

# A Teoria da Resposta ao Item como proposta metodológica no exame de seleção dos cursos técnicos integrados do IFRN

---

THIAGO VALENTIM MARQUES<sup>I</sup>

ELIARDO GUIMARÃES DA COSTA<sup>II</sup>

<http://dx.doi.org/10.22347/2175-2753v14i42.3595>

## Resumo

A Teoria da Resposta ao Item (TRI) vem ganhando cada vez mais espaço em testes de larga escala no Brasil, uma vez que este método adota independência entre itens e indivíduos. Assim, o objetivo geral deste trabalho é mostrar a possibilidade de se estimar a proficiência dos 21.039 candidatos que realizaram o Exame de Seleção 2020 do IFRN por meio da TRI, verificando se é mais vantajoso adotar essa metodologia em vez da Teoria Clássica dos Testes (TCT). Os resultados evidenciaram que a classificação final do certame sofreu alteração em praticamente todas as posições e isso se deve ao fato de a TRI utilizar a coerência pedagógica ao estimar a proficiência dos respondentes. Este trabalho se mostra útil para que gestores possam visualizar as vantagens de se utilizar a TRI no acesso discente.

**Palavras-chave:** Avaliação educacional. Curva característica do item. Ensino Técnico Integrado. Teoria Clássica dos testes. TRI.

Submetido em: 14/07/2021

Aprovado em: 23/03/2022

---

<sup>I</sup> Instituto Federal do Rio Grande do Norte (IFRN), Natal (RN), Brasil; <https://orcid.org/0000-0001-7806-4096>; e-mail: [thiago.valentim@ifrn.edu.br](mailto:thiago.valentim@ifrn.edu.br)

<sup>II</sup> Universidade Federal do Rio Grande do Norte (UFRN), Natal (RN), Brasil; <https://orcid.org/0000-0003-4528-0379>; e-mail: [eliardo.costa@ufrn.br](mailto:eliardo.costa@ufrn.br)

# Item Response Theory as a methodological proposal in the IFRN integrated technical course selection exam

## **Abstract**

The Item Response Theory (IRT) has been gaining more space in large-scale tests in Brazil, since this method adopts independence between items and individuals. Thus, the main goal of this work is to show the possibility of estimating the proficiency of the 21,039 took examinees of the IFRN 2020 Selection Examination through the IRT, checking if it is more advantageous to adopt this methodology instead of the Classical Test Theory (CTT). The results showed that the final classification of the contest changed in practically all positions, and this is because the IRT uses pedagogical coherence when estimating the proficiency of the respondents. This work proves to be useful for managers to visualize the advantages of using IRT in student access.

**Keywords:** Educational evaluation. Item characteristic curve. Integrated Technical Education. Classical Test Theory. IRT.

# La Teoría de Respuesta al Ítem como propuesta metodológica en el examen de selección de curso técnico integrado de IFRN

## **Resumen**

La Teoría de Respuesta al Ítem (TRI) viene ganando cada vez más espacio en las pruebas de gran escala en Brasil, ya que este método adopta la independencia entre ítems e individuos. Así, el objetivo general de este trabajo es mostrar la posibilidad de estimar la competencia de los 21.039 candidatos que tomaron el Examen de Selección IFRN 2020 a través de la TRI, comprobando si es más ventajoso adoptar esta metodología en lugar de la Teoría Clásica de los Tests (TCT). Los resultados mostraron que la clasificación final del concurso cambió en prácticamente todas las posiciones y esto se debe a que el TRI utiliza la coherencia pedagógica al estimar la competencia de los individuos. Este trabajo es útil para que los administradores visualicen las ventajas de usar IRT en el acceso de los estudiantes.

**Palabras clave:** Evaluación Educativa. Curva Característica del Ítem. Educación Técnica Integrada. Teoría Clásica de los Tests. TRI.

## Introdução

A avaliação objetiva dos fenômenos psicológicos é constituída pela utilização de instrumentos e testes psicométricos. O início do desenvolvimento destes instrumentos se deu no início do século XIX, paralelamente com o avanço da ciência positivista e da procura por medidas objetivas e válidas para o desenvolvimento de pesquisas clínicas. Nesse contexto, surge a necessidade de métodos que avaliassem as propriedades psicométricas dos instrumentos, sendo que na década de 1900, se deu o desenvolvimento da Teoria Clássica dos Testes (TCT) a partir das contribuições de Binet e Spearman (SARTES; SOUZA-FORMIGONI, 2013).

As análises e interpretações em exames de seleção tradicionalmente utilizam a TCT, isto é, os resultados são associados à prova como um todo e expressos em função de escores padronizados (ANDRADE; TAVARES; VALLE, 2000). A comparação entre indivíduos ou grupos de indivíduos somente é possível quando eles são submetidos às mesmas provas, o que torna a TCT defasada quando se quer avaliar desempenhos em anos distintos (KLEIN, 2009). Além disso, a condição típica de exames dessa natureza é maximizar respondentes de habilidades medianas, o que torna os instrumentos bem menos apropriados para avaliar respondentes com habilidades superiores ou de pouca habilidade (PASQUALI, 2009).

A TCT apresenta como característica marcante o fato de que todas as questões fornecem a mesma quantidade de informação sobre o conhecimento que os participantes dominam. De fato, nota-se que essa não é a melhor opção metodológica, haja vista que há itens que são mais informativos do que outros no contexto do que está sendo avaliado (INEP, 2012). Diante deste problema, a Teoria da Resposta ao Item (TRI) surge como instrumento avaliativo de testes a partir de meados da década de 80 como alternativa à TCT (PASQUALI; PRIMI, 2003). A TRI propõe modelos para os traços latentes, que são características não observáveis do indivíduo (PASQUALI, 2009; ANDRADE; TAVARES; VALLE, 2000).

A TRI adota dois axiomas fundamentais: (i) o desempenho dos respondentes em um item se explica em função de um conjunto de traços latentes (também chamados de proficiência); e (ii) a relação entre o desempenho no item e o conjunto dos traços latentes pode ser descrita pela Curva Característica do Item (CCI), na qual indivíduos de maior proficiência terão maior probabilidade de responder corretamente o item (PASQUALI, 2009). O fato de a TRI captar essas informações torna

essa metodologia muito mais eficiente para se medir a proficiência dos candidatos em avaliações (INEP, 2012). No Brasil, a avaliação de testes educacionais aplicados em larga escala vem utilizando cada vez mais a TRI. Por exemplo, no Sistema de Avaliação da Educação Básica (Saeb) e no Exame Nacional do Ensino Médio (ENEM).

Ao utilizar a TRI, em vez da TCT, como metodologia em avaliações educacionais surgem uma série de benefícios no planejamento de testes. Um deles está relacionado com a possibilidade de as instituições criarem um banco de itens previamente calibrados com uma amostra suficientemente grande e heterogênea da população de respondentes (HAMBLETON; SLATER, 1997; TAVARES, 2013). Com a TRI também é possível elaborar instrumentos avaliativos distintos para o mesmo exame, o que torna possível a comparabilidade entre os anos.

No que diz respeito ao ingresso dos alunos nos cursos técnicos integrados do Instituto Federal de Educação, Ciência e Tecnologia do Rio Grande do Norte (IFRN), há um Exame de Seleção anual, promovido pela própria instituição, composto por vinte questões objetivas de Língua Portuguesa (LP), vinte questões objetivas de Matemática e uma redação. A nota de classificação de um candidato é calculada pela média aritmética a partir de um escore padronizado para cada um dos conhecimentos avaliados, ou seja, o IFRN adota a TCT para ranquear os candidatos e selecionar os classificados dentro do número estipulado de vagas.

Diante do exposto, tem-se o seguinte problema de pesquisa: do ponto de vista pedagógico, seria mais vantajoso para o IFRN utilizar a TRI como metodologia para a construção de escalas avaliativas e classificatórias no Exame de Seleção? A motivação para a escolha deste tema se deu pela possibilidade de aplicar um método mais eficiente no que diz respeito a seleção pela aptidão dos candidatos que farão parte dos cursos técnicos integrados do IFRN. Acredita-se na hipótese de que a aplicação da TRI como metodologia seja um método mais adequado para ranquear os candidatos no Exame de Seleção.

