

Análise dos Parâmetros de Itens de Matemática à luz da Teoria Clássica dos Testes (TCT) e da Teoria de Resposta ao Item (TRI)

Analysis of the Parameters of Mathematical Items in the view of the Classical Test Theory (CTT) and the Item Response Theory (IRT)

Wendel Melo Andrade¹

Jorge Carvalho Brandão²

Maria José Costa dos Santos³

RESUMO

Objetiva-se com este artigo analisar os parâmetros de itens de matemática praticados nas avaliações externas, fundamentando-se nos pressupostos da TCT e da TRI. O procedimento metodológico utilizado contou com elementos de pesquisa bibliográfica e de pesquisa documental, por meio da realização de uma análise de itens do 5º ano do ensino fundamental, publicados em boletim pedagógico. Dentre os parâmetros estudados, destaca-se a capacidade de discriminação do item, a habilidade necessária para respondê-lo e a probabilidade de resposta correta dada por respondentes de baixa habilidade. Nas análises, foram considerados os preceitos da unidimensionalidade e interdependência local, a habilidade do respondente e os índices de discriminação, de dificuldade e de coeficiente bisserial. Percebeu-se que os procedimentos oriundos da TCT e da TRI possibilitam a identificação da proficiência do estudante, pois é possível estabelecer uma relação entre a probabilidade de um aluno responder corretamente a um item e suas habilidades na área do conhecimento avaliado.

PALAVRAS-CHAVE: Itens de Matemática. Teoria Clássica dos Testes. Teoria de Resposta ao Item. Análise de Parâmetros.

¹ Doutorando em Educação pela Universidade Federal do Ceará (UFC). E-mail: professorwendelmelo@gmail.com. ORCID: <https://orcid.org/0000-0001-8921-7326>.

² Doutor em Educação pela Universidade Federal do Ceará (UFC). E-mail: profbrandao@ufc.br. ORCID: <https://orcid.org/0000-0003-4818-135X>.

³ Pós-Doutora em Educação pela Universidade Estadual do Rio de Janeiro (UERJ). E-mail: mazzesantos@ufc.br. ORCID: <https://orcid.org/0000-0001-9983-1653>.



ABSTRACT

This article aims to analyze the parameters of mathematics items practiced in external evaluations, based on the assumptions of CTT and IRT. The methodological procedure used included bibliographic research and documentary research elements, as an analysis of items from the 5th year of elementary school published in a pedagogical bulletin was carried out. Among the parameters studied, the item's ability to discriminate stand out, the ability to answer it and the probability of correct answer given by respondents with low ability. In the analyzes, we considered unidimensionality and local interdependence precepts, respondent's ability and discrimination, difficulty and biserial coefficient indices. We noticed that the procedures originating from CTT and IRT enable the student's proficiency identification, since it is possible to establish a relationship between the probability of a student correctly answering an item and his/her abilities in the evaluated knowledge area.

KEYWORDS: Mathematics Items. Classical Test Theory. Item Response Theory. Parameter Analysis..

Introdução

No Brasil, o Sistema de Avaliação da Educação Básica (SAEB) constitui um conjunto de avaliações externas em larga escala. Ele vem sendo realizado desde o início dos anos 1990 e acontece nos anos ímpares. Atualmente, avalia competências na área de Língua Portuguesa e Matemática e é aplicado no 2º, 5º e 9º anos do Ensino Fundamental (EF) e no 3º ano do Ensino Médio (EM). Os dados do SAEB possibilitam ao Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira (INEP) realizar um diagnóstico da educação básica brasileira e de fatores que possam interferir no desempenho dos estudantes (BRASIL, 2019).

Em sintonia com o SAEB, o Estado do Ceará criou, em 1992, o Sistema Permanente de Avaliação da Educação Básica do Ceará (SPAECE). Esse sistema de avaliação externa em larga escala tem por objetivo fornecer subsídios para a formulação, reformulação e monitoramento das políticas educacionais com fins de melhoria da qualidade da educação do Estado. Também tem sido utilizado para diagnosticar os resultados em nível de estado, municípios e escolas. Atualmente o SPAECE ocorre anualmente e, assim como o SAEB, também avalia estudantes do 2º, 5º e 9º anos do EF e do 3º ano do EM, contemplando habilidades em Língua Portuguesa e Matemática (CEARÁ, 2007).

Tanto o SAEB quanto o SPAECE oferecem medidas de proficiência acerca do progresso do sistema de ensino como um todo e, em particular, de cada escola, atendendo a dois propósitos principais: o de prestar contas à sociedade sobre a eficácia dos serviços educacionais oferecidos à população e o de fornecer subsídios para o planejamento das escolas em suas atividades de gestão e de ações pedagógicas (CEARÁ, 2012).

Com base nas análises dos seus resultados, esses sistemas de avaliação utilizam-se dos preceitos da Teoria Clássica dos Testes (TCT) e da Teoria de

Resposta ao Item (TRI) para identificar o desempenho de cada aluno, escola, município, estado e até mesmo do país, posicionando-os numa escala de proficiência que varia de 0 a 500.

Em 2017, por exemplo, o resultado do SAEB em matemática, no 5º ano do EF, no Brasil, foi de 224,1. No entanto, para melhor compreender esse resultado, é preciso entender que, no SAEB, a escala de proficiência de matemática está subdividida em 10 níveis, sendo que a proficiência de 224,1 está situada no limite superior do Nível 4, indicando que um estudante, nesse nível, tratando-se do tema espaço e forma, provavelmente é capaz de reconhecer retângulos em meio a outros quadriláteros e de reconhecer a planificação de uma pirâmide dentre um conjunto de planificações, além de outras habilidades dos demais temas, que são: números e operações, grandezas e medidas e tratamento da informação (BRASIL, 2019).

No Quadro 1, pode-se observar a distribuição dos níveis da escala de proficiência em matemática para o 5º e 9º anos do EF e para o 3º do EM, no SAEB.

Quadro 1 – Distribuição de níveis de proficiência de matemática no SAEB

NÍVEIS	5º ano EF	9º ano EF	3º ano EM
Nível 1	Até 149	Até 224	Até 249
Nível 2	150 – 174	225 – 249	250 – 274
Nível 3	175 – 199	250 – 274	275 – 299
Nível 4	200 – 224	275 – 299	300 – 324
Nível 5	225 – 249	300 – 324	325 – 349
Nível 6	250 – 274	325 – 349	350 – 374
Nível 7	275 – 299	350 – 374	375 – 399
Nível 8	300 – 324	375 – 399	400 – 424
Nível 9	325 – 349	400 – 424	425 – 449
Nível 10	Acima de 350	Acima de 425	Acima de 450

Fonte: adaptado do Relatório SAEB (BRASIL, 2019)

Já em 2018, por exemplo, o resultado do SPAECE em matemática, no 5º ano do EF da rede municipal do Ceará foi de 237,0, sendo que, assim como no SAEB, para melhor compreensão desse resultado, precisa-se entender que, no SPAECE, a escala de proficiência está subdividida em padrões de desempenho, sendo eles: muito crítico, crítico, intermediário e adequado. Com esse resultado de 237,0, o 5º ano do EF da rede municipal do Ceará, em matemática, em 2018, situou-se no padrão de desempenho intermediário. Nesse padrão de desempenho, os estudantes revelaram ter consolidado as habilidades consideradas mínimas e essenciais para sua etapa de escolaridade (CEARÁ, 2019).

