

# A psicometria moderna como viés para reflexões no Ensino de Química

## Modern psychometrics as a bias for reflections in Chemistry Teaching

Vinícius da Silva Carvalho<sup>1</sup>, Andréia Francisco Afonso<sup>2</sup>

**RESUMO:** As avaliações de larga escala no contexto educacional têm como um dos principais objetivos investigar o desenvolvimento de habilidades e competências do estudante por meio de itens. O teste contínuo e consolidado na educação brasileira, empenhado em avaliar como estudantes concluintes do Ensino Médio as mobilizam em situações-problema é o Exame Nacional do Ensino Médio. Porém, esta é uma prática pouco explorada por profissionais da educação. Desse modo, o presente artigo tem por objetivo apresentar informações que podem ser obtidas a partir dos atributos psicométricos de um item de Química da prova de Ciências da Natureza do referido exame, aplicados no ano de 2017. Para tanto, utilizamos o método da triangulação para confrontar os dados da base de microdados referentes aos itens com sua respectiva análise pedagógica. Por fim, concluímos que a leitura pedagógica, a partir de dados psicométricos, sugere ao professor e ao gestor, suporte para reflexões sobre o ensino ao indicar a dificuldade dos estudantes, além de subsidiar apoio para pesquisas futuras que visem conhecer a aprendizagem dos estudantes em Química e assim, propor ações para melhorias no processo educacional.

**PALAVRAS-CHAVE:** Teoria de Resposta ao Item, Química, Enem

### ABSTRACT:

One of the main objectives of large-scale assessments in the educational context is to investigate the development of student skills and competencies through items. The continuous and consolidated test in Brazilian education, committed to evaluating how high school graduates mobilize them in problem situations, is the National High School Exam. However, this is a practice little explored by education professionals. Therefore, this article aims to present information that can be obtained from the psychometric attributes of a Chemistry item from the Natural Sciences test of that exam, applied in 2017. To this end, we used the triangulation method to compare the data from the microdata base relating to the items with their respective pedagogical analysis. Finally, we conclude that the pedagogical reading, based on psychometric data, suggests to the teacher and manager support for reflections on teaching by indicating students' difficulties, in addition to providing support for future research aimed at understanding student learning in Chemistry and thus, propose actions to improve the educational process.

**KEYWORDS:** Item Response Theory, Natural Sciences/Chemistry, Enem

### INTRODUÇÃO

<sup>1</sup> Universidade Federal de Juiz de Fora (UFJF), Aluno de Pós-Graduação do Departamento de Química da UFJF, <http://orcid.org/0000-0002-4132-4736>, e-mail: [vinicius.scq@gmail.com](mailto:vinicius.scq@gmail.com).

<sup>2</sup> Universidade Federal de Juiz de Fora (UFJF), Professora Adjunta: <http://orcid.org/0000-0002-0107-2490>; e-mail: [andrea.afonso@ufjf.br](mailto:andrea.afonso@ufjf.br)

Com o avanço da tecnologia, as ferramentas, cada vez mais sofisticadas, têm ocupado espaço na atuação de diversas profissões e servido de auxílio para aprimorar as estratégias utilizadas, como, por exemplo, os diagnósticos na medicina que estão cada vez mais precisos. As ferramentas de tecnologias de informação e comunicação também avançam no campo educacional e adentram a escola de variadas formas. Uma delas, a ser tratada neste artigo, possibilita conhecer os resultados tratados pela psicomетria moderna, como a Teoria de Resposta ao Item (TRI), que é aplicada nas avaliações de grande escala (PASQUALI, 2017).

As avaliações de grande escala, também conhecidas como avaliações externas, requerem o uso de metodologias e ferramentas específicas de análise que possibilitam a manutenção da comparabilidade e confiabilidade dos resultados (GATTI, 2016). Essas características são importantes à medida que os resultados permitem acompanhar o desempenho dos estudantes ao longo dos anos. Porém, essas ferramentas estão distantes do conhecimento do professor, principal agente de mudanças na sala de aula.

Um dos motivos para esse distanciamento é a ausência de um processo formativo para os professores – seja inicial ou continuado - para obtenção e leitura de informações e dos microdados. É relevante esse conhecimento, pois, a partir dos dados, são fornecidas informações sobre o desempenho dos estudantes e a sua proficiência média, como intenção de acompanhar a qualidade da educação nas diferentes redes e sistemas de ensino.

Apesar da destacada relevância, as escolas recebem um relatório pedagógico contendo informações estatísticas com números e gráficos, para que os professores tenham conhecimento sobre o ensino ofertado. A partir dele, pode ocorrer a reflexão sobre ações que se direcionem para a melhoria da qualidade da aprendizagem do educando. Contudo, sem formação para leitura desses dados, a continuidade dos trabalhos, a progressão do ensino e a implementação de políticas públicas mais assertivas ficam prejudicadas.

A partir desse contexto, compreendemos a necessidade de instrumentalizar os docentes, independente da área de atuação, para a leitura dos dados estatísticos associados a uma análise pedagógica. Entretanto, neste artigo, o foco central são os docentes de Química, que compõe a área de Ciências da Natureza, embora também seja útil para professores de Biologia e Física.

O principal motivo se fundamenta nas avaliações de grande escala em Ciências que, por vezes, não apresentaram continuidade na divulgação de resultados ao longo da história da educação brasileira, devido à ausência de uma matriz de referência consolidada (MEDEIROS et al., 2017). Diante desse fato, é possível inferir que os docentes dessa área, quando participam de discussões sobre os resultados de avaliações em grande escala, ou, como se costuma traduzir no Brasil, de larga escala, partem de experiências de outras áreas já consolidadas, como a Matemática e a Língua

Portuguesa (GARCIA et al., 2019), o que pode contribuir pouco para a reflexão sobre a própria prática docente.

Atentos a implementação da Meta 7.7 do Plano Nacional de Educação para o período de 2014 a 2024, percebemos que a falta de interação do professor com os resultados de desempenho dos estudantes vai de encontro a proposta no documento, no qual há o seguinte compromisso:

[...] aprimorar continuamente os instrumentos de avaliação da qualidade do ensino fundamental e médio, de forma a englobar o ensino de ciências nos exames aplicados nos anos finais do ensino fundamental, e incorporar o Exame Nacional do Ensino Médio, assegurada a sua universalização, ao sistema de avaliação da educação básica, bem como apoiar o uso dos resultados das avaliações nacionais pelas escolas e redes de ensino para a melhoria de seus processos e práticas pedagógicas (BRASIL, 2014).