Assim, a presente pesquisa, classificada como quantitativa (GONÇALVES, 2005), tem como objetivo mostrar a possibilidade de se estimar a proficiência dos 21.039 candidatos que realizam o Exame de Seleção 2020 do IFRN por meio da Teoria da Resposta ao Item, e assim ranqueá-los de forma mais coerente. Os dados utilizados nesta pesquisa são referentes ao Exame de Seleção aplicado em 2019 relativos às

provas de LP e Matemática. Além disso, pretende-se mostrar a utilidade da TRI para gestores de exames de seleção e fomentar o seu uso nos respectivos certames.

### Uma breve abordagem da Teoria Clássica dos Testes

O objetivo da TCT é explicar o resultado total de uma avaliação, expresso pelo escore total ( $T$ ) obtido pela quantidade de itens assinalados corretamente (PASQUALI, 2009). Por exemplo, se uma prova é composta por 20 itens e um indivíduo acerta 16 deles, então, pela TCT, tenta-se explicar a seguinte pergunta: o que este 16 significa para este candidato com relação à prova? Desta forma, a TCT se preocupa em produzir tarefas (ou testes) de qualidade para que essa pergunta possa ser respondida.

De maneira geral, o modelo da TCT, elaborado pelo psicólogo inglês Charles Edward Spearman, é dado pela Equação 1:

$$T = V + E, \quad (1)$$

em que  $T$  é o escore bruto ou empírico do indivíduo, que é a soma dos pontos obtidos no teste;  $V$  é o escore verdadeiro, ou seja, a inteligência real do sujeito que o teste quer medir; e  $E$  o erro de medida. Ressalta-se que os erros são explicados por um conjunto de fatores estranhos, como por exemplo defeitos do próprio teste, estereótipos e vieses do sujeito, fatores históricos e ambientais aleatórios (PASQUALI, 2009; 2017; 2018).

Ao utilizar a TCT algumas métricas se destacam na divulgação dos resultados e na análise dos itens, tais como: escore padronizado, índice de facilidade, coeficiente de discriminação, entre outros. Quando um processo seletivo utiliza a TCT, geralmente a classificação é realizada a partir do escore padronizado calculado para cada candidato. Por exemplo, o Exame de Seleção do IFRN utiliza o escore padronizado conforme a Equação 2:

$$Escore = \frac{NT_{candidato} - MédiaNT_{candidatos}}{DP} \times 100 + 500, \quad (2)$$

em que  $NT_{candidato}$  é a nota do candidato na respectiva prova, cujo valor é de 100 pontos distribuídos igualmente entre todos os itens;  $MédiaNT_{candidatos}$  é a média das notas dos candidatos a um mesmo *campus*/curso/turno;  $DP$  é o desvio-padrão das notas dos candidatos a um mesmo *campus*/curso/turno (INSTITUTO FEDERAL DE EDUCAÇÃO, CIÊNCIA E TECNOLOGIA DO RIO GRANDE DO NORTE, 2019).

É notória a utilidade da TCT para o desenvolvimento de métricas focadas em avaliações psicológicas, haja visto que ainda continua sendo amplamente utilizada. Por outro lado, muitos estudos têm discutido suas limitações há décadas. Em particular, Sartes e Souza-Formigoni (2013) elencou algumas limitações da TCT, por exemplo: (i) todas as medidas são dependentes da amostra dos indivíduos que responderam o instrumento; (ii) exames diferentes com índice de facilidade e discriminação diferentes geram resultados diferentes para os mesmos indivíduos; (iii) quando o mesmo conhecimento for medido por dois testes diferentes, implica que os resultados não são expressos na mesma escala, o que gera impedimento para se realizar uma comparação direta. A TCT também apresenta fragilidades do ponto de vista estatístico, uma vez que não é possível avaliar a significância de um item para a decisão de incluí-lo ou excluí-lo do teste (ARANTES, 2016).

### **A Teoria da Resposta ao Item**

A TRI é considerada como um dos métodos modernos da teoria dos testes e, como o próprio nome sugere, a unidade de análise é a resposta do item de maneira individual. A TRI vem ganhando destaque tanto na adoção desse método em testes de larga escala (por exemplo o ENEM) quanto em trabalhos científicos. Neste método, o desempenho do respondente em cada item é o efeito e os traços latentes são a causa. Trata-se de um modelo matemático que utiliza métodos estatísticos nas estimações dos parâmetros que compõem o modelo. De maneira geral, essa metodologia busca representar a probabilidade de um indivíduo acertar um item,  $P(\theta)$ , como função dos seus traços latentes,  $\theta$ , na área do conhecimento avaliada (PASQUALI, 2009; ANDRADE; TAVARES; VALLE, 2000; BAKER; KIM, 2017).

A modelagem de traços latentes em avaliações educacionais relaciona os itens de uma prova (variáveis observáveis) com traços hipotéticos não observáveis ou aptidões, que seriam inerentes ao sujeito e responsáveis pela resposta ao item (PASQUALI; PRIMI, 2003). Nota-se que a TRI adota uma independência do instrumento de medida em relação ao objeto que se deseja medir e essa vantagem em relação à TCT justifica a sua utilização na área de educação em vários países (TAVARES, 2013). Uma grande vantagem da TRI é a possibilidade da elaboração de provas distintas para o mesmo exame, ou seja, é possível aplicar provas distintas em diferentes períodos do ano com o mesmo grau de dificuldade.

É razoável partir da premissa de que um indivíduo com maior proficiência em um item terá maior probabilidade de acertar esse item do que um respondente de menor proficiência. Considerando que a resposta do  $i$ -ésimo respondente ao  $k$ -ésimo

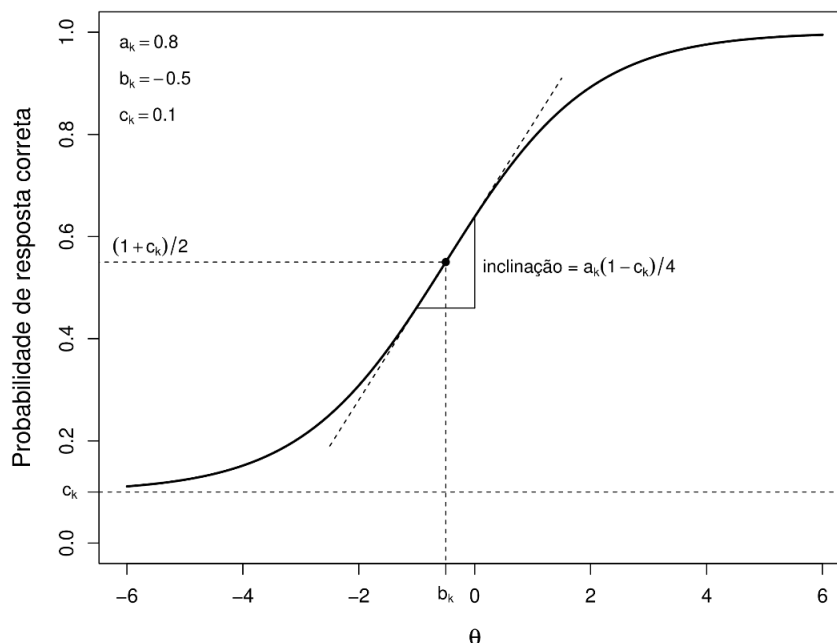
item, denotada por  $Y_{ik}$ , pode assumir dois valores (em geral, 0 para incorreto e 1 para correto), a relação entre a probabilidade de acerto,  $P(Y_{ik} = 1|\theta_i)$ , e a proficiência,  $\theta_i$ , pode ser modelada por uma função matemática monótona crescente em formato de S chamada Curva Característica do Item (CCI) (ANDRADE; TAVARES; VALLE, 2000; BAKER, 2001; PASQUALI, 2009; AYALA, 2013; BAKER; KIM, 2017; PAEK; COLE, 2019).