No Quadro 2, pode-se observar a distribuição dos padrões de desempenho na escala de proficiência em matemática para o 5º e 9º anos do EF e para o 3º ano do EM, no SPAECE.

Quadro 2 – Padrões de desempenho da escala de proficiência de matemática do SPAECE

	MUITO CRÍTICO	CRÍTICO	INTERMEDIÁRIO	ADEQUADO
5º ano EF	Até 149	150 a 199	200 a 249	Acima de 250
9º ano EF	Até 224	225 a 274	275 a 324	Acima de 325
3º ano EM	Até 249	250 a 299	300 a 349	Acima de 350

Fonte: adaptado de Boletim SEDUC (CEARÁ, 2019)

Para a elaboração de seus testes padronizados, esses sistemas se utilizam de Matrizes de Referências, que são “recortes” de propostas curriculares de ensino, elaboradas por especialistas em avaliação e nas áreas específicas do conhecimento. Estes testes padronizados utilizam, predominantemente, questões objetivas e de múltipla escolha, o que denota uma de suas limitações. Além disso, o fato de não contemplarem toda a proposta curricular acentua as restrições dessas avaliações. Em análises realizadas na Provinha Brasil, que também adota uma Matriz de Referência para sua elaboração, Silva, Bellemain e Borba (2016, p. 740) constataram que “o recorte realizado pela Matriz de Referência e as limitações de um teste padronizado restringem as possibilidades da avaliação”. No entanto, nessas Matrizes de Referências, encontram-se os descritores, que são as habilidades avaliadas pelos instrumentos das avaliações externas.

Quadro 3 – Síntese das Matrizes de Referências de matemática do SAEB e SPAECE

MATRIZ	TEMAS	DESCRITORES
SAEB (5º ano EF)	Espaço e forma	D1 até D5
	Grandezas e Medidas	D6 até D12
	Números e operações/Álgebra e funções	D13 até D26
	Tratamento da informação	D27 e D28
SPAECE (Matriz única para todos os anos)	Interagindo com números e funções	D1 até D45
	Convivendo com a geometria	D46 até D63
	Vivenciando as medidas	D64 até D72
	Tratamento da informação	D73 até D78

Fonte: adaptado pelos autores com base nas Matrizes do SAEB e SPAECE

No Quadro 3, pode-se observar uma síntese dos temas que contemplam a Matriz de Referência de matemática do SAEB e do SPAECE para o 5º ano do EF, com o indicativo de seus descritores.

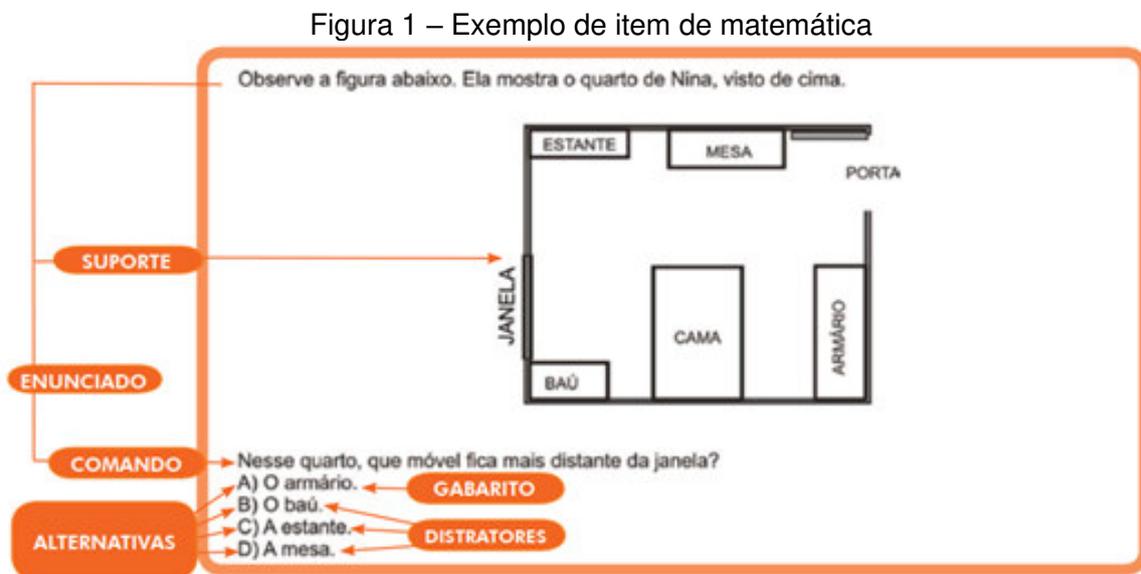
Santos e Ortigão (2016, p. 65) destacam que:

Uma Matriz de Referência é composta por um conjunto de descritores que explicitam dois pontos básicos do que se pretende avaliar: o conteúdo programático a ser avaliado em cada período de escolarização e o nível de operação mental necessário para a realização de determinadas tarefas.

As habilidades indicadas pelos descritores de uma Matriz de Referência servem de base para a elaboração dos itens, que são as questões que compõem uma avaliação em larga escala. Em geral eles são constituídos por enunciado, suporte, comando e alternativas de respostas, que podem ser as erradas, chamadas

de distratores, ou a correta, denominada de gabarito. Os itens são elaborados com o propósito de avaliar apenas uma habilidade, buscando abordar uma única dimensão do conhecimento.

Na Figura 1, pode-se observar um exemplo de item de matemática com seus principais elementos.



Fonte: Guia de elaboração de itens de matemática (CAED, 2008, p. 19)

Atualmente, os sistemas de avaliações externas e em larga escala têm se configurado a principal referencial para o trabalho com os alunos, direcionando o exercício das práticas docentes, muitas vezes, exclusivamente, para o treino dos estudantes para essas avaliações.

Santos e Ortigão advertem sobre isso, ao destacarem que:

[...] a partir das avaliações externas, algumas escolas reforçam o sentido de ensinar e aprender para o teste, e isso tem direcionado as atividades de ensino para os conteúdos que serão avaliados, e são desconsiderados os demais conteúdos do currículo, pois não serão objetos de avaliação, e com isso, o ensino passa a ser, na escola, o treino para as avaliações externas, e isso pode provocar o estreitamento curricular [...]. Em alguns casos, a Matriz de Referência desses testes passa a ser considerada, na escola, como currículo. (SANTOS; ORTIGÃO, 2016, p. 63-64).

Esse fenômeno provoca reflexões sobre as implicações e impactos acarretados por essas avaliações nos diferentes segmentos do sistema educacional, levantando assim a necessidade de conhecer mais sobre os procedimentos que compõem as avaliações e, principalmente, sobre a TCT e a TRI, a começar pelo seu elemento base, que é o item.