À medida que os objetivos educacionais em Ciências da Natureza se tornam mais definidos com a implementação da Base Nacional Comum Curricular (BNCC) do Ensino Médio em 2018, surge um cenário propício para avanços significativos na educação científica no Brasil. O Ministério da Educação (MEC) tem demonstrado um interesse crescente em elevar a proficiência dos estudantes brasileiros no Programa Internacional de Avaliação de Estudantes (PISA). Contudo, observa-se um descompasso no cumprimento da Meta 7 do Plano Nacional de Educação (PNE), que visa melhorar a qualidade do ensino.

Nesse contexto, antecipa-se que, na próxima década, haverá um esforço contínuo por parte do governo federal para manter e, possivelmente, expandir as avaliações de grande escala em Ciências no Brasil. Essa perspectiva coloca em destaque a necessidade de preparar adequadamente os professores de Ciências da Natureza, em particular os de Química, ~~e os docentes de Ciências da Natureza, em geral~~, para interpretar e utilizar os dados e resultados apresentados nos relatórios de feedback das avaliações. Mas, para isso, é preciso preparar os licenciados para a compreensão dos dados e resultados presentes nos relatórios de devolutivas.

Essa questão suscita discussões interessantes sobre a criação da TRI, já que ela pode se transformar em uma ferramenta educacional poderosa, que possibilita prever aspectos sobre a capacidade cognitiva dos estudantes (SPAZZIANI, 2019, CARVALHO; AFONSO, 2023). Assim, propomos, nesse artigo, contemplar dois objetivos, que serão apresentados em dois momentos. No primeiro, faremos uma breve apresentação sobre o processo de desenvolvimento das métricas envolvidas em testes educacionais, uma vez que este está diretamente relacionado à busca pelo processo de aprendizagem. Já no segundo momento, discorreremos sobre a metodologia envolvida na TRI, a partir de uma sequência de eventos traçados pelos autores deste artigo, tratando do modelo logístico de três parâmetros, o mais sofisticado e utilizado por avaliações consolidadas na

atualidade, como o Exame Nacional do Ensino Médio (Enem). Por fim, o artigo propõe uma análise das possíveis informações que podem ser extraídas de um item aplicado na edição do Enem de 2017.

## PRIMEIRO MOMENTO

Iniciaremos essa seção com os estudos realizados no século XX, por meios dos quais foi possível se chegar ao conceito de inteligência, que recebeu contribuição expressiva dos trabalhos de Alfred Binet e Théodore Simon, desencadeando pesquisas sobre a coleta de informações em testes psicométricos e em avaliações escolares. (BINET; SIMON, 1904)

A motivação para os trabalhos de Alfred Binet e Théodore Simon surge com a preocupação do governo francês em identificar crianças com dificuldade na aprendizagem. O governo francês, a partir dos resultados das avaliações em grande escala, discriminava os estudantes por comparação e os classificava como crianças normais e subnormais. As subnormais eram divididas em categorias, sendo uma delas a dos estudantes “*unstable*”, aqueles mais rebeldes que não seguiam ordens, e a outra, dos alunos “*backward intelligence*” que entendemos ser os estudantes que apresentavam alguma dificuldade na aprendizagem. (BINET; SIMON, 1904)

Para oferecer condições diferenciadas de ensino, Binet e Simon (1904) resignificaram o conceito de inteligência e afirmaram que os indivíduos a apresentam de formas diferenciadas, direcionadas às funções mentais complexas, como por exemplo: memória, imaginação, atenção, compreensão ou apreciação estética. Ao considerar esses aspectos, eles criaram a Escala de Inteligência Binet-Simon (TEIXEIRA, 2019).

A Escala de Inteligência Binet-Simon era constituída por 30 itens, distribuídos por ordem crescente de dificuldade e dirigida a crianças dos 3 aos 12 anos. Os itens foram agrupados de acordo com a proporção de acertos das crianças por faixas etárias (ALMEIDA, 1988) e o método foi ajustado para incluir as manifestações de inteligência.

O método psicológico que temos desenvolvido é composto de provas de dificuldade crescente que nós apresentamos em tópicos aos sujeitos. Estes testes foram todos examinados e testados, e foi possível perceber as causas dos erros que eles apresentam e do significado dos resultados obtidos. [...] A série de testes é ajustada de forma a incluir a totalidade da gama de manifestações intelectuais, desde o mais baixo, o mais elementar, o que parece ser uma fixação constante em um objeto em movimento (BINET; SIMON, 1905, p. 508-509).

Porém, os trabalhos de Thurstone (1928, p.547) apontavam que a teoria praticada para verificar a inteligência humana apresentava graves problemas, por ser “um instrumento de medida,

na sua função de medir, não pode ser seriamente afetado pelo objeto de medida. Na extensão em que sua função de medir for assim afetada, a validade do instrumento é prejudicada ou limitada”. (Tradução nossa).

O instrumento que realizava essa medida, e que esteve presente em testes até 1950, é conhecido hoje, pelos pesquisadores, como Teoria Clássica dos Testes (TCT) (PRIMI; PASQUALI, 2003). Quando a TCT é aplicada em um teste educacional, os resultados são estimados a partir da soma dos acertos de cada item que compõe o instrumento de avaliação.

Segundo Sartes e Souza–Formigoni (2013), a partir dessa teoria, foi possível operacionalizar as duas principais propriedades psicométricas dos instrumentos: a validade e a confiabilidade. Para as autoras, a validade de um instrumento pode ser definida como a capacidade do teste em medir aquilo a que se propõe. Já a confiabilidade que lhe é dada, é o grau de concordância entre as múltiplas medidas de desempenho de um mesmo sujeito, sendo possível compará-las inter e intra indivíduos.

Esses estudos reforçavam os problemas apontados por Thurstone antes da década de 1930, ou seja, que o instrumento construído ainda dependia do objeto medido. Pasquali (2020) esclarece que na TCT, os resultados que permitem aferir a inteligência do sujeito seriam diferentes se utilizássemos, por exemplo, as Matrizes Progressivas de Raven (SPM) ou Wechsler Adult Intelligence Scales (Wais).

Contudo, o mesmo autor explica que a superação desse problema surge a partir do desenvolvimento da Teoria do Traço Latente, desenvolvida por Lord (1952) e com a contribuição dos estudos de Rasch (1960). Para Pasquali (2020), a demora para solucionar a independência do instrumento de medida com o objeto, e iniciar uma nova teoria que concentrava os dados na resposta ao item (TRI), estava na complexidade de algoritmos matemáticos que os modelos dessa teoria exigiriam. A tecnologia computacional da época era insuficiente para resolvê-lo.

