017 e 0,13 e 0,20, em 2018.

O único item que classificado como fácil e com discriminação moderada, sob a ótica da TRI, foi o item 155 da edição de 2017 do ENEM. Este item também apresentou a menor probabilidade de acerto ao acaso para essa edição do exame: 0,04. A Figura 13 ilustra, em azul, o comportamento gráfico da função que associa a probabilidade de acerto do

referido item com seus parâmetros psicométricos, conforme Equação 1, denominado Curva Característica do Item (CCI). Além disso, em rosa, têm-se a Curva de Informação do Item (CII), relacionada à precisão com que se pode estimar o nível de conhecimento de um respondente (Andrade; Tavares & Valle, 2000). Ressalta-se que, de acordo com Pasquali (2018), as maiores informações são trazidas para os respondentes cuja habilidade encontra-se em torno do valor correspondente ao índice de dificuldade do item.

Figura 13 - Curvas Característica e de Informação da Questão 155 do ENEM 2017



Fonte: Baseado nos microdados do ENEM (Brasil, 2017).

É importante ressaltar, também, que os parâmetros dos itens utilizados pelo ENEM não são, necessariamente, os aqui estimados. A calibração dos itens do ENEM é feita baseada em uma escala construída com base nos resultados do ENEM 2009, o primeiro a adotar a metodologia TRI. Nas análises aqui realizadas, a escala foi construída com base nos valores da média aritmética e do desvio-padrão obtidos em cada amostra.

Não houve problemas quanto ao ajuste do modelo e as diferentes técnicas utilizadas se mostraram, em geral, coerentes entre si, confirmando o rigor dos métodos adotados neste trabalho. Em uma análise global, pôde-se perceber que os itens analisados apresentaram qualidade adequada para compor o exame.

5 Considerações Finais

As avaliações standardizadas têm se constituído como um importante diagnóstico para o ensino e a aprendizagem, se mostrando incisivas para a implementação de políticas públicas voltadas à melhoria da qualidade da educação. Considerando a importância que este tipo de avaliação tem assumido no cenário educacional, tornam-se imprescindíveis as análises referentes a sua estruturação e a qualidade de seus itens.

A Psicometria oferece uma ampla variedade de ferramentas dirigidas à estas análises. Em sua vertente moderna, considera a Teoria de Resposta ao Item como principal procedimento metodológico de construção e análise técnica de itens. A integração desta metodologia no ENEM, em 2009, possibilitou a seleção de itens mais adequados e informativos, considerando o contexto em que se deseja medir a proficiência do respondente.

Estas características garantem um exame mais confiável e eficiente, garantindo a qualidade dos resultados obtidos e a isonomia entre os participantes. Além disso, a adoção da TRI no ENEM permitiu a comparabilidade entre os resultados de respondentes submetidos a provas ou a edições diferentes do exame. Esta comparabilidade pode ser útil, por exemplo, para avaliar novas metodologias de ensino ou o comportamento de uma política pública adotada, ao longo dos anos.

Os cálculos que envolvem o índice de dificuldade dos itens permitem avaliar pontos críticos de aprendizagem que devem ser considerados no ensino do conteúdo. O alto valor para este índice indica que o assunto ao qual ele se refere deve ser cuidadosamente abordado pelos professores, que devem repensar suas práticas de ensino, muitas das vezes tradicional.

Já a análise da discriminação permite selecionar os itens que possuem maior competência em classificar os respondentes por nível de proficiência. Embora não seja um aspecto comum a ser analisado em sala de aula, é um componente essencial nas avaliações competitivas, que visam identificar os respondentes com maior grau de conhecimento em determinada área, como olimpíadas de conhecimento ou exames vestibulares.

A análise do conteúdo dos itens, por sua vez, pode auxiliar os professores de e que ensinam Matemática a enxergarem as dificuldades pedagógicas dos estudantes, apontando possíveis lacunas existentes na aprendizagem da disciplina. Além disso, permite aos docentes avaliarem se as metodologias de ensino por eles adotadas tem sido eficazes.