Nesse contexto, objetiva-se analisar os parâmetros de itens de matemática praticados nas avaliações externas. De modo análogo, busca-se ampliar os conhecimentos sobre os conceitos e procedimentos relacionados à TCT e à TRI. Para tanto, este estudo apoia-se nos trabalhos de Andrade, Tavares e Valle (2000), Pasquali (2018), Pasquali e Primi (2003), Couto e Primi (2011), Klein (2013), Moreira Junior (2010), Cunha (2014), Mendonça (2012), Tavares (2013), Coelho (2014) e Silva (2015).

Este artigo está organizado em cinco seções, sendo a primeira esta introdução. Na segunda seção, apresenta-se a metodologia desta pesquisa. Na terceira, são levantadas reflexões sobre a TCT e a TRI, buscando entender seus preceitos e principalmente os parâmetros do item. Na quarta, são analisados itens de matemática a partir desses parâmetros, e por fim, na quinta, tecem-se as considerações finais.

Metodologia

Do ponto de vista da sua natureza, esta pesquisa se configura como sendo do tipo aplicada, pois tem o propósito de levantar conhecimentos acerca da análise de itens de matemática, qualificando e ampliando os conhecimentos sobre este assunto (PRODANOV; FREITAS, 2013).

Em decorrência do objetivo a ser alcançado, procedeu-se a uma pesquisa exploratória, com a finalidade de proporcionar mais informações sobre o assunto em questão, possibilitando sua melhor compreensão. De acordo com Gil (2011, p. 57), “as pesquisas exploratórias são desenvolvidas com o objetivo de proporcionar uma visão geral acerca de determinado fato”.

Adotaram-se, inicialmente, procedimentos de uma pesquisa bibliográfica, tendo como foco trabalhos já publicados sobre a temática em estudo, buscando-se aprofundamento nos preceitos da TCT e da TRI. Para tanto, realizaram-se pesquisas em livros, artigos, dissertações e teses, pois, como leciona Severino (2007, p. 122), nesse tipo de procedimento metodológico, o pesquisador:

[...] utiliza-se de dados ou de categorias teóricas já trabalhados por outros pesquisadores e devidamente registrados. Os textos tornam-se fontes dos temas a serem pesquisados. O pesquisador trabalha a partir das contribuições dos autores dos estudos analíticos constantes nos textos.

Posteriormente, adotaram-se procedimentos da pesquisa documental, e realizou-se uma análise de itens de matemática do SPAECE, utilizados na avaliação do 5º ano do EF e publicados em boletim pedagógico (CEARÁ, 2005). Foram

selecionados, desse documento, dois itens para a análise dos parâmetros observados na TCT e na TRI. Como critério de escolha dos itens, considerou-se relevante, para os resultados objetivados por este estudo, tomar um item de dificuldade baixa e outro com maior dificuldade.

Compreendendo a TCT e a TRI

Para uma melhor compreensão acerca da TRI, é preciso entender inicialmente sobre a TCT, que é um dos modelos mais comuns adotados em processos avaliativos. Esse procedimento teórico baseia-se em resultados obtidos em provas a partir de escores brutos ou padronizados. Logo, nesse tipo de análise de resultado, cada item ou questão do instrumento de avaliação tem um único escore, sendo que o respondente adquire esse escore ao acertar o item e não adquire caso erre sua resposta. Com base nos escritos de Pasquali e Primi (2003), a TCT está relacionada à Psicometria Clássica, e nesse contexto, percebe-se uma limitação desse modelo de análise, pois independentemente do nível de dificuldade do item, sendo fácil ou difícil, ao acertá-lo, o respondente terá obtido o mesmo escore.

Em muitos casos, as avaliações de aprendizagens praticadas pelos professores de matemática no ambiente escolar, apresentam características semelhantes à TCT, pois, em geral, cada questão da avaliação tem o mesmo escore, independentemente de sua dificuldade. Por exemplo, em uma prova de matemática com dez questões, na qual cada questão vale um ponto, um estudante que acerte apenas uma questão de nível fácil terá a mesma nota de um estudante que acerte também apenas uma questão, porém de maior dificuldade.

Tradicionalmente, a TCT vem sendo muito utilizada em testes de seleção de indivíduos, tais como concursos públicos ou vestibulares.

Tavares (2013, p. 65) destaca que a análise dos resultados de uma avaliação em larga escala pode ser realizada de duas formas: com a TCT, avaliando os sujeitos a partir dos escores dos itens respondidos; e com a TRI, que, segundo o autor, consiste em um:

[...] conjunto de modelos matemáticos em que a probabilidade de resposta a um item é modelada como função da proficiência do aluno e de parâmetros que expressam certas propriedades dos itens, pois, quanto maior a proficiência do aluno, maior a probabilidade de acertar o item.

Ainda buscando estabelecer um paralelo entre a TCT e a TRI, Tavares (2013, p. 63) elucida que, na TCT, “a comparação entre indivíduos somente é possível

quando eles são submetidos às mesmas provas”. Isso evidencia uma limitação da TCT, mostrando sua dificuldade em comparar resultados obtidos por indivíduos que não são submetidos ao mesmo teste. A TRI surge então para sanar essa lacuna, pois, como ressaltam Andrade, Tavares e Valle (2000, p. 3):

[...] ela permite a comparação entre populações, desde que submetidas a provas que tenham alguns itens comuns ou, ainda, a comparação entre indivíduos da mesma população que tenham sido submetidos a provas totalmente diferentes.

É por causa dessa capacidade de comparação que os sistemas educacionais conseguem estabelecer parâmetros de evolução dos rendimentos de um ano para o outro ou de uma série para outra, ou até mesmo entre escolas públicas e particulares. No entanto, é preciso entender que essa possibilidade de comparação se dá apenas no âmbito daquilo que se propõe à avaliação externa, ou seja, das habilidades apontadas nas Matrizes de Referências dos sistemas de avaliação.

De acordo com Moreira Júnior (2010, p. 138),

a TRI surge no mundo na década de 1950 com os trabalhos de Lord (1952), que desenvolveu os primeiros modelos. Atualmente, existem centenas de aplicações e milhares de trabalhos publicados sobre a TRI no mundo.

Nas últimas décadas, as aplicações da TRI na área da Avaliação Educacional têm se ampliado, estando cada vez mais presentes nos modelos de análises dos resultados de avaliações externas em larga escala.

No Brasil, a TRI foi utilizada pela primeira vez em 1995, na análise dos dados do SAEB. Devido a ela, os desempenhos dos alunos de 4^a e 8^a séries do EF e de 3^a série do EM foram colocados em uma única escala de conhecimento, podendo ser comparados entre si. A partir de então, a TRI passou a ser utilizada também em outros sistemas de avaliação, como o Exame Nacional do Ensino Médio (ENEM), em 2009 (SILVA, 2015).

De acordo com os escritos de Pasquali (2018), Couto e Primi (2011) e Andrade, Tavares e Valle (2000), os vários modelos de análise propostos pela TRI dependem fundamentalmente de três fatores: (I) a natureza do item; (II) o número de populações envolvidas; e, (III) a quantidade de traços latentes que está sendo medida.

Sobre a natureza do item (I), têm-se os modelos de itens dicotômicos e não dicotômicos. Os itens dicotômicos se caracterizam por serem de múltipla escolha, em que se apresenta apenas uma possibilidade de acerto. Eles são amplamente utilizados nas avaliações de larga escala. Quanto aos itens não dicotômicos,

caracterizam-se por serem abertos, de resposta livre, possibilitando mais de uma categoria de acerto (COUTO; PRIMI, 2011).