Mas com a evolução da tecnologia e a disponibilidade de *softwares* cada vez mais sofisticados, a TRI ganha destaque e passa a compor o cenário nas avaliações de grande escala na década de 1990. Portanto, a TRI passa a ser a teoria do traço latente aplicada para verificar a habilidade ou desempenho de um sujeito. Pascoali e Primi (2003) compreendem a teoria do traço latente como:

Uma família de modelos matemáticos que relaciona variáveis observáveis (itens de um teste, por exemplo) e traços hipotéticos não-observáveis ou aptidões, estes responsáveis pelo aparecimento das variáveis observáveis ou, melhor, das respostas ou comportamentos emitidos pelo sujeito que são as variáveis observáveis. (p.102)

Com base nesse conceito, é possível estimar a habilidade ou o nível do traço latente de um sujeito se conhecermos as características das variáveis observáveis, ou seja, a resposta dada por esse sujeito em um item, pois essa se torna constante na equação e, portanto, solucionável.

Seguimos para a próxima seção, onde abordaremos com um pouco mais de detalhes a TRI.

## SEGUNDO MOMENTO

Segundo Soares (2005), dentre as vantagens de se utilizar a TRI sobre as técnicas clássicas estão:

A melhor análise de cada item que constitui o instrumento de avaliação (ou medida), considerando suas características estatísticas específicas na produção das escalas, como as que medem a capacidade de discriminar os indivíduos e as dificuldades dos itens; facilita, também, a interpretação da escala produzida e permite conhecer, diretamente, quais itens estão produzindo a informação gerada ao longo do *continuum* de valores, segundo o qual a escala é construída, tipicamente o conjunto dos números reais; além disso, a TRI permite a comparabilidade dos resultados produzidos para grupos de indivíduos diferentes, mesmo quando instrumentos (parcialmente) diferentes são aplicados; finalmente, a TRI permite um tratamento natural de casos com dados faltantes (p.84)

Além dessas vantagens, Pasquali (2020) apresenta três principais avanços que a TRI trouxe para o tratamento dos resultados nos testes educacionais: 1- A possibilidade de calcular o nível de aptidão do sujeito, independente da amostra de itens utilizados, ou seja, *not test-dependent*, que significa que a habilidade do sujeito não depende do teste; 2- Reconhecer a dificuldade do item e estabelecer uma discriminação entre estudantes aptos dos inaptos, assim, se diz ter *not group-dependent*, ou seja, os parâmetros são independentes dos sujeitos; 3- É possível emparelhar itens com as aptidões dos sujeitos para produzir uma escala de comparação.

Diante dessas reconhecidas vantagens, a TRI é um conjunto de modelos matemáticos que busca associar a probabilidade de um indivíduo acertar determinado item aos seus traços latentes, em determinada área de conhecimento que se pretende medir. Dessa forma, o modelo garante a unidimensionalidade quando apenas uma aptidão responsável pela realização da tarefa (item) será avaliada. Logo, há a independência do desempenho do estudante na resolução de uma tarefa em relação ao desempenho em outra, visto que são estatisticamente independentes.

Os itens que avaliam apenas uma habilidade são dicotômicos e quando aplicados a um único tipo de população, como no caso do Enem, que avalia os estudantes concluintes do Ensino Médio, é possível utilizar o modelo logístico unidimensional de três parâmetros (ML3). O modelo matemático ML3 será mais detalhado em relação aos outros modelos, nesse artigo, por ser o mais utilizado na atualidade (ANDRADE et al., 2000) e por possibilitar a instrução de uma leitura de

dados mais atualizada. Andrade et. al. (2000) definem o modelo matemático ML3 pela seguinte expressão:

$$P_{(U_{ij}=1|\theta_j)} = ci + (1 - ci) \frac{1}{1 + e^{-Dai(\theta_j - bi)}}$$

Considerando  $i = 1, 2, 3, \dots, I$ , e  $j = 1, 2, 3, \dots, n$ , sabe-se que:

$U_{ij}$  é uma variável dicotômica que assume os valores 1, quando o indivíduo  $j$  responde corretamente o item  $i$ , ou 0 quando o indivíduo  $j$  não responde corretamente ao item  $i$ .

$\theta_j$  representa a habilidade (traço latente) do  $j$ -ésimo indivíduo.

$P(U_{ij} = 1|\theta_j)$  é a probabilidade de um indivíduo  $j$  com habilidade  $\theta_j$  responder corretamente o item  $i$  e é chamada de Função de Resposta do Item – FRI.

$a_i$  é o parâmetro de discriminação do item  $i$ .

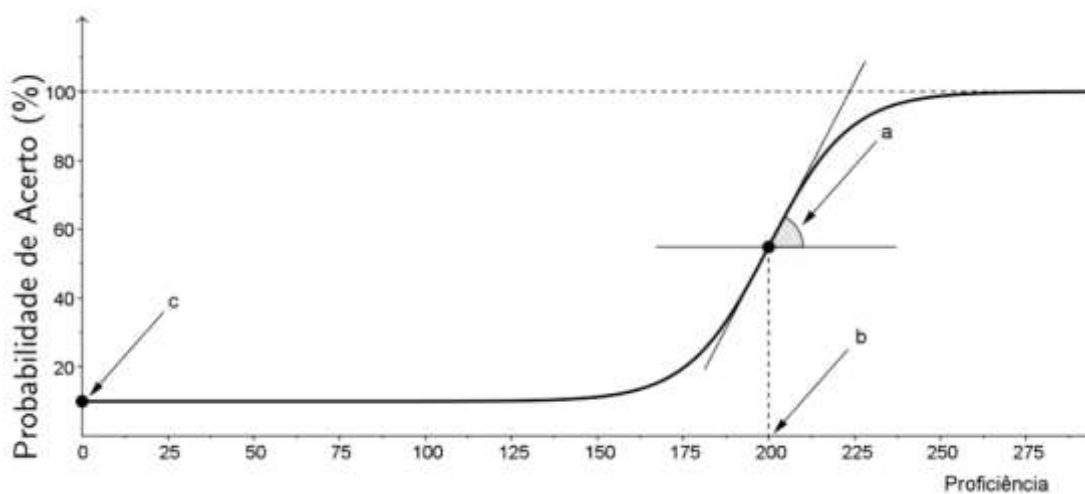
$b_i$  é o parâmetro de dificuldade (ou de posição) do item  $i$ , medido na mesma escala da habilidade.

$c_i$  é o parâmetro do item que representa a probabilidade de indivíduos com baixa habilidade responderem corretamente o item  $i$  (muitas vezes, referido como a probabilidade de acerto casual).