O Modelo Logístico de Três Parâmetros (ML3), desenvolvido por Birnbaum (1968), frequentemente utilizado para descrever a CCI dos itens em avaliações educacionais é apresentado na Equação 3:

$$P(Y_{ik} = 1|\theta_i, a_k, b_k, c_k) = c_k + (1 - c_k) \frac{e^{a_k(\theta_i - b_k)}}{1 + e^{a_k(\theta_i - b_k)}}, \quad (3)$$

sendo  $a_k$  o parâmetro de discriminação,  $b_k$  o parâmetro de dificuldade e  $c_k$  o parâmetro de acerto casual, relativos ao k-ésimo item do teste. Gráficamente  $a_k$  está relacionado com a inclinação,  $b_k$  a abscissa quando a probabilidade de resposta correta é 0,5 e  $c_k$  a interseção da CCI com o eixo das ordenadas (Figura 1). A estimação do traço latente geralmente é realizada por Máxima Verossimilhança via algum método iterativo (Newton-Raphson ou Scoring de Fisher) ou procedimentos bayesianos (ANDRADE; TAVARES; VALLE, 2000; BAKER, 2001; AYALA, 2013; BAKER; KIM, 2017; PAEK; COLE, 2019).

Figura 1 – Exemplo da Curva Característica do Item de acordo com o Modelo Logístico de Três Parâmetros



Fonte: Os autores (2021).

Quando se fala em ter informação, significa dizer que se sabe algo sobre um objeto ou um tópico específico. Na Estatística, Sir Ronald Aylmer Fisher propôs uma definição de informação como o recíproco da variância do estimador de um parâmetro (AYALA, 2013). Com base nisso, Lord (1980) demonstrou que a informação fornecida pelo item  $k$  no nível de habilidade  $\theta_i$ , considerando o modelo ML3, é dada pela Equação 4:

$$I_k(\theta_i) = a_k \frac{Q(\theta_i)}{P(\theta_i)} \left[ \frac{P(\theta_i) - c_k}{1 - c_k} \right]^2, \quad (4)$$

em que  $P(\theta_i) = P(Y_{ik} = 1 | \theta_i, a_k, b_k, c_k)$  e  $Q(\theta_i) = 1 - P(\theta_i)$ . Nota-se que os parâmetros  $a_k$ ,  $b_k$  e  $c_k$  influenciam no aumento da informação do item nos seguintes casos: (i) quando  $b_k$  se aproxima de  $\theta_i$ ; (ii) quanto maior for  $a_k$ ; e (iii) quanto mais  $c_k$  se aproxima de 0.

A função de informação do teste é definida pela soma das informações fornecidas pelos  $K$  itens do teste (Equação 5):

$$I(\theta_i) = \sum_{k=1}^K I_k(\theta_i), \quad (5)$$

também sendo possível a representação através do erro padrão de medida, ou ainda, erro padrão de estimação (na TRI), dada por (Equação 6):

$$EP(\theta_i) = \frac{1}{\sqrt{I(\theta_i)}}. \quad (6)$$

tanto a função de informação quanto o erro padrão de estimação são relevantes para avaliar o teste como um todo e geralmente são representadas na mesma janela gráfica (ANDRADE; TAVARES; VALLE, 2000; PASQUALI, 2009; BAKER; KIM, 2017; PAEK; COLE, 2019).

No que diz respeito à escala do traço latente, teoricamente os valores de  $\theta$  podem variar de  $-\infty$  a  $+\infty$ , mas geralmente se usa uma escala padrão (0,1), ou seja, centrada no zero e desvio padrão igual a uma unidade para que a partir disso seja utilizada uma outra escala por meio de uma transformação linear. Ressalta-se que o



ENEM utiliza a escala (500, 100) com base nas provas do exame realizadas no ano de 2009 e o Saeb adota a escala de proficiência em torno de (250, 50) para todos os anos de aplicação nas séries escolares (ANDRADE; TAVARES; VALLE, 2000; ANDRADE; LAROS; GOUVEIA, 2010; INEP, 2012).

## Procedimentos metodológicos

### Caracterização do IFRN

O IFRN atualmente tem cerca de 28 mil alunos matriculados, divididos em 21 *campi* que estão distribuídos por todas as regiões do estado (Figura 2). Os cursos ofertados pela instituição, de maneira presencial e à distância, atendem várias modalidades, destacando-se as seguintes: Técnico Integrado – que concatena o curso técnico ao ensino médio; Técnico Subsequente – que promove o ensino técnico a quem já possui o ensino médio; Graduação – nas formas tecnológica e licenciatura; e Pós-Graduação – com especialização, mestrado e doutorado (INSTITUTO FEDERAL DE EDUCAÇÃO, CIÊNCIA E TECNOLOGIA DO RIO GRANDE DO NORTE, 2021).

Figura 2 – Localização dos 21 *campi* do Instituto Federal de Educação, Ciência e Tecnologia do Rio Grande do Norte



Fonte: Os autores (2021).

### *Dados disponibilizados pelo IFRN*

As informações referentes aos dados do Exame de Seleção 2020, que ocorreu em 2019, foram disponibilizadas pelo IFRN por meio da Coordenação de Acesso Discente, para qual um termo de confidencialidade, sigilo e compromisso de uso exclusivamente acadêmico foi assinado pelo primeiro autor. O banco de dados contempla diversas variáveis relacionadas aos 21.039 candidatos presentes, tais como: *campus*, *oferta*, *turno*, *inscrição*, *candidato*, *escola pública*, *renda familiar*, *etnia*, *pessoa com deficiência*, *lista*, *escore*, *nota*, *status* e *gabarito*. Ressaltasse que não é objetivo deste estudo avaliar os conteúdos abordados nos itens do Exame de Seleção e o que de fato tenha motivado os resultados que serão evidenciados. Além disso, dois itens foram anulados após recursos de candidatos: item 7 da prova de LP e item 26 da prova de Matemática. Assim, ambas as provas contêm 19 itens com a numeração original mantida.

### *Modelo utilizado na aplicação da TRI*

Uma vez consideradas as respostas dos itens de LP e Matemática de maneira dicotomizada, 1 para acerto e 0 para erro, adotou-se o modelo ML3 para modelar a probabilidade de acerto. Pelo fato do Exame de Seleção do IFRN adotar a TCT para o cálculo dos escores, os itens deste exame não foram calibrados antes do certame, isto quer dizer que tanto os parâmetros dos itens quanto os traços latentes dos candidatos tiveram que ser estimados. Para isto, foi adotado o método EAP (do inglês *expected a posteriori*) e o método da quadratura, que consiste em substituir a integral por uma soma finita de áreas de retângulos (ANDRADE; TAVARES; VALLE, 2000).

### *Análises estatísticas*

As análises estatísticas foram realizadas com o apoio computacional da linguagem R versão 4.0.3 (R CORE TEAM, 2021) com o auxílio dos pacotes *readxl* (WICKHAM; BRYAN, 2019) para ler planilhas do *Microsoft Excel*; *tidyverse* (WICKHAM et al., 2019) com o objetivo de realizar filtros, obter estatísticas sumarizadas dos dados e produzir gráficos mais elaborados; e *Itm* (RIZOPOULOS, 2006) para estimar os parâmetros dos itens e os traços latentes dos respondentes. Itens que apresentaram valores de  $a_k < 0,7$  foram considerados inadequados e removidos, haja vista que foram pouco informativos e possivelmente não relacionados com as proficiências dos candidatos (TEZZA; BORNIA; ANDRADE, 2011).

## Resultados e discussões

### Estimação dos parâmetros dos itens

As estimações dos parâmetros dos 19 itens de LP e dos 19 itens de Matemática do Exame de Seleção do IFRN, considerando o modelo ML3 da TRI na escala (0, 1), além das respectivas informações máximas, são apresentadas na Tabela 1. Os itens 6, 13, 16 e 20 de LP apresentaram inconsistências no que se refere a qualidade do parâmetro  $a_k$  e isto implicou na retirada destes itens no momento das estimações das proficiências dos candidatos nesta disciplina. Além disso, os valores de  $a_k$  foram predominantemente maiores para os itens de Matemática, o que implicou em itens mais informativos do que os de LP.