Sobre o número de populações envolvidas (II), pode-se ter uma ou mais de uma. No entanto, é importante entender que, na área da Avaliação Educacional, é comum que uma população seja definida por características que podem variar dependendo do objetivo do estudo. Assim, pode-se tomar como populações distintas, por exemplo, os alunos do 5º ano e do 9º ano do EF. Como também é possível considerar populações distintas, alunos de uma mesma série, mas de períodos de aplicação diferentes, por exemplo, alunos do 5º ano do EF do ano de 2016 e do ano de 2017. Entendido o conceito de população, é válido compreender o significado de grupo, como sendo uma amostra da população. Por exemplo, da população de alunos do 3º ano do EM, pode-se selecionar um grupo de alunos do turno da manhã e outro grupo de alunos do turno da tarde (ANDRADE; TAVARES; VALLE, 2000).

Sobre a quantidade de traços latentes que está sendo medida (III), pode-se ter um ou mais de um. De acordo com Pasquali (2018), o traço latente pode ser entendido como sendo o grau de satisfação ou maturidade de um indivíduo perante uma habilidade ou proficiência. Os modelos de avaliação em larga escala, em geral, utilizam-se da unidimensionalidade. Isso quer dizer que são compostos por itens que, supostamente, devem estar medindo um único traço latente. Em outras palavras, deve haver apenas uma habilidade responsável pela realização de um item do teste, ou pelo menos uma habilidade considerada dominante (TAVARES, 2013).

Mendonça (2012) nos ajuda a compreender o sentido de habilidade adotado pela TRI, relacionando-o com o conceito de variável latente. O autor afirma que:

A habilidade é uma variável latente, ou seja, uma variável que não pode ser medida diretamente, diferentemente de variáveis como: peso, altura, temperatura etc. Por isso, variáveis como: ansiedade, satisfação, inteligência, conhecimento, que não são diretamente aferidas, são classificadas como latentes. Esse tipo de variável é aferido a partir de variáveis secundárias observáveis relacionadas a ela. No caso da competência, a variável secundária observada é a resposta dada pelo examinado a um item. (MENDONÇA, 2012, p. 20).

É importante entender que a habilidade necessária para a realização de um item se dá de modo independente para com os outros itens. Sendo assim, cada habilidade deve se apresentar não dependendo de outra habilidade, ratificando o que Andrade, Tavares e Valle (2000) definem por independência local. Portanto,

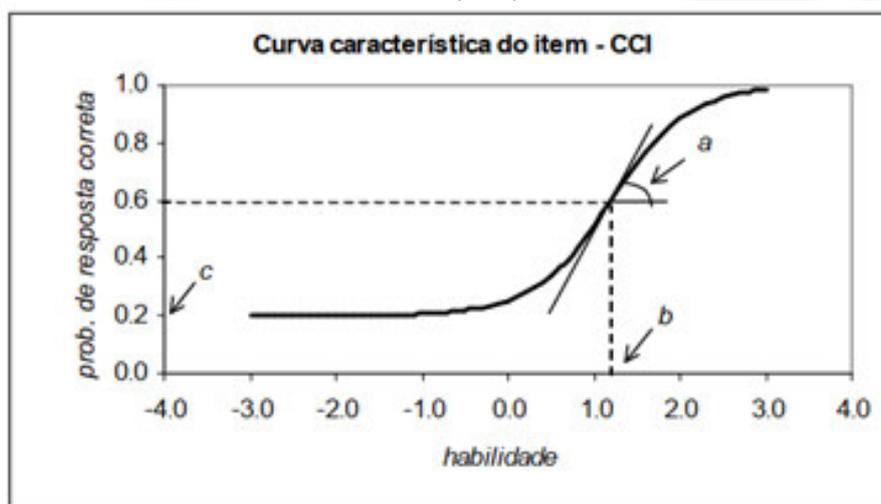
pelo pressuposto da independência local, um item não deve estabelecer relação de dependência com nenhum outro item do teste, mesmo entre habilidades.

Para que se tenha um item que atenda os preceitos da unidimensionalidade e da independência local, é preciso que haja um rigoroso critério na sua elaboração, e, para isso, ele deve respeitar os quesitos que serão postos em análise, como estar devidamente relacionado a um descritor da Matriz de Referência (COELHO, 2014).

Pasquali (2018) postula que o item, uma vez elaborado, é revisado nos pressupostos da TRI, sendo analisado considerando três modelos de parâmetros, os quais são: (a) a discriminação; (b) a habilidade; e, (c) a probabilidade de resposta correta dada por respondentes de baixa habilidade.

A análise desses parâmetros ocorre por meio de modelos matemáticos e estatísticos que geram para cada item a chamada Curva Característica do Item (CCI). Ela se baseia no fato de que indivíduos com maior habilidade possuem maior probabilidade de acertar o item e que essa relação é não linear (TAVARES, 2013).

Figura 2 – Curva Característica do Item (CCI) de um item com cinco alternativas



Fonte: Andrade, Tavares e Valle (2000, p. 11)

Na Figura 2, observa-se um exemplo de CCI para um determinado item, com cinco alternativas.

Encontram-se, na CCI, os três parâmetros analisados, pela TRI, no item, que se detalha a seguir, com base nos trabalhos de Pasquali (2018), Cunha (2014), Andrade, Tavares e Valle (2000) e Coelho (2014).

O parâmetro (a) se refere à capacidade de discriminação do item. Ele é proporcional ao ângulo formado pela reta tangente ao gráfico no seu ponto de inflexão com a reta que passa por esse ponto e é paralela ao eixo da habilidade. Assim, baixos valores de (a), que fazem o gráfico se apresentar menos inclinado,

indicam que o item tem pouco poder de discriminação, ou seja, alunos com habilidades diferentes têm, aproximadamente, a mesma probabilidade de responder corretamente ao item; e valores altos de (a), que fazem o gráfico se apresentar mais inclinado, indicam que o item tem alto poder de discriminação, sendo capaz de distinguir melhor os alunos que possuem menor e maior habilidade.

O parâmetro (b) se refere à habilidade necessária para responder ao item. Ele corresponde à projeção ortogonal do ponto de inflexão do gráfico no eixo da habilidade. Nele, quanto maior for o valor de (b), maior será a habilidade requerida para acertar o item. Seu valor indica a habilidade necessária para que se tenha uma probabilidade de acerto igual a 0,6, ou 60%. De acordo com Cunha (2014, p. 42), esse parâmetro está diretamente relacionado à proficiência do indivíduo, o autor assevera que:

Quanto maior a dificuldade de uma questão, maior deve ser a habilidade de um aluno para respondê-la corretamente. Como a dificuldade do item é fixa, o que varia é o nível de proficiência, que determinará a probabilidade de acerto da questão.