$D$  é um fator de escala, constante e igual a 1. Utiliza-se o valor 1,7 quando se deseja que a função logística forneça resultados semelhantes ao da função ogiva normal.

Em uma representação gráfica, o funcionamento desse modelo de três parâmetros é representado pela Curva Característica do Item (CCI) (Gráfico 1).

Gráfico 1: Elucidação para explicar a CCI modelo logístico de três parâmetros



Fonte: Carlos (2016, p. 38).

Com base nessa representação gráfica, é possível observar que quanto maior for a discriminação do item (parâmetro  $a$ ), maior será a inclinação da curva. Isso significa que quanto maior esse parâmetro, melhor será a distinção entre as probabilidades de dois indivíduos com diferentes proficiências acertarem o item. Já a dificuldade do item (parâmetro  $b$ ) está associada à proficiência do estudante, portanto, item com maior nível de dificuldade exige nível elevado de proficiência por parte do estudante. Por fim, o parâmetro  $c$ , que indica a possibilidade de um estudante com baixa proficiência acertar o item, pode ser observado na altura onde a função se inicia, nesse gráfico, por exemplo,  $c = 10\%$ . Portanto, quanto menor for o acerto ao acaso em um teste, maior será a quantidade de informações que o item irá oferecer. Andrade et. al (2000) concluem que a quantidade de informações que se pode obter vai depender se, o valor da proficiência se aproxima do parâmetro  $b$  do item, se o valor de discriminação for elevado e o valor do parâmetro  $c$  for baixo.

Vale ressaltar que nos testes educacionais produzidos por itens dicotômicos, ou seja, que admitam apenas duas opções de resposta, certa ou errada, os autores também comentam sobre o coeficiente de correlação bisserial. Carlos (2017) afirma que a leitura da bisserial possibilita estimar se os estudantes que acertam os itens são os que demonstram melhor desempenho ao longo do teste. Para atestar a informação do desempenho do estudante ao longo do teste, o autor diz que é realizado um cálculo da correlação ponto-bisserial que aponta, em caso de valor negativo, que há possibilidade de haver problemas técnicos ou conceituais no item, visto que os estudantes com melhor desempenho estão sendo atraídos para alternativas erradas.

Luckesi (2018) garante que a leitura da curva gerada a partir de dados estatísticos pode e deve ser um recurso fundamental para o gestor de uma ação, pois revela o retrato do resultado obtido, servindo como meio para motivar uma investigação avaliativa. Tratando-se de disciplinas como as de Ciências da Natureza, que caminha por uma base de dados descontínua, na perspectiva de estimar o desempenho de estudantes brasileiros em Biologia, Física e Química, por meio da TRI, consideramos relevantes os dados de avaliações mais consolidadas nessa área, como o Enem.

## **A LEITURA DA PROFICIÊNCIA E DO DESEMPENHO DOS ESTUDANTES EM CIÊNCIAS DA NATUREZA NO ENEM VIA TRI**

No Brasil, a TRI é usada desde 1995 nas provas do Sistema Nacional de Avaliação da Educação Básica (Saeb), que mede o desempenho de estudantes do Ensino Fundamental e Médio. A partir de 2009, ela começou a ser aplicada para tratar os resultados do Enem com o objetivo de garantir a comparação das notas do Exame daquele ano com os seguintes.



Cientes que os termos habilidades e competências são correntes em documentos educacionais, como nos currículos, documentos de referência curriculares (como a Base Nacional Comum Curricular), nas matrizes de referência e relatórios pedagógicos de avaliações de grande escala, é importante destacar as suas contribuições para a compreensão dos dados estatísticos. Por isso, é necessário ampliar a discussão sobre esses dois conceitos.

Segundo Brito et al. (2000), a habilidade é uma característica psicológica singular, que trabalha em favor dos domínios de uma determinada atividade, de forma rápida e fácil. Assim, na área educacional, o estudante demonstra possuir habilidade ao atribuir respostas corretas frente a uma situação-problema, que envolve a mobilização de conhecimentos. Já as competências, os autores entendem como a realização de tarefas que requerem habilidades cognitivas específicas.

Primi et al. (2001) argumentam que a qualidade das experiências de aprendizagem ao longo da trajetória acadêmica de um estudante é crucial na transformação de habilidades em competências. Portanto, pela TRI, é possível estimar essas informações por meio dos resultados de proficiência e de desempenho. Segundo os mesmos autores, a proficiência designa a elevada aptidão do estudante em uma certa habilidade. Já o desempenho está intimamente ligado à proficiência e representa o grau/nível da realização de tarefas que requerem do estudante o desenvolvimento de habilidades cognitivas específicas. Dessa forma, a apresentação de uma escala de proficiência, representa os marcos do desempenho dos estudantes em uma determinada área do conhecimento, ao longo de um teste de proficiência.

Nesse sentido, surge a ideia dos níveis âncoras e a técnica conhecida como ancoragem, definida como níveis âncora, os pontos selecionados na escala de proficiência, que são possíveis de serem interpretados pedagogicamente (ANDRADE et al., 2000).

## **METODOLOGIA E O TRATAMENTO DOS DADOS**

Para a interpretação pedagógica de um item do Enem da edição de 2017, houve a necessidade da integração das abordagens quantitativas e qualitativas para compor a natureza deste estudo. Para isso, foi necessário renunciarmos a um método único e utilizar o método da triangulação, também conhecido por métodos mistos. Sobre os métodos mistos, Creswell e Clark (2015), os caracterizam pela mistura de dois procedimentos metodológicos, nos quais:

[...] o pesquisador coleta e analisa de modo persuasivo e rigoroso tanto os dados qualitativos quanto os quantitativos (tendo por base as questões de pesquisa); mistura (ou integra ou vincula) as duas formas de dados concomitantemente, combinando-os (ou misturando-os) de modo sequencial, fazendo um construir o outro ou incorporando um no outro; dá prioridade a uma ou a ambas as formas de

dados (em termos do que a pesquisa enfatiza); usa esses procedimentos em um único estudo ou em múltiplas fases de um programa de estudo; estrutura esses procedimentos de acordo com visões de mundo filosóficas e lentes teóricas; e combina os procedimentos em projetos de pesquisa específicos que direcionam o plano para a condução do estudo (CRESWELL; CLARK, 2015, p. 22).