Como meio de exemplificar a aplicação dessas técnicas de análises de itens, este trabalho se propôs a analisar a qualidade de 11 itens referentes às Medidas de Tendência Central que estiveram presentes no ENEM de 2016 a 2018. Embora introdutório, espera-se que os resultados obtidos possam esclarecer alguns conceitos da psicometria e auxiliar os professores de Matemática no que se refere ao ensino de destes tópicos da estatística na educação básica. Também espera-se que este trabalho sirva como um instrumento de difusão dessas técnicas na análise da qualidade de itens, em todas as áreas do conhecimento.

Referências

- Alnabhan, M. & Harwell, M. (2001). Psychometric challenges in developing a college admission test for Jordan. *Social Behavior and Personality: an international journal*, Palmerston North, NZ, 28 (5), 445-458. doi: <https://doi.org/10.2224/sbp.2001.29.5.445>
- Andrade, D. F., Tavares, H. R. & Valle, R. C. (2000, fevereiro). Teoria da resposta ao item: conceitos e aplicações. In: 14., 2000, Caxambu. *Anais do Simpósio Nacional de Probabilidade e Estatística*, Caxambu, MG, Brasil, 14. Recuperado de: https://docs.ufpr.br/~aanjos/CE095/LivroTRI_DALTON.pdf
- Arteaga, P., Batanero, C., Gea, M. M. & Contreras, J. M. (2015)., Conocimiento de la estadística y los estudiantes en futuros profesores: un estudio exploratorio. *Journal Internacional de Estudos em Educação Matemática*, 8 (3), 98-125. doi: <https://doi.org/10.17921/2176-5634.v8n3>
- Baker, F. B. (2001). *The basics of item response theory*. ERIC Clearinghouse on Assessment and Evaluation, University of Maryland, College Park, MD.
- Bock R. D. (1972). Estimating item parameters and latent ability when responses are scored in two or more nominal categories. *Psychometrika*, New York, 37 (1), 29-51. doi: <https://doi.org/10.1007/BF02291411>
- Brasil. (1998). Secretaria de Educação Fundamental. *Parâmetros Curriculares Nacionais: Matemática*. Secretaria de Educação Fundamental. Brasília: MEC / SEF.
- Brasil. (2002). *Parâmetros Curriculares Nacionais Para o Ensino Médio: Orientações Educacionais Complementares aos Parâmetros Curriculares Nacionais: Ciências da Natureza, Matemática e suas Tecnologias*. MEC-SEMTEC, Brasília.
- Brasil. (2009). Ministério da Educação. Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira. *Matriz de Referência para o ENEM*. Brasília: INEP/MEC. Recuperado de: http://download.inep.gov.br/download/enem/matriz_referencia.pdf
- Brasil. (2016-2018). Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira. *Microdados do Enem 2016-2018*. Brasília: INEP, 2018. Disponível em: <http://portal.inep.gov.br/web/guest/microdados>.
- Chalmers, R. P. (2012). mirt: A multidimensional item response theory package for the R environment. *Journal of Statistical Software*, 48 (6), 1-29. doi: <https://doi.org/10.18637/jss.v048.i06>
- Chernyshenko, O. S., Stark, S., Chan, K. Y., Drasgow, F., & Williams, B. (2001). Fitting item response theory models to two personality inventories: Issues and insights. *Multivariate Behavioral Research*, 36(4), 523-562. doi: http://doi.org/10.1207/S15327906MBR3604_03
- Costa, M. C., Lima, S. H. O., & Soares, D. J. M. (2020). Uma proposta de análise de itens da prova preparatória para o Enade aplicada aos discentes de engenharia civil do IFMG – Campus Avançado Piumhi. *ForScience*, 8(1), e00706. <http://doi.org/10.29069/forscience.2020v8n1.e706>
- Curcio, F. R. (1989). *Developing graph comprehension: elementary and middle school activities*. Reston, VA: NCTM.
- Drasgow, Fritz, e Lissak, Robin L. (1983). Modified parallel analysis: A procedure for examining the latent dimensionality of dichotomously scored item responses. *Journal of Applied Psychology*, 68(3), 363-373. doi: <http://doi.org/10.1037/0021-9010.68.3.363>
- Fernandes, J. A., Sousa, M. V. & Ribeiro, S. A. (2004). O ensino de estatística no ensino básico e secundário: Um estudo exploratório. In: Fernandes, J. A., Sousa, M. V. & Ribeiro, S. A. (Orgs.). *Ensino e Aprendizagem de Probabilidades e Estatística*. (Atas do 1.º Encontro de Probabilidades e Estatística na Escola, pp. 165-193.). Braga: Centro de Investigação em Educação da Universidade do Minho.
- Hair Jr., Joseph F., Rolph E. Anderson, Ronald L. Tatham e