O parâmetro (c) representa a probabilidade de um aluno com baixa habilidade responder corretamente ao item. Ele está associado com a probabilidade de acerto ao acaso. Logo, quanto maior for o valor de (c), maior será a probabilidade de um estudante acertar o item pelo “chute”. De um modo geral, um item com cinco alternativas deverá ter valor de (c) próximo de 0,2, o que corresponde a um percentual de 20% de acerto, caso um aluno venha a responder ao item no “chute”, ou seja, 100% dividido por cinco alternativas. Item, com cinco alternativas, que apresenta valor de (c) acima de 0,2 indica que o gabarito se apresentou mais atrativo para os estudantes com baixa proficiência, e, nos casos em que o (c) apresenta valor abaixo de 0,2, tem-se um item em que o gabarito se apresentou menos atrativo para os estudantes de baixa proficiência. Logo, valores de (c) mais baixos sugerem itens mais bem elaborados, pois a chance de um aluno de baixa proficiência acertá-los no “chute” é menor.

Compreendidos os significados dos valores dos parâmetros de um item, (discriminação, habilidade e probabilidade de acerto ao acaso), surge então a seguinte questão: como identificar estes valores para cada item?

A resposta para esta questão está no processo de estimação, que é uma das etapas mais importantes da TRI, pois é por meio da estimação que se identificam os parâmetros do item e com eles o nível de proficiência necessária para respondê-lo corretamente.

Na prática, a estimação ocorre com a pré-testagem dos itens elaborados, a partir da realização de uma avaliação contendo itens novos em uma pequena população. De acordo com Klein (2013), a precisão no processo de estimação dependerá da quantidade de itens aplicados e da população, ou seja, quanto mais itens tiver o teste e quanto maior for a população do pré-teste, maior será a precisão na identificação dos seus parâmetros.

Para o processo de estimação, a TRI utiliza modelos matemáticos e estatísticos envolvendo cálculos probabilísticos de acerto e de erro do item pela população que realiza o pré-teste. Os modelos geralmente utilizados são denominados de Estimadores de Máxima Verossimilhança (EMV), pois, com eles, é possível estimar os parâmetros dos itens e as habilidades dos indivíduos simultaneamente. Os EMV consistem em funções não lineares que relacionam o número de acertos e os parâmetros de dificuldade do item, sendo que, para cada modelo logístico, de 1, 2 ou 3 parâmetros, usam-se equações específicas que possibilitam maximizar o valor da função de verossimilhança, ou seja, a confirmação dos valores dos parâmetros do item e da habilidade de seus respondentes (KLEIN, 2013).

Compõem este processo de estimação do item procedimentos estatísticos também relacionados à TCT, que determinam índices que variam de 0 a 1, dentre os quais se destacam: (I) O Índice de Dificuldade do Item (Difi), obtido pelo percentual de acerto do gabarito do item, sendo considerado item fácil aquele que possui índice próximo ou superior a 0,65, e difícil aquele próximo ou inferior a 0,30; (II) O Índice de Discriminação do Item (Discr), obtido a partir da diferença entre os percentuais de acertos dos 27% de alunos de melhor desempenho e dos 27% dos alunos de pior desempenho. Um Índice de Discriminação muito baixo, próximo ou menor que 0,25, indica que o item não discrimina adequadamente os alunos de melhor e pior desempenho. Para que o item discrimine bem os alunos que sabem dos que não sabem, seu índice deve ser superior e o mais distante possível de 0,25; (III) O Índice de Coeficiente Bisserial (Bisse), obtido pela correlação bisserial entre o acerto no item e o número de acertos no teste. Para um item ser considerado bom, seu valor deve ser positivo e maior que 0,30, sendo o mesmo valor do coeficiente bisserial do gabarito do item, enquanto que, nos distratores, os valores devem ser negativos e distantes de zero, indicando que as alternativas erradas não se apresentam atrativas aos estudantes de melhor desempenho (CEARÁ, 2006).

Quando o item passa pelo processo de estimação, afirma-se que ele está calibrado e já pode ser aplicado em uma avaliação de larga escala, pois, com o seu auxílio, e utilizando-se dos preceitos da TRI, é possível identificar a proficiência dos respondentes (KLEIN, 2013).

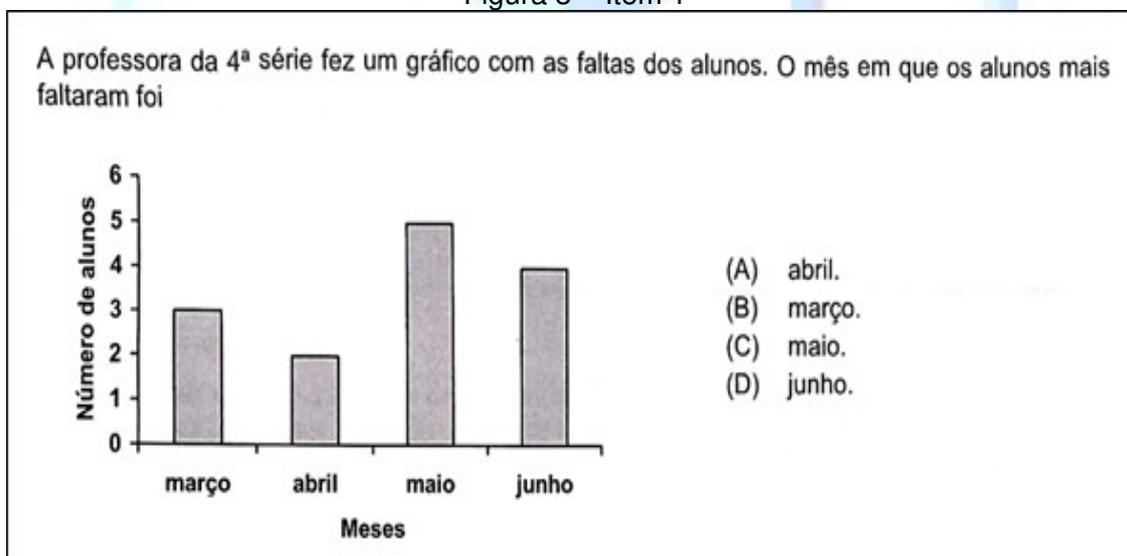
Na seção seguinte, analisam-se alguns itens de matemática com base nos parâmetros estudados.

Análise dos parâmetros de itens de matemática

Nesta análise, observou-se o item em sua estrutura e formulação, considerando-se os preceitos da unidimensionalidade e interdependência local. Também se buscou interpretá-lo com base nos parâmetros relacionados ao Índice de Dificuldade do Item (Difi), ao Índice de Discriminação do Item (Discr), ao Índice de Coeficiente Bisserial (Bisse) e aos parâmetros (a), (b) e (c) apontados na CCI, indicando a proficiência necessária para responder corretamente ao item, relacionando-a com algumas habilidades da Matriz de Referência associadas a essa proficiência.

Procede-se, inicialmente, à análise do item indicado na Figura 3.

Figura 3 – Item 1



Fonte: Ceará (2005, p. 15)

O item em questão pertence ao tema 4, tratamento da informação, da Matriz de Referência do 5º ano do EF do SPAECE e aborda o descritor D74, que versa sobre a habilidade de ler informações apresentadas em gráficos de barras ou colunas.