Compreendemos que a escolha por esse método não está na validade ou confluência dos resultados, mas sim na possibilidade de olhar para o objeto de pesquisa e entendê-lo, para além de uma única perspectiva. Para isso, o desenvolvimento deste estudo, ou seja, o percurso metodológico, foi delineado em duas etapas. Na primeira etapa, realizamos um estudo quantitativo a partir da base de dados do Enem e a segunda etapa teve uma abordagem qualitativa, quando identificamos as sentenças descritoras (habilidades) associadas aos itens obtidos pela base e promovemos a resolução da atividade.

Cientes de que as Bases de Dados são fontes de registro e informação eletrônica, pesquisáveis, de modo interativo ou conversacional, através de um computador (MODESTO et al, 2006), compreendemos que devem servir para promover o acesso e fornecer à sociedade informações atualizadas, precisas e confiáveis, atendendo às necessidades do público-alvo e criando uma memória científica a partir de mecanismos eficientes de recuperação dessas informações.

Dessa forma, este artigo se apoiou na consulta e no tratamento de dados da Base de Dados do Enem<sup>1</sup>, disponibilizada pelo Instituto Nacional de Estudos e Pesquisas Anísio Teixeira (Inep). Essa Base oferece diversas informações sobre a escola e o estudante<sup>3</sup>, tais como o município, a inscrição da escola, se ele é concluinte do Ensino Médio ou se é um treineiro (estudante fora da etapa regular de ensino), a porcentagem de acerto, os parâmetros a, b e c bisseriais, o ponto de ancoragem, entre outros. A partir dessas informações, selecionamos as variáveis de interesse para identificar as habilidades de Química, avaliadas pelo Enem das edições de 2013 a 2017, que os estudantes concluintes do Ensino Médio apresentaram bom e baixo desempenho.

A seguir, detalharemos as etapas que vão subsidiar o entendimento do leitor sobre o acesso aos microdados e o tratamento realizado para a seleção de variáveis de interesse para esta pesquisa.

**Etapa 1:** Tratamento da base de dados do Enem do ano de 2017, realizando a seleção de variáveis e tabulando dados estatísticos.

Para iniciar a Etapa 1 deste estudo, foi realizado o download da base de dados do Enem de 2013 para familiarização com os microdados. Utilizamos o sistema de codificação *American Standard Code for Information Interchange* (ASCII), que transforma caracteres em números,

---

<sup>3</sup> Esses dados são disponibilizados ao público em geral pelo acesso ao portal do Inep, de forma gratuita pelo link: <https://www.gov.br/Inep/pt-br/acesso-a-informacao/dados-abertos/microdados/Enem>

facilitando a leitura dos dados. Para a análise estatística e manipulação dos dados, empregamos softwares amplamente reconhecidos, como o *Statistical Analysis System* (SAS) e o *Statistical Package for the Social Sciences* (SPSS). Estes programas são essenciais em diversas áreas, incluindo Ciências Sociais, negócios, pesquisa científica e saúde, devido à sua capacidade de processar e analisar dados complexos.

Um foco particular foi dado à variável que identifica a habilidade correspondente a cada item do teste. Essa variável é crucial para relacionar as habilidades especificadas com os microdados do exame. Além disso, realizamos uma análise detalhada dos itens do caderno azul, na seção de Ciências da Natureza, concentrando-nos em itens que abordavam estequiometria. Buscamos associar esses itens às descrições das habilidades na matriz de referência do Enem.

Para acessar os arquivos de dados, que estavam no formato de arquivo comprimido (zip), utilizamos o software descompactador 7-zip, conforme instruções do site do Inep. Os dados foram extraídos e analisados utilizando o IBM SPSS *Statistics* 22, uma versão específica do SPSS. Esse software permitiu a leitura de variáveis-chave, como número de inscritos, código do município de inscrição, unidade federativa, status de conclusão do Ensino Médio, código da escola, presença na prova, nota, vetor de resposta em Ciências da Natureza, e o código do caderno.

Utilizamos o *software* Programa R na versão 64 x 3.3.1, um ambiente de programação amplamente usado em estatística e análise de dados, para calcular a média e o desvio-padrão da proficiência dos estudantes ao longo do teste. Esses cálculos também forneceram a base para determinar os parâmetros ‘a’, ‘b’ e ‘c’ da TRI.

A Tabela 1, anexado ao documento, detalha de forma organizada os processos adotados para o tratamento da base de dados do Inep.

Tabela 1. Descrição dos processos para o tratamento da base de dados do Inep.

<b>Etapa</b>	<b>Ações</b>	<b>Descrição das ações do processo</b>
1	Leitura dos Dados	Utilização do <i>software</i> SPSS para leitura e análise inicial dos dados do Enem.
2	Filtragem de Concluintes	Seleção de dados específicos de estudantes que concluíram o Ensino Médio.
3	Análise de Frequência	Avaliação da frequência com que os estudantes participaram das edições do Enem.
4	Organização do Vetor de Resposta	Classificação do vetor de resposta, correspondente às respostas corretas.
5	Agrupamento de Itens	Criação de um sistema de identificação para reorganizar os itens nas diferentes versões coloridas do exame.
6	Análise de Acerto ou Erro	Avaliação do desempenho dos estudantes em cada item, utilizando códigos de leitura para acertos (1) e erros (0).
7	Cálculo do Percentual de Acerto	Determinação do percentual de acertos tanto para o item como para cada alternativa.
8	Análise por Descritor	Calculando o percentual de acertos associado a cada

		descritor do teste.
9	Média de Proficiência	Cálculo da média de proficiência dos estudantes baseado em seu desempenho no teste.
10	Média e Desvio Padrão dos Itens Acertados	Calculando a média e o desvio padrão para os itens acertados.
11	Cálculo de Parâmetros Estatísticos	Realização dos cálculos das bisseriais e dos parâmetros 'a', 'b' e 'c' da TRI.
12	Classificação dos Itens por Disciplina	Agrupamento dos itens de acordo com a disciplina correspondente, em diálogo com a etapa inicial de leitura dos dados.

Fonte: Elaborado pelo pesquisador.

Após concluir a sequência inicial de análise de dados, procedemos com a eliminação dos registros de estudantes que participaram como treineiros (sequência 2), focando nossa atenção nos concluintes do Ensino Médio. Esta seleção foi feita com base na variável numérica "0", conforme indicado pelo dicionário de variáveis, identificando os estudantes que concluíram o Ensino Médio (sequência 2). Posteriormente, estimamos a frequência de participação desses alunos nas edições do Enem de 2017 (sequência 3).

Em seguida, analisamos a proficiência desses estudantes com base em suas notas, calculando a média de proficiência em Ciências da Natureza tanto em nível nacional quanto regional. Além disso, utilizamos o vetor de resposta do estudante (sequência 4), presente no exame, para coletar dados sobre a frequência de escolha de cada alternativa em cada item.