- William C. Black. (2005). *Análise multivariada de dados* (5a ed.). Porto Alegre: Bookman.
- Hambleton, R. K., Swaminathan, H. & Rogers, H. J. (1991). *Fundamentals of item response theory*. Newbury Park: Sage Publications.
- Moreira, J. V. B., Rosa, L. M. F., Soares, D. J. M., & Soares, T. E. A. (2020). Análise Clássica de Avaliações: Um Estudo de Caso/Test Classical Analysis: a case study. *Brazilian Journal of Development*, 6(2), 7962-7970. doi: <http://doi.org/10.34117/bjdv6n2-197>
- Muñiz, J. (1990). *Teoría de respuesta a los ítems: Un nuevo enfoque en la evolución psicológica y educativa*. Madrid: Ediciones Pirámide, S. A.
- Muthén, B. O., Kao, C. F. & Burstein, L. (1991). Instructionally Sensitive Psychometrics: Application of a New IRT – Based Detection Technique to Mathematics Achievement Test Items. *Journal of Educational Measurement*, New Jersey, 28 (1), 1-22. doi: <https://doi.org/10.1111/j.1745-3984.1991.tb00340.x>
- Pasquali, Luiz, & Primi, Ricardo. (2003). Fundamentos da teoria da resposta ao item: TRI. *Avaliação Psicológica*, 2(2), 99-110. Recuperado em 08 de abril de 2020, de http://pepsic.bvsalud.org/scielo.php?script=sci_arttext&pid=S1677-04712003000200002&lng=pt&tln=pt.
- Pasquali, Luiz. (2009). Psicometria. *Revista da Escola de Enfermagem da USP*, 43(spe), 992-999. doi: <https://doi.org/10.1590/S0080-62342009000500002>
- Pasquali, Luiz. (2018). *TRI–Teoria de resposta ao item: Teoria, procedimentos e aplicações*. Curitiba: Editora Appris.
- R Core Team. (2020). *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing. www.R-project.org/.
- Rabelo, Mauro. (2013). *Avaliação educacional: fundamentos, metodologia e aplicações no contexto brasileiro*. Rio de Janeiro: SBM, 29, 30-31.
- Revelle, W. & Revelle, M. W. (2015). *Package ‘psych’*. The Comprehensive R Archive Network, 2015.
- Rizopoulos, D. (2006). ltm: An R package for latent variable modeling and item response theory analyses. *Journal of statistical software*, 17(5), 1-25. doi: <https://doi.org/10.18637/jss.v017.i05>
- Soares, D. J. M. (2018) *Teoria clássica dos testes e teoria de resposta ao item aplicadas em uma avaliação de Matemática básica* (Dissertação de Mestrado). Universidade Federal de Viçosa. Viçosa, MG, Brasil. Recuperado de: <https://www.locus.ufv.br/handle/123456789/18404>
- Tôrres, F. C. (2015). *Uma aplicação da teoria de resposta ao item em um simulado de Matemática no modelo ENEM* (Dissertação de Mestrado). Programa de Mestrado Profissional em Matemática em Rede Nacional, Universidade de Brasília, Brasília, Brasil. Recuperado de: <https://repositorio.unb.br/handle/10482/19102>
- Travitzki, R. (2017). Avaliação da qualidade do Enem 2009 e 2011 com técnicas psicométricas. *Estudos em Avaliação Educacional*, 28(67), 256-288. doi: <http://dx.doi.org/10.18222/eae.v28i67.3910>.