Sobre o item, pergunta-se o mês em que os alunos mais faltaram, portanto, os respondentes poderiam obter sua solução pela análise da quantidade de alunos

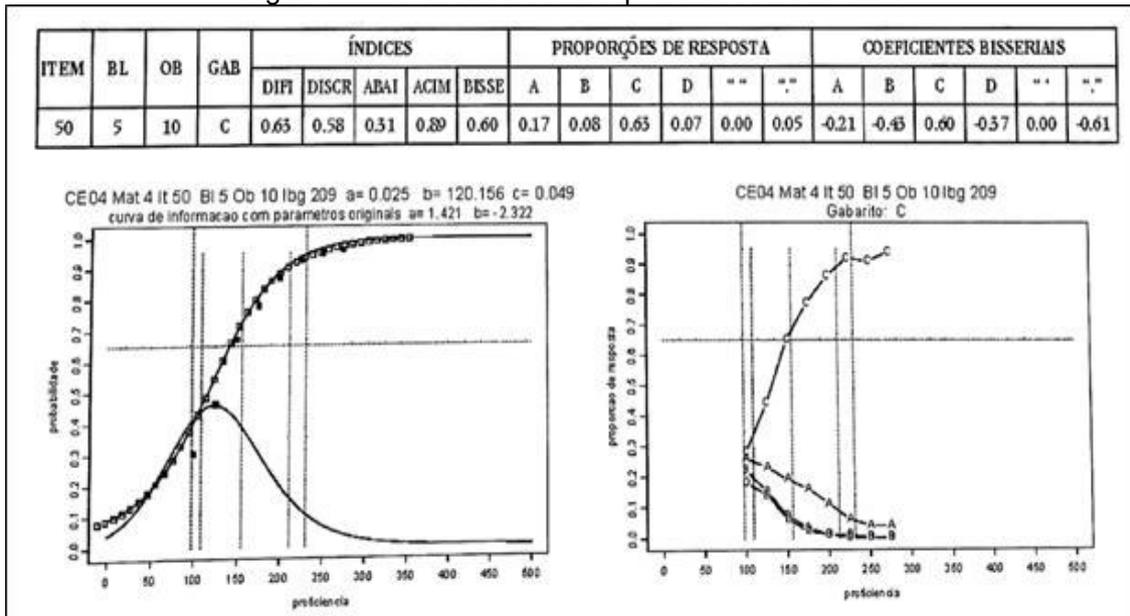
indicada no eixo vertical ou, compreendendo que o gráfico relaciona o mês e a quantidade de alunos faltantes, bastaria que fosse feita uma análise visual das alturas das colunas, não precisando realizar a leitura dos valores no eixo vertical. Assim, os respondentes identificariam o mês de maio como aquele em que houve mais faltas, correspondendo a resposta correta à alternativa C.

Observa-se que, no gráfico utilizado como suporte para o item, o eixo horizontal está ordenado cronologicamente segundo os meses em que foram realizadas as pesquisas dos alunos faltantes. Isso fez com que as colunas se apresentassem com alturas aleatórias, ou seja, não aparecessem ordenadas de modo crescente nem decrescente, o que oferece ao item uma pequena complexidade na sua interpretação, porém, nada que possa elevar a sua dificuldade.

Nota-se que o item atende aos aspectos da unidimensionalidade e da independência local, uma vez que busca avaliar uma única habilidade, ou, pelo menos, uma habilidade dominante, que é a de interpretar as informações do gráfico de barras. Sua solução também não está vinculada a nenhum outro item. Isso implica em dizer que o desempenho do estudante nesse item não afeta o seu desempenho em outro item, pois cada item é respondido exclusivamente em função do nível do traço latente do respondente (PASQUALI, 2018; PASQUALI; PRIMI, 2003; COUTO; PRIMI, 2011; ANDRADE; TAVARES; VALLE, 2000; TAVARES, 2013).

Na Figura 4, destacam-se a CCI e os índices dos parâmetros de estimação desse item. Vê-se, nessa figura, que o item apresentou Índice de Dificuldade (Difi) igual a 0,63, indicando que houve 63% de acertos pelos respondentes durante o pré-teste, o que o aponta como relativamente fácil. Isso também pode ser observado pela posição do gráfico da CCI, que está situado mais à esquerda, evidenciando que a projeção do seu ponto de inflexão no eixo da proficiência está localizado no valor $b=120,156$, indicando que um estudante com proficiência aproximadamente igual a 120, que é um valor relativamente baixo numa escala de 0 a 500, já possui 60% de chance de acertar esse item. Assim, reforça-se novamente a afirmação de que o item tem dificuldade baixa.

Figura 4 - CCI e índices dos parâmetros do Item 1



Fonte: Ceará (2005, p. 15)

O valor do Índice de Discriminação (Discr), igual a 0,58, mostra que o item tem a capacidade de separar adequadamente os alunos de melhor e pior desempenho. Isso também pode ser percebido pela elevada inclinação do gráfico na CCI, evidenciando um item com parâmetro de discriminação (a) alto.

Pelo Índice de Coeficiente Bisserial (Bisse), igual a 0,60, e pelo valor do parâmetro (c) abaixo de 0,1, percebe-se que o item é considerado bom, pois a alternativa correta não atraiu respondentes de baixa proficiência. Ao analisar os valores dos coeficientes bisseriais das alternativas, constatam-se valores negativos e relativamente distantes de zero para os distratores, o que confirma que o item é bom, pois as alternativas erradas atraíram apenas os alunos de baixo desempenho. Nota-se também que o distrator A, por ter apresentado um valor negativo mais próximo de zero (-0,21), foi a alternativa que mais atraiu os alunos de baixo rendimento, uma vez que eles, provavelmente, podem ter entendido que as colunas do gráfico do item indicavam a quantidade de alunos presentes e não os ausentes, sendo que esta má interpretação do item pode tê-los levado a entenderem que abril seria o mês com a menor quantidade de alunos presentes, consequentemente o mês em que os alunos mais faltaram.

Com base na habilidade necessária para responder ao item, pode-se constatar que ele requer do estudante uma proficiência aproximadamente igual a 120. Analisando esse valor pela distribuição dos níveis de proficiência de matemática no SAEB (Quadro 1), identifica-se que um aluno com essa proficiência está situado no Nível 1, que vai até 149. Analisando-se, contudo, pelos padrões de

desempenho do SPAECE (Quadro 2), um aluno com essa proficiência está situado no padrão de desempenho muito crítico, indicando que o estudante ainda apresenta carência de aprendizagem em relação às habilidades previstas para sua etapa de escolaridade.

De acordo com Ceará (2007), estudantes com proficiência 120, tratando-se do tema tratamento da informação, são capazes, apenas, de localizar informações relativas à maior ou menor frequência de elementos, em gráficos de colunas. Quanto às demais habilidades dos outros temas da Matriz de Referência, entre outras, eles são apenas capazes de resolver problemas simples envolvendo adição com pequenas quantidades; até já calculam o resultado de adição com reserva, mas ainda não dominam as demais operações; também já conseguem determinar a área de figuras desenhadas em malhas quadriculadas, porém apenas por meio da contagem simples.

Segue-se, agora, com a análise do item indicado na Figura 5.

Figura 5: Item 2

Observe as fichas do jogo de João. Elas estão desorganizadas.