Considerando que o Enem oferece cadernos de provas de diferentes cores (azul, amarelo, branco e rosa), com os itens distribuídos de maneira variada, optamos por utilizar o caderno azul como referência para todas as edições analisadas, seguindo o procedimento adotado por Carlos (2017). Essa decisão nos levou a reorganizar os itens, reagrupando as informações geradas pela Base de dados de cada caderno. Para referência, adotamos a nomenclatura "ITEM\_1" para o primeiro item do caderno azul de Ciências da Natureza e seguimos a ordem crescente para os demais itens. Portanto, foi possível realizar uma análise detalhada de acertos e erros para cada item do teste, abrangendo toda a amostra (conforme descrito na sequência 6 do procedimento). Essa análise incluiu a avaliação do percentual de acertos e a frequência de escolha de cada distrator (sequência 7). Com os dados devidamente preparados, avançamos para a etapa de cálculo da correlação ponto-bisserial entre as alternativas e as respostas corretas. Para isso, utilizamos o pacote "ltm" (*Likelihood Theory Models*) no *software* R. Esse pacote é uma biblioteca estatística especializada que oferece ferramentas para modelar itens em testes e avaliações educacionais, empregando modelos da TRI na função "*descript*".

Em seguida, procedemos ao cálculo dos parâmetros 'a', 'b' e 'c' da TRI, utilizando também o pacote "ltm", mas através da função "tpm". Esta função é específica para ajustar modelos de TRI

aos dados de testes como o Enem. Com base nesses parâmetros, determinamos o ponto de ancoragem de cada item, bem como a média de proficiência e o desvio-padrão dos estudantes. Estas medidas nos permitiram distribuir os alunos em diferentes níveis de desempenho, conforme as sequências 8, 9, 10 e 11 do nosso processo analítico. Estas etapas são fundamentais para uma compreensão aprofundada do desempenho dos estudantes e para a avaliação precisa da qualidade dos itens do teste.

Agrupamos todos esses dados em uma planilha, objetivando uma análise qualitativa mais acessível e sistemática. Para isso, estabelecemos categorias específicas que nos auxiliaram a identificar nos itens os conceitos pertinentes às áreas de Biologia, Física ou Química (conforme descrito na sequência 12 do processo). Essa categorização foi um passo essencial para transformar os dados quantitativos em informações pedagógicas aplicáveis, especialmente no campo de estudos de Química. Acreditamos que essa abordagem nos permitiu alcançar um nível de tratamento de dados em que a interpretação dos números estatísticos pode ser efetivamente convertida em compreensões pedagógicas valiosas, contribuindo para o aprimoramento do ensino e da aprendizagem em Química.

Na próxima seção, Etapa 2, descreveremos os critérios adotados para realizar as análises pedagógicas dos itens do Enem. Essa etapa é crucial, pois, ao combinar essas análises com os dados estatísticos já coletados, expandimos significativamente nosso entendimento sobre os conteúdos de Química abordados no exame. Esta abordagem integrada possibilita uma visão mais ampla e profunda dos tópicos químicos, contribuindo para uma compreensão mais completa dos desafios e oportunidades presentes na avaliação e no ensino de Química.

**Etapa 2:** Seleção do item que avalia conteúdo Químico, relacionado às habilidades previstas na matriz de referência do Enem.

Nesta Etapa 2, buscamos selecionar, por meio do ponto de ancoragem, um item da edição de 2017, com conteúdo químico. Após o reconhecimento desse item por meio da Base de dados, propomos a determinação da sentença descritora, que descreve a habilidade que mobiliza conhecimentos próprios da Química, a ser avaliada. Para isso, consideramos o domínio cognitivo, o conteúdo avaliado e o contexto envolvido na situação-problema. Nesse sentido, os resultados, a serem apresentados, são análises pedagógicas que engloba uma discussão sobre o seu domínio cognitivo, o conteúdo, o contexto, os mobilizadores utilizados como suporte e as informações estatísticas disponibilizadas pelo Inep à luz da TRI.

## RESULTADOS E DISCUSSÃO

### Leitura a partir das informações estatísticas de um item do Enem aplicado na edição de 2017

Para esclarecer a leitura desses dados, de modo prático, apresentamos um item aplicado no Enem, na avaliação de Ciências da Natureza. A Figura 1, a seguir, apresenta o item 42, do caderno azul, aplicado na edição de 2017.

Figura 1 - Item 42, Enem 2017.

Uma das estratégias para conservação de alimentos é o salgamento, adição de cloreto de sódio (NaCl), historicamente utilizado por tropeiros, vaqueiros e sertanejos para conservar carnes de boi, porco e peixe.

O que ocorre com as células presentes nos alimentos preservados com essa técnica?

- A O sal adicionado diminui a concentração de solutos em seu interior.
- B O sal adicionado desorganiza e destrói suas membranas plasmáticas.
- C A adição de sal altera as propriedades de suas membranas plasmáticas.
- D Os íons  $\text{Na}^+$  e  $\text{Cl}^-$  provenientes da dissociação do sal entram livremente nelas.
- E A grande concentração de sal no meio extracelular provoca a saída de água de dentro delas.

Fonte: Inep, 2017.

A contextualização aplicada nesse item é a técnica de conservação de alimentos por salgamento. Essa é uma técnica milenar que tem sido utilizada ao longo da história para preservar alimentos perecíveis, como carnes, peixes e vegetais. O processo envolve a adição de sal aos alimentos, o que reduz a atividade da água e cria um ambiente hostil para o crescimento de microorganismos responsáveis pela deterioração dos alimentos. Portanto, é uma contextualização do tipo pessoal, que envolve a atitude do aluno para conservar o alimento.

Nota-se também que o item explora a integração entre as áreas de ciência biológica e química, mobilizando do aluno o tipo de conhecimento de conteúdo dessas áreas no campo procedimental. Nesse sentido, o aluno que marca o gabarito, letra E, deverá ter em mente que a elevada concentração de sal no meio extracelular, também conhecida como hipertonicidade, provoca a saída de água das células por osmose. A osmose é o movimento de água através de membranas celulares semipermeáveis, do meio menos concentrado (hipotônico) para o meio mais

concentrado (hipertônico), a fim de igualar as concentrações de solutos (como o sal) nos dois lados da membrana.

Quando uma célula é exposta a um meio extracelular hipertônico, a concentração de solutos fora da célula é maior do que dentro dela. Como resultado, a água tende a se mover para fora da célula em uma tentativa de diluir a concentração de solutos no meio extracelular. Isso pode levar à perda de água da célula e causar a contração ou murchamento celular, um processo conhecido como plasmólise em células vegetais ou crenação em células animais.