Tabela 1 - Parâmetros estimados dos itens de Língua Portuguesa e Matemática do Exame de Seleção 2020 do IFRN para o modelo ML3 da TRI

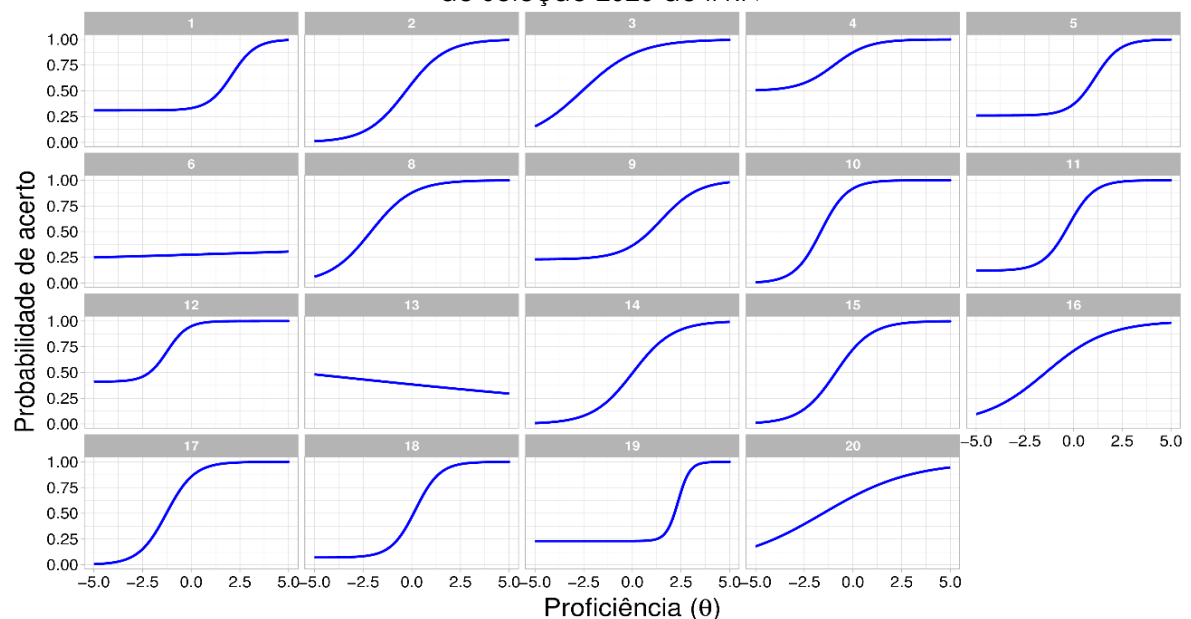
Língua Portuguesa					Matemática				
Item $k$	$a_k$	$b_k$	$c_k$	$\max\{I_k(\theta)\}$	Item $k$	$a_k$	$b_k$	$c_k$	$\max\{I_k(\theta)\}$
1	1,64	2,09	0,31	0,37	21	1,75	-0,34	0,14	0,58
2	0,98	-0,29	0,00	0,24	22	1,64	0,19	0,14	0,51
3	0,70	-2,57	0,00	0,12	23	3,24	1,17	0,46	1,04
4	1,12	-0,97	0,50	0,11	24	2,20	1,84	0,30	0,67
5	1,57	1,11	0,26	0,37	25	3,34	1,33	0,16	2,03
6	0,03	34,56	0,02	0,00	27	1,61	0,45	0,30	0,36
8	0,94	-2,10	0,00	0,22	28	1,95	0,29	0,24	0,60
9	1,04	1,49	0,23	0,17	29	1,25	2,11	0,27	0,24
10	1,47	-1,65	0,00	0,54	30	3,12	0,87	0,26	1,46
11	1,52	-0,25	0,12	0,45	31	2,60	-0,01	0,35	0,84
12	1,90	-1,24	0,41	0,40	32	2,98	2,27	0,26	1,33
13	-0,08	-5,92	0,00	0,00	33	2,11	0,35	0,18	0,79
14	0,97	0,03	0,00	0,24	34	2,70	0,78	0,25	1,13
15	1,10	-0,88	0,00	0,30	35	2,50	1,67	0,25	0,97
16	0,63	-1,42	0,00	0,10	36	2,37	0,58	0,19	0,98
17	1,40	-1,27	0,00	0,49	37	1,95	2,24	0,22	0,63
18	1,60	0,19	0,07	0,56	38	2,04	0,49	0,21	0,69
19	3,31	2,35	0,23	1,77	39	2,20	0,51	0,21	0,81
20	0,44	-1,53	0,00	0,05	40	1,61	1,00	0,12	0,52

Fonte: Os autores (2021).

No que diz respeito à dificuldade,  $b_k$ , os itens de Matemática se mostraram, de forma predominante, mais difíceis do que os de LP e variaram dentro do intervalo  $(-0,34; 2,27)$ . Ademais, percebe-se que os itens 6 e 13 apresentaram dificuldades razoavelmente discrepantes (34,56 e -5,92) quando comparados aos demais itens que compuseram a prova. Sobre o parâmetro  $c_k$ , houve convergência do algoritmo de estimação dos parâmetros do modelo para os dados das duas provas quando foi adotado a probabilidade máxima de acerto casual igual a 0,5. Todos os itens de Matemática apresentaram  $c_k > 0$ , enquanto dez itens de LP tiveram  $c_k = 0$ . Isto sugere que os itens de LP são menos sujeitos a chutes, ou seja, acerta o item quem realmente possui a proficiência. Por outro lado, isso não acontece em Matemática, o que de certa forma impactou em uma diminuição nas respectivas informações máximas.

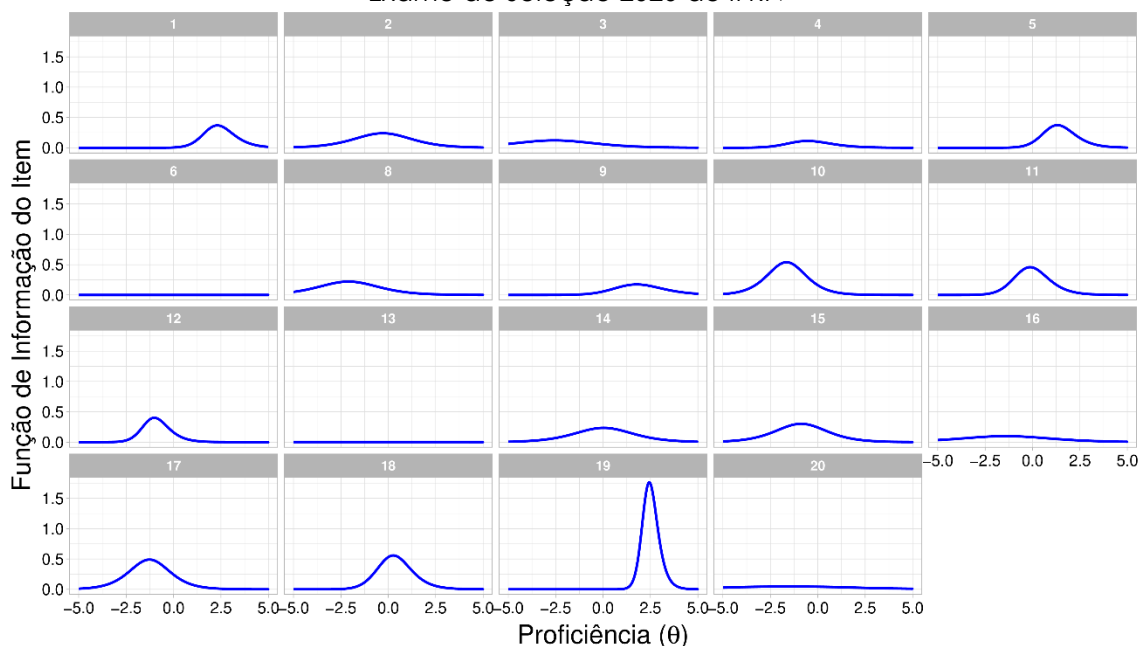
Para facilitar a interpretação dos parâmetros discutidos, as CCI e as curvas de informação dos itens de LP e Matemática são apresentadas nas Figuras 3-6. Conforme comentado anteriormente, quatro itens de LP (6, 13, 16 e 20) apresentaram informação praticamente nula e foram excluídos. Em contrapartida, o item 18 é o que apresenta maior informação e está em torno do traço latente 2,5, ou seja, respondentes cuja aptidão está a 2,5 desvios-padrão da média 0.