3,6
 0,9
 1,8
 2,7

Para organizá-las em ordem crescente, João deve arrumá-las da seguinte maneira:

(A) 0,9 3,6 2,7 1,8

(B) 1,8 2,7 3,6 0,9

(C) 0,9 1,8 2,7 3,6

(D) 3,6 2,7 1,8 0,9

Fonte: Ceará (2005, p. 30)

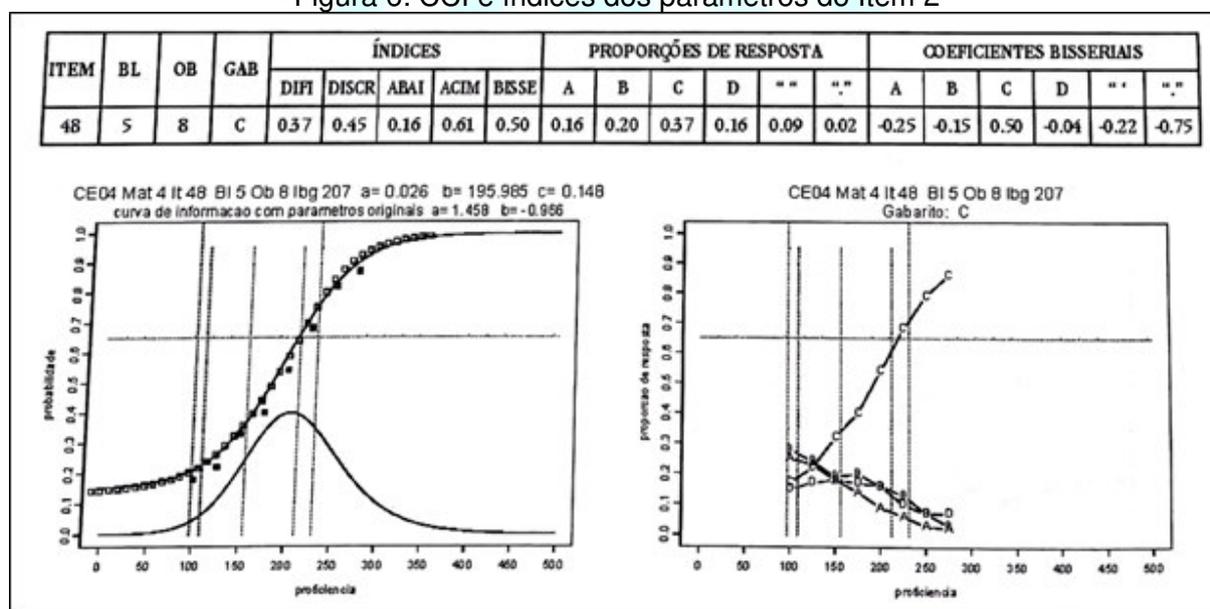
Este item, do 5º ano do EF, pertence ao tema interagindo com números e funções, da Matriz de Referência do SPAECE, contemplando o descritor D14, que compreende a habilidade de comparar números racionais na forma fracionária ou decimal.

Para resolver o item, os respondentes precisariam comparar os quatro números racionais na forma decimal. No caso, bastaria que eles comparassem as partes inteiras, que são todas diferentes, para depois ordenarem os números do menor para o maior, encontrando a sequência 0,9; 1,8; 2,7 e 3,6, identificada na alternativa C.

Assim como no item anterior, percebe-se o atendimento ao postulado da unidimensionalidade, pois para solucionar esse item o aluno utilizará, prioritariamente, uma habilidade, ou, pelo menos, uma aptidão dominante, que no caso é a habilidade de comparar números na forma decimal. Também se nota o atendimento ao pressuposto da independência local, haja vista o item não estabelecer relação de dependência com outro item, nem mesmo entre habilidades (PASQUALI, 2018; PASQUALI; PRIMI, 2003; COUTO; PRIMI, 2011; ANDRADE; TAVARES; VALLE, 2000; TAVARES, 2013).

Prossegue-se com a análise do item a partir da observação, na Figura 6, da sua CCI e dos índices dos parâmetros levantamento estatístico durante a estimação do item.

Figura 6: CCI e índices dos parâmetros do Item 2



Fonte: Ceará (2005, p. 30)

Pela observação do Índice de Dificuldade do Item (Difi), igual a 0,37, que é baixo, averigua-se que o item teve um pequeno percentual de acerto, ou seja, que apenas 37% dos respondentes o acertaram, logo, ele apresentou certa dificuldade, podendo até ser considerado mais difícil, pois o valor do Difi se aproximou de 0,30.

Analisando o Índice de Discriminação do Item (Discr), igual a 0,45, percebe-se que o item discrimina bem, pois apresentou valor superior a 0,25. Na CCI, também se evidencia esse fato pela elevada inclinação do seu gráfico, que é característica de um item que tem parâmetro de discriminação (a) mais alto.

O Índice de Coeficiente Bisserial (Bisse), igual a 0,50, com valor superior a 0,30, e o valor do parâmetro (c) entre 0,1 e 0,2, mostra que o gabarito do item não atraiu estudantes de baixa proficiência. Esse fato também pode ser constatado pelos

valores negativos dos coeficientes bisseriais dos distratores, evidenciando que eles atraíram apenas alunos de baixo rendimento. No entanto, é válido ressaltar que o distrator D, que apresentou valor $-0,04$, muito próximo de zero, foi a alternativa errada que mais atraiu os estudantes de baixa proficiência, provavelmente, pelo fato de essa alternativa trazer os números decimais em ordem decrescente, levando os alunos que ainda não consolidaram os conceitos de crescente e decrescente a marcarem essa alternativa.

Quanto à habilidade necessária para responder ao item, observando o gráfico na CCI, e pelo valor de $b=195,985$, compreende-se que um estudante com proficiência aproximada de 196 já possui 60% de chance de acertar esse item. Analisando esse valor pela distribuição dos níveis de proficiência do SAEB (Quadro 1), percebe-se que um aluno com essa proficiência está no Nível 3, que vai de 175 até 199. Quando se analisa pelos padrões de desempenho do SPAECE (Quadro 2), averigua-se que um aluno com essa proficiência está situado no padrão de desempenho crítico, evidenciando que o estudante ainda não demonstra um desenvolvimento adequado das habilidades esperadas para sua etapa de escolaridade.

Com base em Ceará (2007), os alunos nesse nível de proficiência (196), tratando-se do tema números e operações, já são capazes de, entre outras habilidades, localizarem números racionais em sua forma decimal, em uma reta numérica graduada em que estão expressos diversos números naturais consecutivos com dez subdivisões entre eles. Nos outros temas da Matriz de Referência, os estudantes com essa proficiência, entre outras habilidades, apontam: reconhecer as propriedades básicas de um polígono, associando-os aos seus respectivos nomes; já determinam o horário de início e fim de um evento, porém no formato de hora inteira; e já reconhecem o maior valor em uma tabela de dupla entrada cujos dados possuem até duas ordens.

Considerações finais

Durante as seções deste artigo, apresentaram-se os pressupostos da TCT e da TRI, estudando-se os parâmetros do item e, também, realizando-se uma análise de itens de matemática a partir desses parâmetros, com o propósito de contribuir para um melhor entendimento dos processos que permeiam os sistemas de avaliações externas e em larga escala, principalmente aqueles que envolvem o item e sua estimação.