Outro mecanismo importante, é transpor em uma sentença descritora a forma como o aluno deve mobilizar esse conteúdo/ procedimento, possível de observar com precisão, o que, de fato, está sendo avaliado no item 42. Para este item, utilizamos a Taxonomia de Bloom Revisada de Anderson e Krathwohl (2001). Essa taxonomia é frequentemente usada como um guia para o planejamento de atividades de ensino e avaliações, pois ajuda os professores e gestores a esclarecer os objetivos de aprendizado avaliados e a criar experiências de aprendizado que promovam o desenvolvimento de habilidades cognitivas em seus alunos.

Com base nesse documento e na taxonomia de Bloom et al. (1977), a hierarquia cognitiva do item está no campo do Avaliar. Segundo Bloom et al. (1977), avaliação estima Julgamentos a respeito do valor do material e dos métodos para certos propósitos. Para Ferraz e Belhot (2010, p. 429) avaliar está relacionado ao estudante realizar julgamentos baseados em critérios e padrões qualitativos e quantitativos ou de eficiência e eficácia. Como por exemplo, "*Avaliar a eficácia de diferentes abordagens no tratamento de uma determinada doença, com base em evidências científicas*".

No caso do item 42, analisado neste artigo, propomos que o estudante deverá "*Avaliar a eficácia da técnica de conservação de alimentos por salgamento, considerando o conceito de osmose*". Segundo os microdados do Inep, essa sentença está relacionada a habilidade H19 – Avaliar métodos, processos ou procedimentos das ciências naturais que contribuam para diagnosticar ou solucionar problemas de ordem social, econômica ou ambiental (BRASIL, 2009). Esta habilidade pretende compor o cenário que avalia a capacidade do aluno em demonstrar a competência 5 - Entender métodos e procedimentos próprios das ciências naturais e aplicá-los em diferentes contextos (BRASIL, 2009).

Após a aplicação do teste, a tabela 2 apresenta as informações estatísticas dos estudantes concluintes do Ensino Médio para esse item, obtidas pelo tratamento da base de dados do Inep, da forma como os dados são, geralmente, apresentados às escolas em relatórios pedagógicos.

Tabela 2. Exemplo de informações de um item do Enem (2017) em um relatório pedagógico

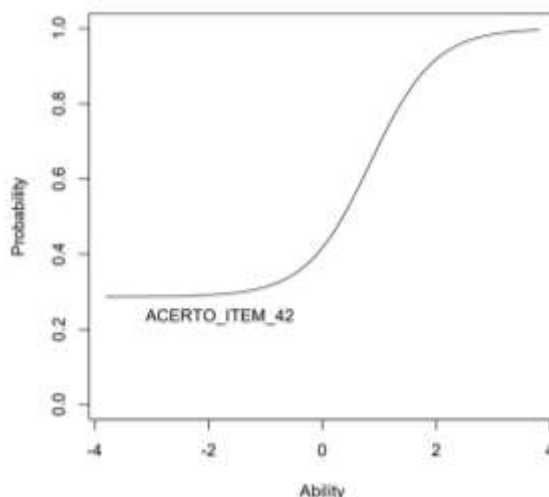
Item: 8757		Gabarito: E		Posição: 42			
Parâmetros da TRI		Análise Clássica		Percentual por opção de resposta		Bisserial das alternativas	
a	0,025	% de acerto	48,1%	A	22,4%	A	-0,17
b	566,7			B	6,9%	B	-0,12
c	0,28			C	13,2%	C	-0,12
Ancoragem	567,5			D	9,0%	D	-0,13
Proficiência média	506,3			E	48,1%	E	+0,34

Fonte: Elaborado pelos autores a partir da base de dados do Inep, 2017.

Com base na Tabela 2, segundo a TCT, o item obteve 48,1% de acerto. Pela TRI, além de conhecer o percentual de estudantes que acertaram, é possível obter mais informações a respeito do comportamento desse grupo de estudantes frente ao item avaliado. A partir da análise da bisserial das alternativas, somente o gabarito, alternativa E, apresentou valor positivo. Com isso, estimamos que os estudantes que acertaram o item, demonstraram ao longo do teste que possuíam proficiência adequada e, portanto, mobilizaram o conhecimento de conteúdo para solucionar o problema da questão. Para análise das bisseriais, caso outra alternativa tivesse apresentado valor positivo, poderia sinalizar que o item apresentaria algum problema técnico, em que alunos com bom desempenho no teste estariam sendo atraídos ou, problema pedagógico, quando se constata as deficiências no ensino do aluno com bom rendimento.

Pelos parâmetros da TRI, é possível construir a curva característica do item (CCI) (Figura 2).

Figura 2 - CCI do item 42, Enem 2017.



Fonte: Elaborado pelos autores.



Pela CCI, notamos pelo parâmetro  $c$  que aproximadamente 29% dos estudantes não atingiram a proficiência adequada para resolvê-lo e, possivelmente, tentaram acertar o item ao acaso. O valor da bisserial corrobora com esta análise, pois não se apresentou positiva para outra alternativa, a não ser o gabarito. Observamos também a inclinação da curva do item, pelo parâmetro “ $a$ ”, nos informa que o item cumpriu com a sua função de discriminar os estudantes que sabem dos que não sabem. Como vimos em Andrade et al. (2000), quanto maior a inclinação da curva, mais precisa será a distinção da população avaliada.

Com base no parâmetro  $b$ , que estima no gráfico a proficiência dos estudantes, a CCI indica dados sobre a dificuldade do item. A inclinação da curva começa a subir a ter um aumento no desvio padrão acima do 0. A escala acima de 0 indica que a proficiência exigida passa de 500 pontos, e, portanto, o item foi considerado acima da média para boa parte dos estudantes que realizaram o teste de Ciências da Natureza. Para melhor compreender, analisamos o dado da ancoragem.

A ancoragem revela que, na escala de 0 a 1000, o item 42 se posicionou na escala em 567,5. Isso significa que os alunos com proficiência superior a esta pontuação demonstraram ter condições de acertarem ao item. Considerando que a média da proficiência dos alunos concluintes do ensino médio no Enem de 2017 no teste de Ciências da Natureza foi de, aproximadamente, 506 pontos, nota-se que a maior parte dos alunos não apresentaram dificuldade em executar a atividade.