Figura 3 – Curvas características dos itens de Língua Portuguesa referentes a prova do Exame de Seleção 2020 do IFRN



Fonte: Os autores (2021).

Figura 4 – Curvas de informação dos itens de Língua Portuguesa referentes a prova do Exame de Seleção 2020 do IFRN



Fonte: Os autores (2021).

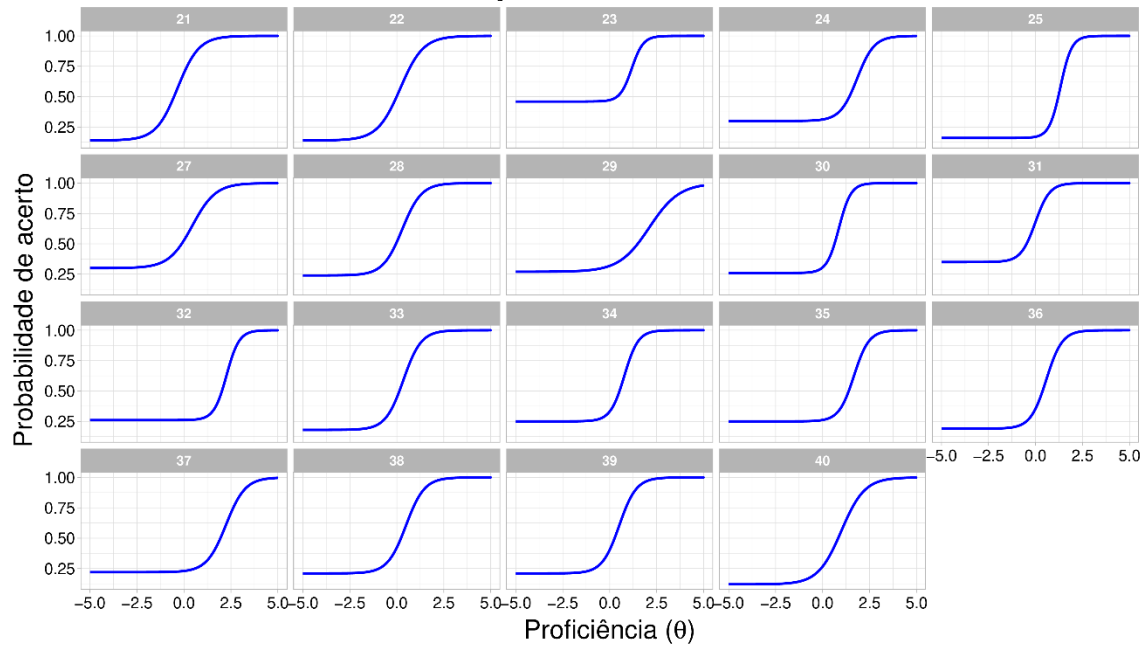
Para Matemática, o item 29 apresenta a menor informação máxima (0,24), ou seja, uma das menores discriminações (1,25), embora seja considerada elevada  $a_k > 1$  (ANDRADE; TAVARES; VALLE, 2000), e a probabilidade de acerto casual relativamente alta (0,27), o que favorece a diminuição da informação. Já o item 25 tem a maior discriminação (3,34) e está entre as menores probabilidades de acerto ao acaso (0,16). Assim, este item apresenta dificuldade 1,33 e a informação máxima está concentrada na vizinhança desse valor.

Os gráficos presentes nas Figuras 3-6 são informativos para se tomar decisões acerca dos itens que irão compor um determinado teste, o que poderá ser feito, na prática, se os itens passarem por uma etapa de calibração com uma amostra de respondentes com as mesmas características da população que será submetida ao teste. Isto possibilita que antes da realização do exame se conheça as faixas de proficiência que se deseja avaliar.

Quando a calibração não é realizada, como no presente estudo e em estudos similares, realizados por Moreira (2020), para a prova de Matemática do Exame de Seleção 2020 dos cursos técnicos subsequentes do IFRN, e Moreira e Nogueira (2020) para a prova da Olimpíada Brasileira de Matemática das Escolas Públicas (OBMEP), se torna possível realizar uma análise dessas faixas de proficiência com a estimação

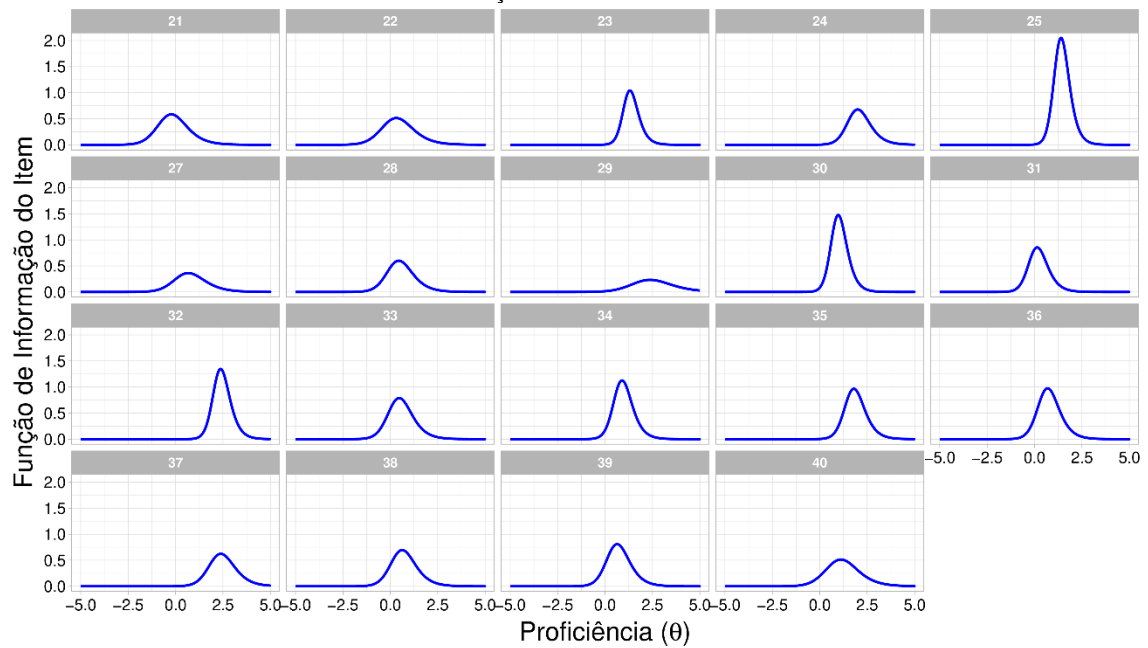
dos parâmetros dos itens após a coleta dos dados referentes às respostas dos candidatos.

Figura 5 – Curvas características dos itens de Matemática referentes a prova do Exame de Seleção 2020 do IFRN



Fonte: Os autores (2021).