Viu-se que a aplicação de procedimentos e técnicas derivadas da TCT e da TRI possibilitam a identificação da variável latente dos estudantes, pois, por meio de modelos matemáticos e estatísticos, percebe-se que é possível estabelecer uma relação entre a probabilidade de um aluno responder corretamente a um item e suas habilidades na área do conhecimento avaliado, tendo como elemento central os itens e não o teste como um todo (COELHO, 2014).

Compreende-se que os estudos provenientes da TRI vão para além das análises apresentadas neste artigo, pois permitem, por exemplo, a comparação entre populações distintas submetidas a provas diferentes, mas com alguns itens comuns, ou ainda, a comparação entre indivíduos da mesma população que tenham sido submetidos a diferentes provas, com ou sem itens comuns (ANDRADE; TAVARES; VALLE, 2000).

Com efeito, obteve-se como pressuposto de reflexão que, para uma interpretação pedagógica mais ampla acerca dos processos de aprendizagens e dos conhecimentos adquiridos pelos alunos, faz-se necessária uma análise que vá para além dos aspectos técnicos envolvidos nos parâmetros de um item.

Almeja-se que as análises aqui apresentadas proporcionem reflexões contributivas sobre os instrumentos e procedimentos adotados pelos sistemas de avaliações externas, auxiliando, assim, numa melhor interpretação de seus resultados, e vindo a complementar outras possibilidades de análises.

Referências

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

BRASIL. Ministério da Educação – MEC. **Relatório SAEB 2017**. Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira – INEP. Brasília: MEC/SEF, 2019.

CAED – Centro de Políticas Públicas e Avaliação da Educação. **Guia de elaboração de itens: Matemática**. Universidade Federal de Juiz de Fora – UFJF. Juiz de Fora: CAED, 2008.

CEARÁ. Secretaria da Educação Básica. **Sistema Permanente de Avaliação da Educação Básica do Ceará (SPAEBCE)**: Relatório Pedagógico – Análise de itens de matemática. Fortaleza: SEDUC/CESGRANRIO, 2005.

_____. Secretaria da Educação Básica. **Sistema Permanente de Avaliação da Educação Básica do Ceará (SPAEBCE)**: Relatório Técnico-Pedagógico de matemática. Fortaleza: SEDUC/CESGRANRIO, 2006.

_____. Secretaria de Educação Básica. **Sistema Permanente de Avaliação da Educação Básica do Ceará: SPAECE 2006**: Relatório Técnico-Pedagógico de matemática. Fortaleza: SEDUC; Rio de Janeiro: Fundação CESGRANRIO, 2007.

_____. Secretaria de Educação Básica. **Sistema Permanente de Avaliação da Educação Básica do Ceará**: Boletim de gestão escolar. Juiz de Fora: CAEd/UFJF, 2012.

_____. Secretaria de Educação Básica. **Sistema Permanente de Avaliação da Educação Básica do Ceará**: Boletim SPAECE 2018 – Relatório executivo. Juiz de Fora: CAEd/UFJF, 2019.

COELHO, E. C. **Teoria da resposta ao item: desafios e perspectivas em exames multidisciplinares**. Tese (Doutorado em Ciências). Universidade Federal do Paraná. Curitiba, 2014.

COUTO, G.; PRIMI, R. **Teoria de Resposta ao Item (TRI)**: conceitos elementares dos modelos para itens dicotômicos. Boletim de Psicologia, 2011, v. LXI, n. 134: p. 1-15. 2011.

CUNHA, D. A. **A teoria de resposta ao item na avaliação em larga escala**: um estudo sobre o Exame Nacional de Acesso do Mestrado Profissional em Matemática em Rede Nacional. Dissertação (Mestrado Profissional em Matemática em Rede Nacional - PROFMAT). Instituto Nacional de Matemática Pura e Aplicada – IMPA, Rio de Janeiro, 2014.

FREITAS, L. C. de. **Responsabilização, meritocracia e privatização**: conseguiremos escapar ao neotecnicismo. Trabalho apresentado no III Seminário de Educação Brasileira, Cedes-Unicamp, 28 fev. a 01 mar. Campinas, 2011.

GIL, A. C. **Métodos e Técnicas de Pesquisa Social**. 6. ed. São Paulo: Atlas, 2011.

KLEIN, R. **Alguns aspectos da teoria de resposta ao item relativos à estimação das proficiências**. Revista Ensaio: Avaliação Política Pública em Educação. 2013, v. 21, n. 78, p. 35-56. 2013.

LORD, F. M. **Uma teoria das pontuações dos testes**. Monografia psicométrica, v. 7, 1952.

MENDONÇA, J. **Análise da eficiência de estimação de parâmetros da TRI pelo Software ICL**. Dissertação (Mestrado em Estatística). Universidade Federal de Lavras – MG. Lavras, 2012.

MOREIRA JUNIOR, F. J. **Aplicações da teoria da resposta ao item (TRI) no Brasil**. Biometric Brazilian Journal, Jaboticabal, v. 24, n. 4, p.137-170, 2010.

PASQUALI, L. **TRI – Teoria de Resposta ao Item**: Teoria, Procedimentos e Aplicações. v. 1. Brasília: Editora Appris, 2018.

PASQUALI, L.; PRIMI, R. **Fundamentos da Teoria da Resposta ao Item – TRI**. Avaliação Psicológica, v. 2. p. 99-110. 2003.

SEVERINO, A. J. **Metodologia do trabalho científico**. 23. ed. São Paulo: Cortez, 2007.

SILVA, F. E. F. **Teoria de Resposta ao Item (TRI) em avaliações de matemática na EEM Professor Gabriel Epifânio dos Reis**. Dissertação (Mestrado Profissional

em Matemática em Rede Nacional - PROFMAT). Universidade Federal Rural do Seminário – UFERSA, Mossoró, 2015.

SILVA, J. A.; BELLEMAIN, P. B.; BORBA, R. E. S. E. **Análise de Itens da Provinha Brasil de Matemática referentes a Grandezas e Medidas**. Perspectivas da Educação Matemática. Seção Temática. INMA/UFMS. v. 9, n. 21, p. 724-743. 2016.

SANTOS, M. J. C.; ORTIGÃO, M. I. R. **Tecendo redes intelectivas na Matemática nos anos iniciais do ensino fundamental**: relações entre currículo e avaliação externa (SPAECE). Revista REMATEC. Universidade Federal do Rio Grande do Norte. Natal, RN: EDUFRN – Editora da UFRN, n. 22, p. 59-72, 2016.

PRODANOV, C. C.; FREITAS, E. C. **Metodologia do trabalho científico: métodos e técnicas da pesquisa e do trabalho acadêmico**. 2. ed. Novo Hamburgo: Feevale, 2013.

TAVARES, C. Z. **Teoria da Resposta ao Item**: uma análise crítica dos pressupostos epistemológicos. Estudos em Avaliação Educacional, São Paulo, v. 24, n. 54, p. 56-76, jan./abr. 2013.

Submetido em abril de 2020.

Aceito em julho de 2020.