Para avançar ainda mais nessa análise, sugerimos a análise pedagógica, que avalia para além dos dados estatísticos, outros elementos, tais como os caminhos cognitivos por trás de cada alternativa, avaliando os possíveis interferentes para avançar na resolução dessa situação. Também um estudo sobre o tipo de suporte e o contexto aplicado ao item. Somados aos dados estatísticos e, replicado aos outros itens do teste, é possível gerenciar um mapa de itens que expressa a capacidade cognitiva dos estudantes concluintes do ensino médio respondentes dessa edição, a fim de compará-las com as gerações futuras.

## CONSIDERAÇÕES FINAIS

Ao longo da história da educação brasileira, as avaliações de Ciências ocorrerem de modo descontínuo, contribuindo pouco com reflexões sobre o ensino de Ciências da Natureza e ainda em menor impacto, no ensino da Química, oculto dentro das avaliações dessa grande área.

A TRI é uma ferramenta interessante para investigar o desenvolvimento dos conceitos de Ciências da Natureza, de modo analítico, a discriminação dos estudantes avaliados, a dificuldade enfrentada pelos estudantes em determinada habilidade e a quantidade de estudantes que precisaram

marcar uma alternativa diferente do gabarito por não estarem aptos para resolver uma situação–problema.

Nota-se que o item conseguiu cumprir com a finalidade em discriminar os alunos que conseguem e ou que não desenvolveram ainda a proficiência para executar a tarefa. Com os parâmetros também podemos observar que o conteúdo de osmose utilizado para analisar procedimentos de conservação de alimentos é um assunto bem relacionado, envolve integração entre os campos da ciência biológica e química, aplicados a saúde, que resgata conteúdos e procedimentos de ciências trabalhados desde os anos finais do ensino fundamental e, provavelmente, devido a esses fatores, pode ter contribuído para que, quase 50% dos respondentes tenham acertado o item.

A bisserial também é um dado relevante, pois sinaliza que ao responderem a atividade, estudantes com bom desempenho ao longo do teste, e que, possivelmente, teriam condição de acertar o item, são atraídos para outras alternativas, o que não foi o caso para o item em estudo. Esse apontamento pode dizer respeito a qualidade do item, ou às falhas conceituais que o estudante possui que o impede de avançar para outro nível de desempenho.

Consideramos que uma das falhas na formação continuada do professor está na ausência de momentos para se discutir aspectos da prática docente a partir dos instrumentos de avaliação, como possibilita as métricas obtidas pelos testes de larga escala tratados pela TRI, até mesmo para produzir relações com o currículo adotado na escola e as metodologias de ensino aplicadas em sala de aula.

Por fim, não pretendemos esgotar as possibilidades dessa ferramenta, mas atrair olhares para a possibilidade de traduzir números estatísticos em informações pedagógicas para o professor. A partir desse contexto, consideramos suficientes as informações apresentadas neste artigo para instrumentalizar os docentes, independente da área de atuação.

## REFERÊNCIAS BIBLIOGRÁFICAS

ALMEIDA, L. S. **Teorias da inteligência**. 1988.

ANDERSON, L. W.; KRATHWOHL, K. R. A. *Taxonomy for learning, teaching and assessing: a revision of Bloom's taxonomy of educational objectives*. New York: Longman, 2001.

ANDRADE, D. F. de; TAVARES, H. R; DA CUNHA VALLE, R. *Teoria da Resposta ao Item: conceitos e aplicações*. ABE, São Paulo, 2000.

BINET, A.; SIMON, T. *Méthodes nouvelles pour le diagnostic du niveau intellectuel des anormaux*. *L'Année Psychologique*, Vol. 11, p. 191–244; 1904.

BINET, A.; SIMON, T. Méthodes nouvelles pour diagnostiquer l'idiotie, l'imbécillité et la débilité mentale. In: Sante de Sanctis (Ed.), **Atti del V congresso internazionale di psicologia tenuto**. p. 507-510. Roma: Forzani. 1905.

BLOOM, B. S. Taxionomia de objetivos Educacionais. 6. ed. Rio Grande do Sul: **Editora Globo**, 1977.

BRITO, M. R.; MUNHOZ, A.; PRIMI, R.; GONÇALVEZ, M. H. et al. Exames Nacionais: uma análise do ENEM aplicado à Matemática. **Avaliação**, Campinas, v. 5, n. 4, p. 445-454, 2000.

GARCIA, P. S.; BIZZO, N.; ROSA, S. S. da. Políticas de avaliação e o desenvolvimento profissional do professor de ciências. **Educação UFSM**, v. 44, 2019.

CARLOS, P. R. O. **Avaliações Externas De Grande Escala**. Clube de Autores, 2017.

CARVALHO, V. D. S; AFONSO, A. F. O uso da psicometria para avaliar procedimentos matemáticos em um item da prova de ciências da natureza e suas tecnologias do Enem 2015: implicações para o ensino de química. **Cuadernos de Educación y Desarrollo**, v. 15, n. 11, p. 14024-14043, 2023.

GATTI, B. A. Avaliação de sistemas educacionais no Brasil. Sísifo: **Revista de Ciências da Educação**, n. 9, p. 7-18, 2009.

LUCKESI, C. C. **Avaliação em educação: questões epistemológicas e práticas**. Salvador: Cortez, 2018.

MEDEIROS, L.; JALOTO, A; DOS SANTOS, A. V. F. A área de ciências nas avaliações internacionais de larga escala. **Estudos em Avaliação Educacional**, v. 28, n. 68, p. 512-537, 2017.

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

PASQUALI, L. **TRI–Teoria de resposta ao item: Teoria, procedimentos e aplicações**. Editora Appris, 2020.

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

PRIMI, R.; SANTOS, A. A. A.; VENDRAMINI, C. M. M. et al. Competências e habilidades cognitivas: diferentes definições do mesmo construto. **Psicologia: Teoria e Pesquisa**, Brasília, v. 17, n. 2, p. 151-159, 2001.

SARTES, L. M A; SOUZA-FORMIGONI, M. L. O. de. Avanços na psicometria: da teoria clássica dos testes à teoria de resposta ao item. **Psicologia: Reflexão e Crítica**, v. 26, p. 241-250, 2013.

SOARES, T. M. Utilização da teoria da resposta ao item na produção de indicadores sócio-econômicos. **Pesquisa operacional**, v. 25, p. 83-112, 2005.

SPAZZIANI, G. P. **Devolutivas pedagógicas construídas a partir das escolhas das alternativas pelos candidatos nos itens de física do ENEM**. 2019. Tese de Doutorado. [sn].

TEIXEIRA, R. A. G. Educação do anormal a partir dos testes de inteligência. **História da Educação**, v. 23, p. e90024, 2019.

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