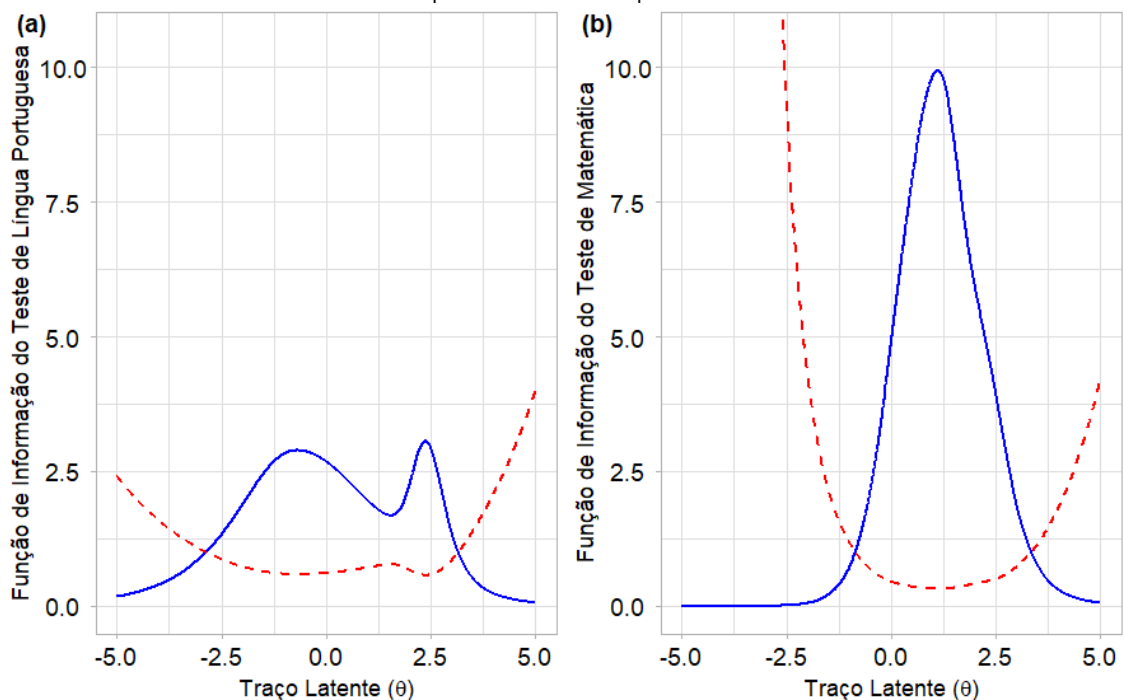
Figura 6 – Curvas de informação dos itens de Matemática referentes a prova do Exame de Seleção 2020 do IFRN



Fonte: Os autores (2021).

No que se refere a função de informação do teste e o respectivo erro padrão para LP (Figura 7a), já com os itens inadequados removidos, percebe-se que esta prova está avaliando melhor indivíduos com habilidade em torno de -2,5 a 2,5. Em Matemática (Figura 7b) há maior concentração das habilidades por volta de -1 e 3, isto é, maior informação captada para indivíduos com habilidades maiores que a média 0. Além disso, apresentou máximo por volta de 10 (cerca de três vezes maior do que o máximo em LP) e concentrada em um intervalo de habilidade razoavelmente menor quando comparada com LP. Por outro lado, o erro padrão de estimação teve crescimento aproximadamente exponencial para habilidades menores que -1.

Figura 7 – Curvas de informação do teste para as provas de Língua Portuguesa (a) e Matemática (b) do Exame de Seleção 2020 do IFRN. A curva tracejada em vermelho representa o erro padrão



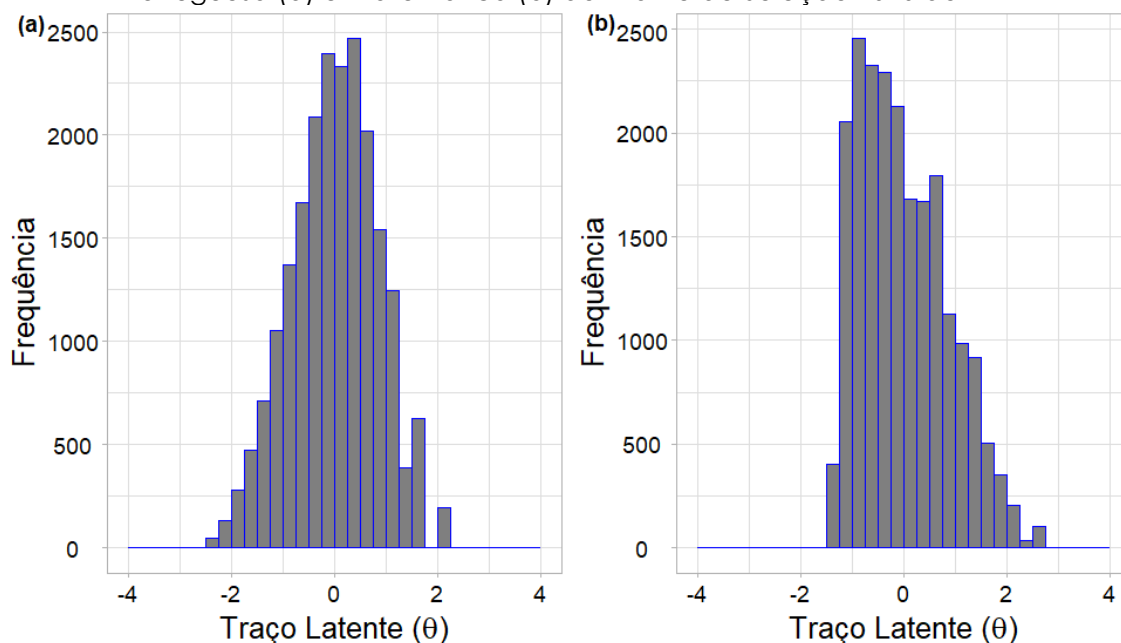
Fonte: Os autores (2021).

#### Análise descritiva da proficiência dos candidatos

No quesito estimação da proficiência em LP, cada respondente teve um valor de  $\theta$  estimado considerando a escala (0,1), conforme histograma apresentado na Figura 8a. As proficiências estimadas sugerem uma distribuição simétrica em torno de 0, desvio padrão 0,84, mínimo -2,43 e máximo 2,07. Já com relação à Matemática, a distribuição de  $\theta$  evidenciada é assimétrica à direita com média 0, desvio padrão

0,86, mínimo -1,34 e máximo 2,54 (Figura 8b), confirmando o que já era esperado dado os maiores valores de  $b_k$  em relação aos itens de LP.

Figura 8 – Histogramas do traço latente  $\theta$  dos 21.039 respondentes dos itens de Língua Portuguesa (a) e Matemática (b) do Exame de Seleção 2020 do IFRN



Fonte: Os autores (2021).

Um gráfico recorrente em livros e artigos sobre TRI é o gráfico de dispersão entre o número de acertos e a proficiência. Esse gráfico é importante para verificar a amplitude da proficiência para cada total de itens acertados. Por outro lado, neste trabalho preferiu-se utilizar *boxplots* (Figura 9), visto que há um melhor resumo descritivo contemplando cinco medidas: mínimo, primeiro quantil, mediana, terceiro quantil e máximo, além da informação sobre valores discrepantes (*outliers*).

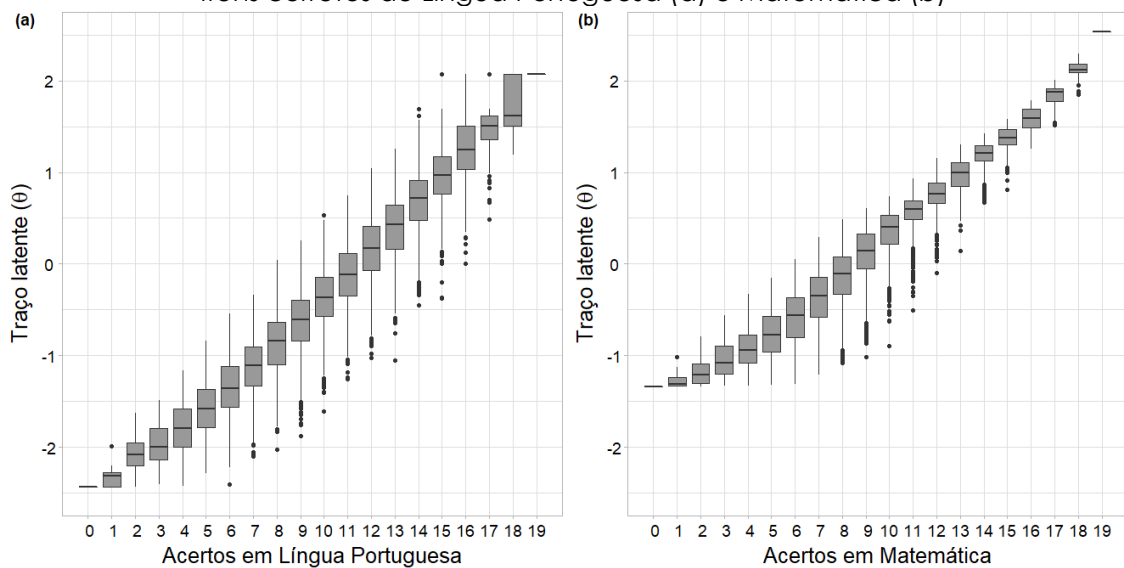
Em ambas as provas, a dispersão dos traços latentes vai aumentando com o aumento de acertos (considerando os 19 itens em cada disciplina) e, gradativamente, vai diminuindo até se tornar 0 para os respondentes que acertaram todos os itens. Além disso, percebe-se a presença de *outliers*, predominantemente na parte inferior dos *boxplots* e uma possível justificativa para esse fato são muitos itens acertados ao acaso. Em contrapartida, os *outliers* presentes na parte superior dos *boxplots* em LP (Figura 9a) podem ser justificados por 10, 14, 15 e 17 acertos majoritariamente por itens que passaram pelo controle de qualidade do parâmetro  $a_k$ .

Também pode ser observado que, por exemplo, candidatos com 5 itens acertados em LP podem ter proficiência maior do que candidatos que acertaram 13



itens e candidatos que acertaram 15 itens, sendo exclusivamente todos considerados adequados pela TRI, tiveram a mesma pontuação dos candidatos que acertaram todos os 19 itens que compuseram a prova (Figura 9a). Em Matemática existem respondentes com 6 acertos que chegam a ter pontuação maior do que respondentes com 12 acertos (Figura 9b). A justificativa para esse fato é a coerência pedagógica, que é evidenciada, por exemplo, no Guia do Candidato do ENEM (INEP, 2012).

Figura 9 – Boxplots da proficiência dos 21.039 respondentes agrupados por quantidade de itens corretos de Língua Portuguesa (a) e Matemática (b)



Fonte: Os autores (2021).

### Comparação entre a TRI e a TCT

O Exame de Seleção do IFRN aplica a TCT ao calcular o escore padronizado dos candidatos, os estratificando por *campus*, curso e turno. Na aplicação da TRI, todos os candidatos podem ter seus escores comparáveis, sem a necessidade de realizar a estratificação. Assim, para efeitos de comparação na classificação final, foi tomado, por exemplo, o resultado da seleção para o Curso Técnico Integrado em Eletrônica, turno vespertino do *Campus* Natal - Zona Norte.

A nota final, via TRI, foi dada pela média aritmética das proficiências  $\theta_i$  em LP e Matemática na escala (500,100), além da nota da redação considerando a correção variando de 0 a 1000 pontos. Ao considerar os 87 candidatos que tiveram a redação corrigida, notou-se mudanças em praticamente todas as posições ao utilizar a TRI como metodologia (Tabela 2), sendo nas posições centrais que se notou as maiores alterações. Por exemplo, o 16º colocado da classificação via TRI ficou em

34° na TCT e o 21° da TRI estava em 40° na TCT. Essa alteração na classificação, mesmo com um número de acertos menor do que candidatos que estiveram em posições mais baixas, pode ser justificada devido a uma maior coerência pedagógica do candidato quando comparado com outros candidatos com mais acertos, ou seja, candidatos que possivelmente tenham acertado itens que exigiram maior proficiência (ou itens inadequados pelo controle de qualidade) e errado itens considerados fáceis. Estes resultados são consistentes com o estudo realizado por Moreira e Nogueira (2020) na comparação na classificação entre a TRI e a TCT na prova da OBMEP.

Tabela 2 – Classificação final de parte dos candidatos do Exame de Seleção do IFRN que optaram pelo Curso Técnico Integrado em Eletrônica, turno vespertino do Campus Natal - Zona Norte ao considerar a TRI como metodologia

Posição TRI	Posição TCT	Acertos PORT.	Acertos MAT.	Escore TRI PORT.	Escore TRI MAT.	Nota da Redação ( $\times 100$ )	Escore final TCT	Escore final TRI
1	1	18	18	707	709	950	721	789
2	3	16	16	657	644	850	660	717
3	11	15	10	604	545	950	609	700
4	4	16	14	618	618	850	641	695
5	8	17	12	707	574	800	625	694
6	6	15	15	583	653	830	635	689
7	9	15	13	562	613	880	625	685
8	16	13	13	560	615	880	601	685
9	2	18	17	657	696	680	664	677
10	7	16	14	618	624	780	629	674
11	5	17	16	707	644	650	637	667
12	19	14	13	610	599	780	595	663
13	12	13	15	525	648	810	608	661
14	24	14	12	594	575	810	591	660
15	18	14	14	604	628	730	596	654
16	34	15	9	592	525	830	578	649
17	33	13	13	578	588	780	583	649
18	10	15	14	565	610	760	613	645
19	27	12	15	517	656	760	587	644
20	22	15	14	650	635	640	592	642
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
83	79	10	6	512	400	330	400	414
84	78	10	6	424	441	350	404	405
85	80	11	9	511	543	100	400	385
86	86	6	4	319	380	340	335	346
87	87	4	4	318	452	240	293	337

Fonte: Os autores (2021).

## Considerações Finais

Este trabalho consistiu em aplicar a TRI aos dados do Exame de Seleção 2020 do IFRN e teve como objetivo geral mostrar a possibilidade de se estimar o traço latente (também chamado de proficiência, aptidão entre outros) dos 21.039 candidatos que realizaram este exame por meio da Teoria da Resposta ao Item em relação às provas de LP e Matemática.

Após a estimação dos parâmetros, quatro itens de LP (6, 13, 16 e 20) foram removidos do processo de estimação das proficiências por não atenderem os referenciais de qualidade para o parâmetro de discriminação, sendo pouco informativos, o que implicou na função de informação do item ser praticamente uma reta horizontal em torno da informação. Em Matemática, os itens se mostraram mais discriminativos e com maior dificuldade em relação à LP.

Também foi notada uma certa variabilidade na proficiência dos candidatos quando comparada com o número de acertos nas provas. Esse é um diferencial da TRI, visto que o modo que  $\theta$  é estimado implica na chamada coerência pedagógica. Além disso, todos os candidatos de todos os campi, cursos e turnos podem ter seus escores comparáveis, sem a necessidade de fazer estratificações para este fim.

No tocante à comparação da TRI com a TCT houve mudança em praticamente todas as posições, sendo nas posições centrais as maiores discrepâncias encontradas. Ressalta-se que a TRI utiliza uma coerência pedagógica ao estimar a proficiência dos respondentes, enquanto a TCT considera que todos os itens trazem a mesma informação para o cálculo dos respectivos escores.

Como limitação de pesquisa, tem-se o fato de considerar a classificação dos candidatos apenas em termos de ampla concorrência, isto é, a Lei de Cotas, que determina o percentual mínimo de 50% das vagas para estudantes de escolas públicas, não foi considerada. Como perspectivas futuras pretende-se incluir procedimentos estatísticos na comparação da TRI com a TCT, realizar estudos sobre funcionamento diferencial dos itens e fatores associados a este contexto, uma vez que informações sociais dos estudantes também constam nos dados do Exame de Seleção do IFRN.

Portanto, este trabalho se torna relevante para que os gestores das instituições que realizam exames de seleções para a entrada de alunos possam conhecer um pouco mais sobre as vantagens e desvantagens de se implementar a TRI nos referidos

certames. Em particular, estudos como este e o realizado por Moreira (2020) mostram que é possível adotar a TRI como metodologia nos processos seletivos de alunos para o ingresso no IFRN, de preferência com a calibração dos itens para que as provas de todos os anos carreguem a mesma carga de dificuldade.

### **Agradecimentos**

Os autores deste trabalho agradecem ao Instituto Federal de Educação, Ciência e Tecnologia do Rio Grande do Norte, em particular à Coordenação de Acesso Discente, pela disponibilização dos dados referentes ao Exame de Seleção 2020.

## Referências

- ANDRADE, D. F. de; TAVARES, H. R.; VALLE, R. da C. *Teoria da resposta ao item: conceitos e aplicações*. São Paulo: SINAPE, 2000. Disponível em: [https://docs.ufpr.br/~aanjos/CE095/LivroTRI\\_DALTON.pdf](https://docs.ufpr.br/~aanjos/CE095/LivroTRI_DALTON.pdf). Acesso em: 20 abr. 2021.
- ANDRADE, J. M. de; LAROS, J. A.; GOUVEIA, V. V. O uso da teoria de resposta ao item em avaliações educacionais: diretrizes para pesquisadores. *Avaliação Psicológica*, São Paulo, v. 9, n. 3, p. 421-435, 2010. Disponível em: <http://pepsic.bvsalud.org/pdf/avp/v9n3/v9n3a09.pdf>. Acesso em: 1 maio 2021.
- ARANTES, L. J. Avaliando a aprendizagem do conceito de energia no ensino médio usando a TRI. Orientador: Ulisses Azevedo Leitão. 2016. 156 f. Dissertação (Mestrado Profissional em Ensino de Física) – Faculdade de Física, Universidade Federal de Lavras, Lavras, MG, 2016. Disponível em: <http://repositorio.ufla.br/handle/1/11022?locale=en>. Acesso em: 1 maio 2021.
- AYALA, R. J. D. *The theory and practice of item response theory*. New York: Guilford Publications, 2013.
- BAKER, F. B. *The basics of item response theory*. Washington, DC: ERIC, 2001.
- BAKER, F. B.; KIM, S. H. *The basics of item response theory using R*. New York: Springer, 2017.
- BIRNBAUM, A. L. Some latent trait models and their use in inferring an examinee's ability. In: LORD, F. M.; NOVICK, M. R. (ed.). *Statistical theories of mental test scores*. [Boston]: Addison-Wesley, 1968.
- BORTOLOTTI, S. L. V.; ANDRADE, D. F. Aplicação de um modelo de desdobramento graduado generalizado-GGUM da teoria da resposta ao item. *Estudos em Avaliação Educacional*, v. 18, n. 37, p. 157-188, 2007.
- GONÇALVES, H. de A. *Manual de metodologia da pesquisa científica*. Rio de Janeiro: Avercamp, 2005.
- HAMBLETON, R. K.; SLATER, S. C. Item response theory models and testing practices: Current international status and future directions. *European Journal of Psychological Assessment*, [S. l.], v. 13, n. 1, p. 21-28, 1997.
- INSTITUTO FEDERAL DE EDUCAÇÃO, CIÊNCIA E TECNOLOGIA DO RIO GRANDE DO NORTE. De escola de aprendizes artífices a instituto federal. In: INSTITUTO FEDERAL DE EDUCAÇÃO, CIÊNCIA E TECNOLOGIA DO RIO GRANDE DO NORTE. *Histórico*. Natal: IFRN, 2021. Disponível em: <https://portal.ifrn.edu.br/institucional/historico>. Acesso em: 15 fev. 2021.
- INSTITUTO FEDERAL DE EDUCAÇÃO, CIÊNCIA E TECNOLOGIA DO RIO GRANDE DO NORTE. Edital nº 29/2019-PROEN/IFRN. Processo seletivo para os cursos técnicos de nível médio na forma integrada. *Diário Oficial da União*, Natal, 2019. Disponível em: <https://portal.ifrn.edu.br/ensino/processos-seletivos/tecnico-integrado/exame-de>

selecao/exame-de-selecao-2019-edital-29-2018-proen-ifrn-1/documentos-publicados/edital-29-2019-exame-de-selecao-cursos-tecnicos-integrados-2020. Acesso em: 30 mar. 2020.

INEP. *Entenda a sua nota no ENEM: guia do participante*. Brasília, DF: Ministério da Educação, 2012. Disponível em: [https://download.inep.gov.br/educacao\\_basica/enem/guia\\_participante/2013/guia\\_do\\_participante\\_notas.pdf](https://download.inep.gov.br/educacao_basica/enem/guia_participante/2013/guia_do_participante_notas.pdf). Acesso em: 15 fev. 2021.

KLEIN, R. Utilização da teoria de resposta ao item no Sistema Nacional de Avaliação da Educação Básica (SAEB). *Meta: Avaliação*, Rio de Janeiro, v. 1, n. 2, p. 125-140, 2009. DOI: <http://dx.doi.org/10.22347/2175-2753v1i2.38>. Disponível em: <https://revistas.cesgranrio.org.br/index.php/metaavaliacao/article/view/38>. Acesso em: 23 mar. 2021.

LORD, F. M. *Applications of item response theory to practical testing problems*. Abingdon: Routledge, 1980.

MOREIRA, A.; NOGUEIRA, C. H. Comparação da classificação na prova da OBMEP por meio da Teoria de Resposta ao Item (TRI) e da Teoria Clássica de Testes (TCT). *Meta: Avaliação*, Rio de Janeiro, v. 12, n. 34, p. 162-180, 2020. DOI: <http://dx.doi.org/10.22347/2175-2753v12i34.2151>. Disponível em: <https://revistas.cesgranrio.org.br/index.php/metaavaliacao/article/view/2151>. Acesso em: 23 mar. 2021.

MOREIRA, C. R. B. *A teoria de resposta ao item e o ensino de matemática na educação básica*. In: SOUZA, S. A. de (org.). *O ensino aprendizagem face às alternativas epistemológicas 3*. Ponta Grossa, PR: Atena, 2020. p. 30-42. Disponível em: <https://www.atenaeditora.com.br/post-ebook/3279>. Acesso em: 23 abr. 2021.

PAEK, I.; COLE, K. *Using R for item response theory model applications*. Abingdon: Routledge, 2019.

PASQUALI, L. *Psicometria*. *Revista da Escola de Enfermagem da USP*, São Paulo, v. 43, n. especial, p. 992-999, 2009. DOI: <https://doi.org/10.1590/S0080-62342009000500002>. Disponível em: <https://www.scielo.br/j/reensp/a/Bbp7hnp8TNmBCWhc7vjbXgm/?format=pdf&lang=pt>. Acesso em: 18 abr. 2021.

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

PASQUALI, L. *TRI - Teoria de resposta ao item: teoria, procedimentos e aplicações*. Curitiba: Appris, 2018.

PASQUALI, L.; PRIMI, R. Fundamentos da teoria da resposta ao item: TRI.: *Avaliação Psicológica*, São Paulo, v. 2, n. 2, p. 99-110, 2003. Disponível em: <http://pepsic.bvsalud.org/pdf/avp/v2n2/v2n2a02.pdf>. Acesso em: 18 abr. 2021.

QGIS DEVELOPMENT TEAM. QGIS *Geographic Information System*: open source Geospatial Foundation project. [S. l.], 2016.

R CORE TEAM. *R: a language and environment for statistical computing*. Vienna, 2021.

RIZOPOULOS, D. Irm: an R package for latent variable modelling and item response theory analyses. *Journal of Statistical Software*, [S. l.], v. 17, n. 5, p. 1-25, 2006. Disponível em: <http://www.jstatsoft.org/v17/i05/>. Acesso em: 15 jan. 2021.

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*, Porto Alegre, v. 26, n. 2, p. 241-250, 2013. DOI: <https://doi.org/10.1590/S0102-79722013000200004>. Disponível em: <https://www.scielo.br/j/prc/a/PfzhXqpV4vzPYgvf75PVwcl/?lang=pt>. Acesso em: 18 abr. 2021.

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, 2013. DOI: <https://doi.org/10.18222/eae245420131902>. Disponível em: <http://publicacoes.fcc.org.br/index.php/eae/article/view/1902/1886>. Acesso em: 18 abr. 2021.

TEZZA, R.; BORNIA, A. C.; ANDRADE, D. F. de. Measuring web usability using item response theory: principles, features and opportunities. *Interacting with Computers*, London, v. 23, n. 2, p. 167-175, 2011.

WICKHAM, H. *et al.* Welcome to the tidyverse. *Journal of Open Source Software*, [S. l.], v. 4, n. 43, 2019.

WICKHAM, H.; BRYAN, J. *readxl: Read Excel Files: R package version 1.3.1*. [S. l.], 2019. Disponível em: <https://CRAN.R-project.org/package=readxl>. Acesso em: 20 jan. 2021.